# Characterization of a new horse transferrin variant

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

**Geoffrey Reayburn Paterson**

1993

## Abstract

Transferrin is a glycoprotein with a molecular weight of approximately 80 kd. Its single polypeptide chain is formed into two lobes and it is able to bind two ferric (Fe III) ions per protein molecule. Horse serum transferrin, like the transferrins of most vertebrate species, exhibits extensive genetic polymorphism.

Transferrin is one of several protein systems used for blood-typing horses. During routine blood typing a new band (designated *) was found. This variant originated from a thoroughbred stallion which was of considerable value as a sire and so it was of interest to characterize this new transferrin variant.

Thoroughbred horses carry genes for only six of the fourteen known transferrin isoforms; D, F1, F2, H2, O and R. The aim of this project was to characterize, by classical amino acid sequence analysis, the * transferrin variant and the parental variants D and F1, from one of which * must have arisen.

All three variant forms (D, F1 and *) were purified. Tryptic digests of the variants were analysed by HPLC and those peaks appearing to differ between the HPLC profiles were sequenced by automated protein sequencing. The sequences obtained confirmed that the protein isolated was a transferrin variant. Further sequencing allowed deduction of the parent transferrin variant.

Two clear sequence differences between the D and F1 variant have been identified. The F1 variant was found to contain an arginine residue at amino acid position 553, whereas the D variant contains a cysteine residue at this position. At position 418 of the F1 variant a serine residue was found and at the same position in the D variant a proline residue was found.

Sequence determination of peptides from the * tryptic digest revealed that a proline residue and a cysteine residue were found at positions 418 and 553 respectively, clearly indicating that the new * phenotype has arisen from the D allele and not the F1 allele.

## Acknowledgements

I would like to thank Miss Diane Bridger, Mr Bruce Farndale, and Dr Ian Anderson of the Equine blood-typing and Research centre for providing the horse blood essential to the project, and analysing the occasional sample with their superior gel techniques.

I am extremely grateful to Mr Julian Reid for giving me much needed support, helpful advice, encouragement and hours of computer time over the last three years.

I wish to thank Mrs Carole Flyger and Mrs Robyn Marston for bearing with my work habits in the laboratory, and keeping me up to date on all important matters.

Many thanks go to Mr Dick Poll for allowing me free rein on the FPLCs.

Finally, I am indebted to Dr Chris Moore, Dr Graham Pritchard, and Dr Malcolm Chick for their continued patience and supervision throughout (and beyond) this investigation.

## List of Abbreviations

| | |
|---|---|
| cDNA | copy deoxyribonucleic acid |
| CDT | carbohydrate-deficient transferrin |
| DEAE | diethylaminoethyl |
| DTE | dithioerythritol |
| EDTA | ethylenediaminetetracetic acid |
| Endo-F | endoglycosidase-F |
| FPLC | fast performance liquid chromatography |
| GlcNac | β-D-N-acetylglucosamine |
| HPLC | high pressure liquid chromatography |
| $MH^+$ | molecular ion |
| $Neu(4,5)Ac_2$ | *N*-acetyl-4-*O*-acetylneuraminic acid |
| Neu5Ac | *N*-acetylneuraminic acid |
| NTA | nitrilotriacetic acid |
| PAGE | polyacrylamide gel electrophoresis |
| pI | Isoelectric point |
| PNGase | Peptide-(N4-acetyl-b-D-glucosaminyl) asparagine Amidase F |
| PTH | phenylthiohydantoin |
| SDS | sodium dodecyl sulphate |
| TEMED | N N N'N'-tetramethyl ethylene-diamine |
| TFA | trifluoroacetic acid |
| TRIS | tris-(hydroxymethyl)-aminomethane |

## Amino acid abbreviations

| Amino Acid | Three letter code | One letter code |
|---|---|---|
| alanine | Ala | A |
| cysteine | Cys | C |
| aspartate | Asp | D |
| glutamate | Glu | E |
| phenylalanine | Phe | F |
| glycine | Gly | G |
| histidine | His | H |
| isoleucine | Ile | I |
| lysine | Lys | K |
| leucine | Leu | L |
| methionine | Met | M |
| asparagine | Asn | N |
| proline | Pro | P |
| glutamine | Gln | Q |
| arginine | Arg | R |
| serine | Ser | S |
| threonine | Thr | T |
| valine | Val | V |
| tryptophan | Trp | W |
| tyrosine | Tyr | Y |

# Table of Contents

## Chapter 3 Results

## Chapter 4 Discussion



## References

# List of Figures

# List of Figures (cont)

# List of Tables

# Chapter 1 Introduction

## 1.1 The transferrin family

The transferrins are a class of non-haem iron binding proteins found in the physiological fluids (and some cells) of many vertebrates. In 1947 Laurell and Ingelman purified the 'red protein' from pig plasma. In the same year they proposed the name transferrin (Holmberg and Laurell, 1947), which since has become the generic name for this protein family. The family contains lactoferrin (found in milk, secreted in tears and by leukocytes), melanotransferrin (plasma membrane of melanocytes), ovotransferrin (in egg-white), and serum transferrin (present in serum and external secretions but also found in lymphatic fluid and cerebrospinal fluid, some cell types, and the milk of some species).

Transferrin is mainly synthesized in the liver. In man the level of serotransferrin in the serum is about 2-4 mg/ml. It has been noted that in horses, transferrin levels change with age; serum transferrin levels of foals are significantly higher than those of adult horses. However no difference has been found in transferrin levels between male and female horses (Schmid et al., 1990). Lower levels have been observed in horses with infections, acute laminitis and cirrhosis of the liver, and raised transferrin levels occur in horses with anemia. (Schmid et al., 1990).

Bezkorovainy (1980) suggested criteria that may be used to classify a protein as a transferrin: (1) it is freely soluble in water, (2) it may bind a maximum of two ferric (Fe III) ions per protein molecule, (3) it consists of a single polypeptide chain which is formed into two domains (or lobes) and has a molecular weight of approximately 80 kd, and (4) the iron-laden form has a characteristic absorption in the visible region of the spectrum (responsible for the characteristic red-brown colour) with a maximum at about 460 nm. Chasteen (1983) suggested additional characteristics relating to the iron-binding property of the transferrins: (1) the N- and C-terminal halves of the single chain polypeptide each contain a similar, but not identical, iron-binding site, (2) iron-binding is strong enough to resist removal in the extracellular fluids, yet weak enough to allow release within specific intracellular compartments, (3) binding of a synergistic anion (usually carbonate or bicarbonate) is essential for iron to bind to the transferrin molecule.

## 1.2 Structure of transferrin

Transferrins are glycoproteins with molecular weights in the range of 72000 to 83000 (Welch, 1990).

The amino acid sequence of human transferrin (676 amino acids) has been determined by MacGillivray et al., (1983). The N-lobe (residues 1-336) and the C-lobe (residues 337-679) can be aligned, by introducing gaps in appropriate locations, revealing a sequence homology of 42% between the two lobes. Gene duplication may give rise to this internal homology.

Human transferrin has a molecular weight of 79570, 6% of which is carbohydrate (Thorstensen and Romslo, 1990). This carbohydrate component comprises two covalently attached polysaccharide moieties, both of which are attached to the C-terminal half of the protein (Spik et al., 1988). The carbohydrate is linked by N-glycosidic bonds to specific asparagine residues in the protein sequence. Human transferrin has 8 disulphide bridges in the N-lobe and 11 in the C-lobe.

Human lactoferrin was the first member of the transferrin family to have its tertiary structure reported. Anderson et al., (1987) elucidated the structure of diferric human lactoferrin at a resolution of 3.2 Å (which has subsequently been refined to 2.8 Å, Anderson et al., 1989).

The tertiary structure of rabbit serum transferrin has been determined by X-ray diffraction analysis at 3.3 Å resolution (Bailey et al., 1988). The structures of human lactoferrin and rabbit serum transferrin were found to be very similar in terms of their overall geometry and folding patterns. Rabbit transferrin is made up of two homologous lobes, each of which binds one ferric ion. The lobes are joined by a linear bridging region. Each lobe is made up of two domains, and the iron binding site is located within the interdomain cleft (see Fig 1.1). The molecule is stabilized by nineteen disulphide bridges. An interesting structural feature is the N-terminal part of domain one of the N-lobe. This region is made up of two $\alpha$-helices and two $\beta$-strands in a planar configuration, stabilized by two disulphide bonds (Bailey et al., 1988). It is a discrete structure situated remotely from the iron binding site and does not include iron-binding.

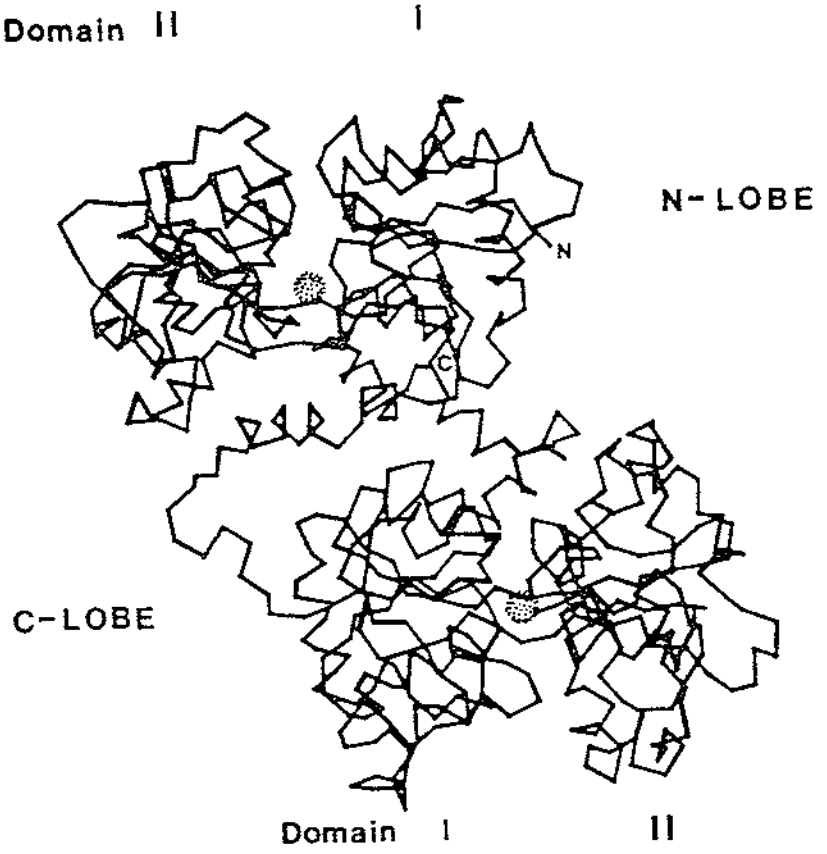


**Figure 1.1** The a-carbon chain structure of rabbit transferrin

The structurally homologous N-lobe and the C-lobe of the rabbit transferrin molecule, each contain two dissimilar domains (domains I and II), showing the transferrin molecule to consist of four structurally distinct regions. A cleft between domains I and II in each lobe contains the iron-binding site. The iron atoms are represented as spheres. From Sarra et al. (1990).

The amino acid sequence of this region is homologous to that of the lymphocyte-derived transforming proteins ChBlymn-I and HuBlym-I, which suggests that this region may be associated with the role of transferrin in the stimulation of cell growth and proliferation (Bailey et al., 1988).

The cDNA sequence for horse transferrin (Carpenter, 1992) encodes a protein of 706 residues, including a signal sequence of 19 amino acids. The amino acid sequence predicted from this cDNA information has the duplicated structure, conserved iron-binding and cysteine residues which are characteristic of the transferrin family. This is consistent with the structural and functional similarities within the family. Horse transferrin has 73.4% amino acid sequence identity to human transferrin, 61.9% identity with human lactoferrin, and 72.8% identity with rabbit transferrin.

## 1.3 Binding of iron to transferrin

The iron binding sites are located in the clefts between the domains that make up each lobe . Structural studies have indicated that the co-ordination of the iron is essentially the same in both lobes. The iron-binding sites are buried deeply within the folded protein. There are five ligands that bind the iron atom. Four of these ligands are provided by amino acid side chains and the fifth is supplied by a non-protein ligand. The four protein ligands are provided by one carboxylate oxygen (Asp60 and Asp395; N- and C-terminal lobes respectively), two phenolate oxygens (Tyr92, Tyr192 and Tyr435, Tyr528) and one imidazole nitrogen (His253 and His597). The non-protein ligand is usually a bicarbonate ion (see Fig 1.2).

In rabbit transferrin the iron is bound by two tyrosine residues (numbers 95 and 188 in the N-lobe), one histidine (249) and one aspartic acid (63) (see Fig 1.2).

The transferrins can bind only one ferric ion per domain, i.e. at the most two ferric ions per protein molecule. Physiologically only a third of all the available iron-binding sites are filled at a time. The iron-binding site of serum transferrin is much closer to the surface of the molecule than is the case in human lactoferrin, and it will tend to lose its iron more readily. Serum transferrin will lose its iron at a pH below about 5 (with optimal binding at approximately pH 8), whereas human lactoferrin loses its iron below pH 3.

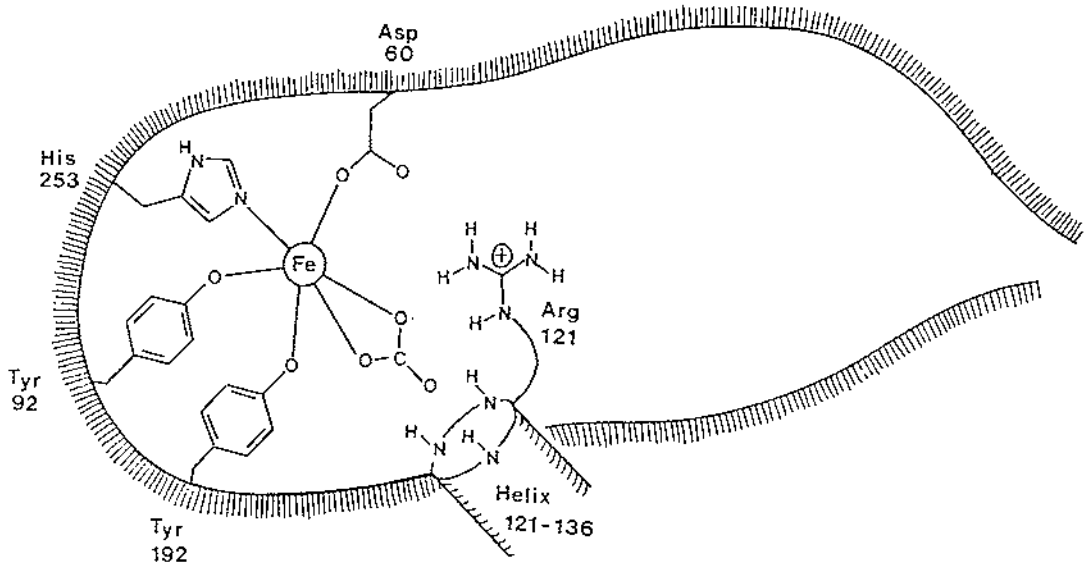


**Figure 1.2** The iron-binding site of human lactoferrin

This diagram represents the N-lobe and the C-lobe of the human lactoferrin molecule, from Anderson et al. (1989). The C-lobe has the same arrangement, as do the N and C-lobes of the rabbit transferrin molecule (Sarra et al., 1990). The ferric ion and carbonate ion are shown.

Transferrin occurs physiologically in four different forms: apotransferrin, monoferric transferrins with iron bound in either the C- or N-lobe, and diferric transferrin (de Jong & van Eijk, 1989).

For each iron atom bound a carbonate ion is also bound and three protons are released. This results in an increase in the net negative charge of the protein and it is easily distinguished from its apo-transferrin counterpart upon electrophoresis. One would thus expect the diferric transferrin to bind more firmly to an anion chromatographic column such as DEAE(diethylaminoethyl)-cellulose. However this is not the case. Iron-saturated serotransferrin is eluted first, followed by serotransferrin containing a single iron atom, and lastly apotransferrin (Lane, 1971). This unexpected behaviour is explained by a surface charge redistribution that occurs upon iron binding to apotransferrin. Chemical modification methods have shown that apotransferrin contains 1 phenolic and 2 amino groups that are unreactive i.e buried, whereas the corresponding numbers for iron-saturated transferrin are 8 and 11 respectively (Bezkorovainy and Zschocke, 1974). These results suggest that iron-saturated transferrin is more compact than apotransferrin. Iron-binding causes a number of other changes in the physico-chemical properties of the transferrins (Bezkorovainy, 1980a).

## 1.4 Biological functions of transferrin

The metabolic significance of this protein is accentuated by the fact that mutations resulting in the absence of transferrin are very rare. The congenital absence of transferrin (atransferrinaemia) has been described in only a small number of children and results in severe anaemia, increased tissue iron stores and early death (Heilmeyer et al., 1961; Goya et al., 1972).

The function of the glycan moieties is unclear. It has been suggested that the carbohydrate groups may have a role in determining the rate of elimination of transferrin from circulation (Wong and Regoeczi, 1977).

### 1.4.1 Iron transport

Iron is an essential nutrient for the growth and survival of all eukaryotic cells. Free iron present in most physiological fluids is likely to be in the insoluble $Fe^{3+}$ oxidation state. Consequently, iron *in vivo* must be bound by proteins which can stabilize it in a soluble

form, making it available for biological processes. Transferrin binds iron deep within clefts in the protein, allowing it to be transported in a soluble form.

Transferrin may have a role in the transport of metal ions other than iron, such as $Al^{3+}$ (van Ginkel et al., 1990). Transferrin facilitates the entry of $Al^{3+}$ into the central nervous system (Roskams & Connor, 1990).

Serum transferrin, as its name suggests, is involved in the transport of iron in the serum and extravascular fluids. From synthesis in the liver the protein enters the circulation system and picks up and delivers iron to various cells around the body via receptor-mediated endocytosis, which is the accepted model for the transport of iron by transferrin (Huebers and Finch, 1987). Iron-bound transferrin recognizes and binds a specific receptor on the cell surface. The iron-transferrin-receptor complex is then internalized within an endocytotic vesicle. The receptor is thought to facilitate the release of iron from transferrin once inside the vesicle due to a lower pH in the vesicle as compared to that at the cell surface where the iron-transferrin-receptor complex was formed. Apotransferrin has a lower affinity for the receptor than iron-saturated transferrin so it is readily released from the receptor after iron removal (Kornfeld, 1969). The transferrin receptor is a glycoprotein of molecular weight 170 000 - 200 000, made up of two identical subunits (Trowbridge and Omary, 1981 (human)).

Iron bound to transferrin is delivered to cells by receptor-mediated endocytosis (Huebers and Finch, 1987). Transferrin, bound to receptors on the cell membrane, is then internalised into a non-lysosomal acidic compartment of the cell where the iron is released.

### 1.4.2 Growth factor

Horse serotransferrin has been identified as the growth promoting factor of human myeloid Leukemia cells (Yoshinari, 1989). It has also been found to be an essential component in serum-free media for the growth of many other cell types including a rat pituitary cell line, GH3 (Hayashi & Sato, 1976), and a human cervical cell line HeLa (Hutchings & Sato, 1978). It is thought that this property of transferrin to promote growth is a consequence of its ability to supply iron. Evidence has been produced to support the theory that iron is essential for DNA synthesis, and that transferrin functions as a growth factor solely by supplying iron (Laskey et al., 1988). However, a specific growth factor function has also been suggested for transferrin that is unrelated to its iron-

binding ability (Brock & Mainou-Fowler, 1983; Seligman, 1983). Tissue specific expression of transferrin also occurs in situations where it is required as a growth factor (Bowman et al., 1988).

Investigations concerning transferrin receptors have revealed that large numbers are found on virtually all dividing cells and receptor numbers decline as cell division ceases. These results suggest that receptor expression is related to cell proliferation. Thus transferrin receptors can only be detected in peripheral blood lymphocytes after antigen growth stimulation. It is now standard procedure to determine the growth potential of *in vivo* tumours by measuring transferrin receptor expression (Crichton, 1991).

The binding of iron to transferrin has been found to alter the functional capabilities of the transferrin molecules. The findings of Murate (1988) show that iron-saturated transferrin is active, while iron-free transferrin is inactive in causing spontaneous differentiation of human erythroleukemia K562 cells (Murate et al., 1988 ).

### 1.4.3 Bacteriostasis

Along with the ability of transferrin to promote growth its antimicrobial activity has also been attributed to its iron-binding properties. It was in fact this activity that led to the discovery of the transferrins. Micro-organisms require iron for growth and must compete for iron with the iron-binding proteins of the host. The virulence of many bacteria has been attributed, in part, to their ability to sequester iron from the environment (Bullen et al., 1982). On the other hand, transferrin, with its high affinity for iron, can retard microbial growth by making iron relatively unavailable. Transferrin also performs a more active role. It is thought to release lipopolysaccharide from the outer membrane of Gram-negative bacteria causing damage to the membrane and altering its permeability (Ellison et al., 1988).

### 1.4.4 Functionally abnormal transferrin isoforms

Several functionally abnormal forms of transferrin have been reported. A functionally deficient isoform of human transferrin has been detected (Evans et al., 1988). An amino acid change of a glycine residue to an arginine residue at position 394 in the type C transferrin resulted in less tight binding of iron to the C-terminal lobe. It is thought that the large sidechain of the substituted arginine residue may interfere with the aspartate

residue at position 392 which is an iron ligand. Another abnormal isoform, also with an iron-binding deficiency, has replaced an asparagine with an isoleucine, the position of which is uncertain (Welch & Langmead, 1990).

Abnormalities in glycosylation have also been found in humans affected by alcohol abuse. Schellenberg (1989) in a comparative evaluation of various diagnostic tests for alcohol abuse concluded that the carbohydrate-deficient transferrin (CDT) assay developed by Stibler et al. (1986) was the best test available for the diagnosis of alcohol abuse. The cause of this carbohydrate deficiency is thought to be due to the accumulation of ethanol and/or its metabolites in the body inducing a reduction in the activity of α-2,6-sialyltransferase responsible for the sialylation of serotransferrin (Malagolini et al., 1989). Altered glycosylation of serotransferrin is also seen in patients with hepatocellular carcinoma, in this case the alteration of the carbohydrate moieties is seen to be various including the increase of highly branched sugar chains (Yamashita et al., 1989).

The finding of a silent allele (no expression) in a thoroughbred mare and transmission to five of her offspring has been reported by Schmid and Braend, (1990). Another silent allele, first found in a thoroughbred stallion and provisionally designated Tf, has an electrophoretic protein migration pattern which is outside the traditionally defined range of transferrin variation (Bowling et al., 1991). The allele behaves codominantly with reduced penetrance or expressivity ("null" expression) in some circumstances.

## 1.5 Genetic Polymorphism

The present study arises from the finding that horse serotransferrin, like the transferrins of most vertebrate species, exhibits extensive genetic polymorphism. Over evolutionary time different variations of the same gene have arisen, due to mutation, resulting in the presence of two or more alleles of a particular gene in a population. The very similar yet non-identical proteins that result from this are termed isoforms or variants.

Transferrin polymorphism has been found in most species studied (Morgan, 1981). Twenty two allelic forms of human transferrin have so far been discovered. Most of the alleles are uncommon, the predominant type C transferrin occurring in more than 90% of the population (Kuehnl and Speilmann, 1978).

These variants can be separated on gel electrophoresis as different amino acid compositions change the overall charge of the proteins. Since polymorphisms are

hereditary, gels can provide a fingerprint for individual identification and parental verification. However one particular polymorphic locus is not enough for this purpose. For blood-typing to be statistically sound, several different loci are necessary. The common loci used for blood-typing horses are; prealbumin (Pr), postalbumin (Pa or Xk), transferrin (Tf), and serum esterase (Es) (Putt & Whitehouse, 1983). Other important genetic markers in horses are; albumin, $\alpha_1$-protease inhibitors, vitamin D binding protein, $\alpha_1$B-glycoprotein (Xk protein), ceruloplasmin, plasminogen, haptoglobin, complement component 3, serum protein 1, and apolipoprotein A-IV (serum protein 2). Juneja et al (1989) give the following reasons as to why these proteins are particularly useful for linkage studies and for routine parentage testings:

(1) a majority of these proteins are highly polymorphic in most breeds
(2) most are visualized by inexpensive general protein staining and
(3) most can be phenotyped in serum samples that have been frozen for years.

### 1.5.1 Polymorphism of horse transferrin

According to Bell et al. (1988) and Cothran et al. (1991) fourteen common transferrin variants have been identified so far in the horse; D, D2, D3, F1, F2, F3, G, H1, H2, J, M, O, R, and X. (D2, D3, F3, G, and X have not been fully characterised).

All variants show autosomal co-dominant inheritance. Electrophoretic patterns of pure transferrins show that each protein has a strong anodal band and weaker cathodal band while several weaker bands can also be detected (Stratil and Glasnak, 1981). The D2 (or C) is unique among transferrin variants since its two major bands are of equal intensity on gel electrophoresis (Scott, 1980; Stratil & Glasnak, 1981). The appearance of several bands of a purified transferrin variant upon electrophoresis is known as transferrin microheterogeneity. The occurrence and cause of this microheterogeneity has been studied in many different species (Stratil and Spooner, 1971; Sratil & Kúbek, 1974; Spooner et al., 1975; Maeda et al., 1980; Tsuji et al., 1984; Chung and McKenzie, 1985). In cattle, it is thought that microheterogeneity is linked to both the structure of the glycan moiety and the amino acid sequence of the polypeptide chain (Maeda et al., 1980). However, there is, as yet, no convincing evidence to show that horse transferrin microheterogeneity is also linked to the modification of the polypeptide chain, whereas there is much evidence for glycan moiety linkage (Coddeville et al., 1989).

From time to time new alleles are found and characterised. Known alleles are also being found in different horse breeds. e.g. D and Hp found in the horse breed *Equus przewalskii* in 1982 (Putt & Whitehouse, 1982). In 1990 it was concluded that the J transferrin isoform could be used as a genetic marker for Spanish Thoroughbred horses (Rodriguez-Gallardo et al., 1991).

Thoroughbred horses carry genes for only six of the fourteen known transferrin isoforms; D, F1, F2, H2, O and R. Among these, five are quite widely distributed in the whole equine species with a marked predominance of the F isoforms. However, in a few breeds, the frequency of the F variant falls below 0.30, and the frequency of the R variant alone or of both the O and R variants reach high frequencies. For example, the R variant in the Döle horse has a frequency of 0.66 (Braend, 1964); together the O and R variants have a frequency of 0.45 in the Icelandic pony (Hesselholt, 1966); and the R variant has a frequency of 0.72 in the Tokara breed Nozawa et al., 1976).

The occurrence and frequencies of twelve transferrin alleles were determined for nineteen different horse breeds. The only alleles found in all nineteen breeds were F2 and R. F1 was found in six breeds, and was most frequent in thoroughbreds. The null allele (Tf) was only found in Anglo-Arabs (Yokohama et al., 1989).

### 1.5.2 The D and R transferrin groups

The fourteen transferrin variants may, by various methods, be divided into two groups. The R group, containing the R, O and M variants, and the D group comprising the rest. The O, R and M transferrins act as homologous antigens while transferrin D antisera recognizes only D group isoforms as homologous antigens (Kaminski et al., 1981) (F3, G and X were not known when testing took place). The D group isoforms move much faster on electrophoresis (towards the anode) than do the R group isoforms (Kaminski et al., 1981). The D group has a strong band to weak band ratio of about four whilst the R group ratio is only around two (Watanabe et al., 1989). Amino acid substitutions or deletions are thought to determine the electrophoretic differences between the two groups of variants. Carpenter (Ph.D thesis, 1992) has detected six amino acid substitutions between the D and R groups. These amino acid substitutions change the overall charge of the protein and therefore account for the difference in electrophoretic mobility of the two groups. The significant immunological partition between the "faster migrating" and "slower-migrating" variants can also be accounted for by these amino acid substitutions as they are clustered on an external part of the molecule (Carpenter, 1992). Kaminski

had previously put forward the hypothesis that an ancestral transferrin gene underwent duplication, and that its two parts have since evolved independently by point mutations or other events (Kaminiski et al., 1989). Carpenter's research serves to substantiate this theory. It has also been noted that horses with the O and R alleles have lower levels of serum iron than those with exclusively other transferrin alleles (Yokohama et al., 1982).

## 1.6 Glycosylation of transferrins

Transferrin is a glycoprotein. Hudson et al. (1973) studied the monosaccharide content of several different species. They determined that bovine transferrin probably contains 1 glycan moiety per polypeptide chain while rabbit and equine transferrin contain 2 glycans and porcine transferrin contains 4 glycans per polypeptide chain.

Human transferrin contains two heteropolysaccharide units. The two carbohydrate moieties of human serum transferrin are located on the C-terminal lobe. These glycans may be both biantennary (82%), both triantennary (1%), or one of each (17%), (Spik et al., 1988). The biantennary glycan of human transferrin is shown in Fig 1.3 (Spik et al., 1975). A schematic diagram of the complete human lactoferrin molecule showing the positions of carbohydrate attachment for a variety of transferrin types including rabbit and human transferrin is shown in Fig 1.4 (Baker et al., 1987).

The carbohydrate content of transferrin varies between species, and also between tissues within a species. Differences in carbohydrate contents tend to be most easily recognized as variations in numbers of sialic acids. Since sialic acid residues are highly charged this will cause protein separation on gel electrophoresis. The degree of complexity of transferrin heterogeneity varies with animal species and sialic acid content is either partly or wholly responsible for it (Chung and McKenzie, 1985). Variation in sialic acid content is responsible for all isoform variation in chicken and sheep (Williams, 1962; Spooner et al., 1975). Some species contain transferrins with no carbohydrate content at all, yet heterogeneity still occurs (e.g. carp Valenta et al., 1976). Amino acid sequence or some other unknown factor must be the cause of heterogeneity in these species. Sialic acids are only partially responsible for the heterogeneity that occurs in cattle and pigs (Stratil and Spooner, 1971; Maeda et al., 1980; Stratil and Kúbek, 1974). Human transferrin type C occurs as three bands upon electrophoresis. Experiments using neuraminidase to cleave the sialic acid residues show that these bands differ in sialic acid content (Regoeczi et al., 1977). Bovine transferrin has 0 to 5 sialic acid residues per

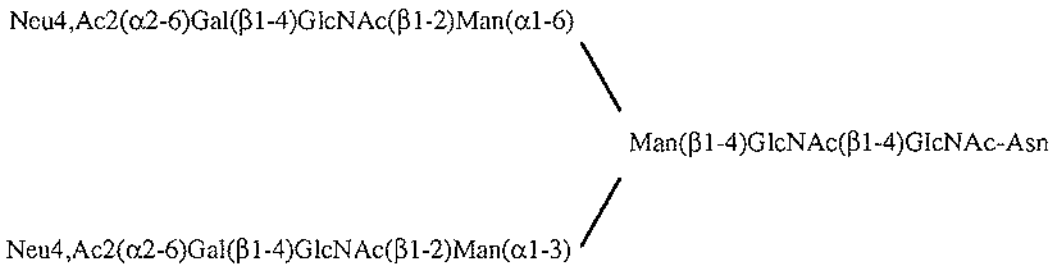


**Figure 1.3** General stucture of the biantenary carbohydrate moiety identified in human serum transferrin

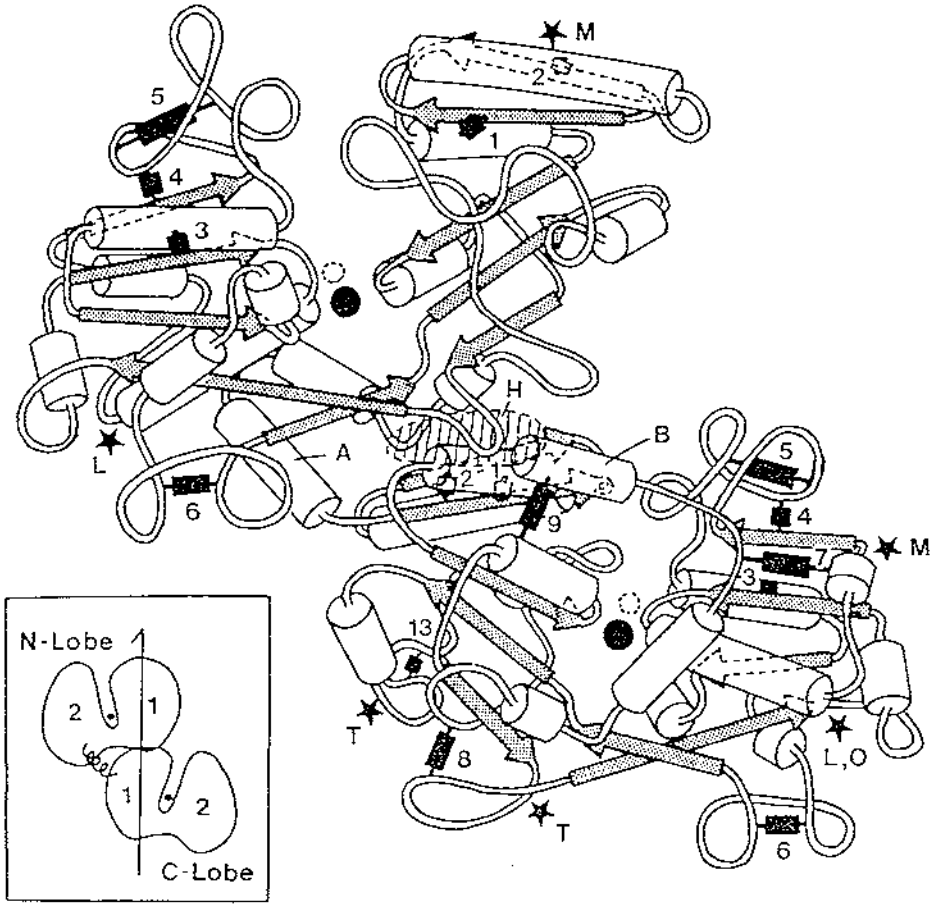


**Figure 1.4** Schematic diagram of the complete human lactoferrin molecule

The positions of carbohydrate attachment are marked. O, ovotransgferrin; T, human transferrin; L, human lactoferrin; R, rabbit transferrin; M, melanotransferrin. The connecting helix is labelled as A, and the C-terminal helix as B. The disulphide bridges are indicated by solid lines, and the iron- and bicarbonate-binding sites as filled or open circles respectively. From Baker et al. (1987).

molecule (Stratil and Spooner, 1971). Rat transferrin has either 2 or 3 sialic acid residues per molecule (Schreiber et al., 1979).

Horse serotransferrin is a glycoprotein containing two N-glycosylation sites in its C terminal domain (Carpenter, 1992). Homozygous transferrin variants can be fractionated into two main bands and a few minor bands. Nomenclature of the two major bands of horse transferrin is analogous to that which was used for cattle transferrin (Stratil and Spooner, 1971). They named the slow component (with respect to the mobility of the bands towards the anode at alkaline pH) 2a and the fast component 4b. The number referred to the probable number of sialic acid residues that could be split off the transferrin glycan moieties by neuraminidase, and the letter referred to the residual differences between the bands after the sialic acids had been removed (Stratil and Glasnák, 1981). Interestingly neuraminidase only partially desialylated transferrins D and R under conditions that were shown to completely remove the sialic acids of bovine transferrin (Chung and McKenzie, 1985). There is no explanation for this but it is also the case for human transferrin C (Wong et al., 1978).

A few years later Stratil et al., (1984), determined that the major and minor bands of the transferrin O variant had the same amino acid compositions but differed in their carbohydrate composition. Component 2a contained 10 moles of sugar per mole of protein (4 hexoses, 4 hexosamines and 2 sialic acids) while component 4b contained 20 moles of sugar per mole of protein (8 hexoses, 8 hexosamines and 4 sialic acids). The carbohydrates were identified as mannose and galactose, N-acetylglucosamine and N-acetylneuraminic acid. Molecular weights of the components 2a and 4b were determined to be 75 200 and 80 500 respectively. This difference in molecular weight, along with the carbohydrate differences, suggested that component 2a had one diantennary glycan, while component 4b had two (Stratil et al., 1984).

In 1989, Coddeville et al., published four different glycan structures (type I, type II A, type II B, and type III) that they identified in components 2a, 4b, and the first minor band studied 5b. This work confirmed the previous finding that 2b had one glycan and 4b had two glycans. The minor variant 5b was also found to contain 2 glycans (i.e. 4 sialic acids). Coddeville et al., (1989) also found that microheterogeneity is dependent not only on the number of glycans but on the nature of the sialic acids present in the biantennary structure. The four following types of sialic acids had previously been found in horse transferrin: Neu5Ac, Neu5Gc, Neu(4,5)Ac2 and *N*-glycolyl-4-*O*-acetylneuraminic acid, in the relative ratio 35: 5: 50: 10, respectively.

Differences in the molecular weights of the three transferrin components 2a, 4b, and 5b are not completely explained by the changes in glycan structure and number suggesting that the polypeptide chain sequences might also differ. As yet there is no other evidence for this but examination at the amino acid sequence level may explain the residual differences in electrophoretic mobility between the transferrin components 2a, 4b, and 5b. Differences in electrophoretic mobility between some of the transferrin variants have been attributed to amino acid substitution. For example the mobility differences between the D and R isoforms have been ascribed to substitution of Asp and Glu residues in the D isoform by Gly residues in the R isoform (Chung and McKenzie, 1985).

The glycosylation of the transferrins may be of functional importance. Upon finding Neu(4,5)Ac2 in horse serum glycoproteins it was proposed that this component could play a role in the inhibition of Asian influenza virus A2 haemagglutination since the influenza virus A2 neuraminidase is able to cleave the linkages of this sialic acid (Pepper, 1968).

Kornfeld (1968), revealed that glycosylation does not seem necessary for the efficient transport of iron by transferrin. However, it has been shown to be important in determining the rate of elimination of transferrin from circulation. Removal is less rapid when the sialic acid content is high (Wong and Regoeczi, 1977). The cause of microheterogeneity may be due to the differentiated loss of carbohydrate groups during circulation in the blood and may also reflect the age of the protein.

## 1.7 Linkage of transferrin genotype to phenotype

One reason for studying polymorphisms is to look for associations between biochemical markers and inherited abnormalities i.e. linkage of genotype to phenotype. This correlation of transferrin genotypes with traits of horses has been attempted by various groups. When transferrin types were compared with racing performance in thoroughbreds, an excess of FF phenotypes was detected among race winners (Osterhoff et al., 1974). On the other hand Andersson et al., (1987) did not find any significant relation between the transferrin type and the racing performance of the Swedish Trotter breed. Weitkamp et al., (1991) found an association between transferrin heterozygosity and endometrial health, an important component of fertility in mares.

The relationship between serotransferrin polymorphism and different growth and reproductive traits in some bovine breeds has been studied. The only significant finding

was the effect on AFO (age at first oestrus). The earliest AFO was found to be in cows homozygous for the E allele (Singh and Choudhary, 1989). Studies on equine reproductive traits in relation to transferrin allelic linkage has shown for four breeds (Arab, Thoroughbred, Russian Trotter and Budyonny) that conception rates are slightly higher when mating parents with different transferrin genotypes, 76-94% as opposed to 50-80% for parents with the same transferrin types (Starodumov, 1989). Conception rates in Arabs were highest for mares homozygous in all four loci tested (transferrin, albumin, esterase, and the D blood group system), whereas for Russian Trotter horses heterozygous mares had the highest conception rate (Dubrovskaya and Starodumov, 1989).

## 1.8 Background to the Present Study

The blood typing of New Zealand-born Thoroughbreds takes place at the Equine Blood Typing and Research Centre located at Massey University. Thoroughbred horses genetically carry only six of the twelve known serotransferrin isoforms; D, F1, F2, H2, O and R. However during routine blood typing (using the standard starch gel electrophoresis technique) a new band was found. The band ran slightly anodal to the H2 transferrin isoform and did not correspond to any of the known transferrins, nor was it present in either parent (both of which were genotypically DF1). The progeny was recognized as having inherited the transferrin variant D along with the new protein and thus given the genotype D* (* is the standard notation given for a new variant not yet recognized as such). Researchers from the Equine Blood Typing and Research Centre have speculated that the new variant arose as a result of a genetic change that occurred at the transferrin locus in the gamete of one or other parent. Analysis of 111 progeny of D* showed the new allele to be codominant with the other alleles, as 55 inherited *, and 56 inherited D, i.e. the alleles separate with frequency of 0.5. This new allele is of international significance as the horse (D*) is very valuable commercially which will therefore ensure the distribution of this new allele throughout the New Zealand and world wide thoroughbred population.

## 1.9 Aims of this study

The aims of this research project were;

(1) to purify the putative transferrin variant * and the transferrin variant F1 from which it may have been derived. The other transferrin variant (D) from which * may have been derived was already available in purified form.

(2) to partially characterise the two transferrin variants D and F1, and the putative transferrin variant *; firstly, to confirm that * is in fact a transferrin variant; secondly, to deduce from which parent transferrin variant (either D or F1) the new variant has been derived; and thirdly, to try to determine specific sequence differences between the * transferrin variant and the parent transferrin variant that would account for the difference in the electrophoretic mobilities between the new isoform and the parent variant.