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 α-helices and two β-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.

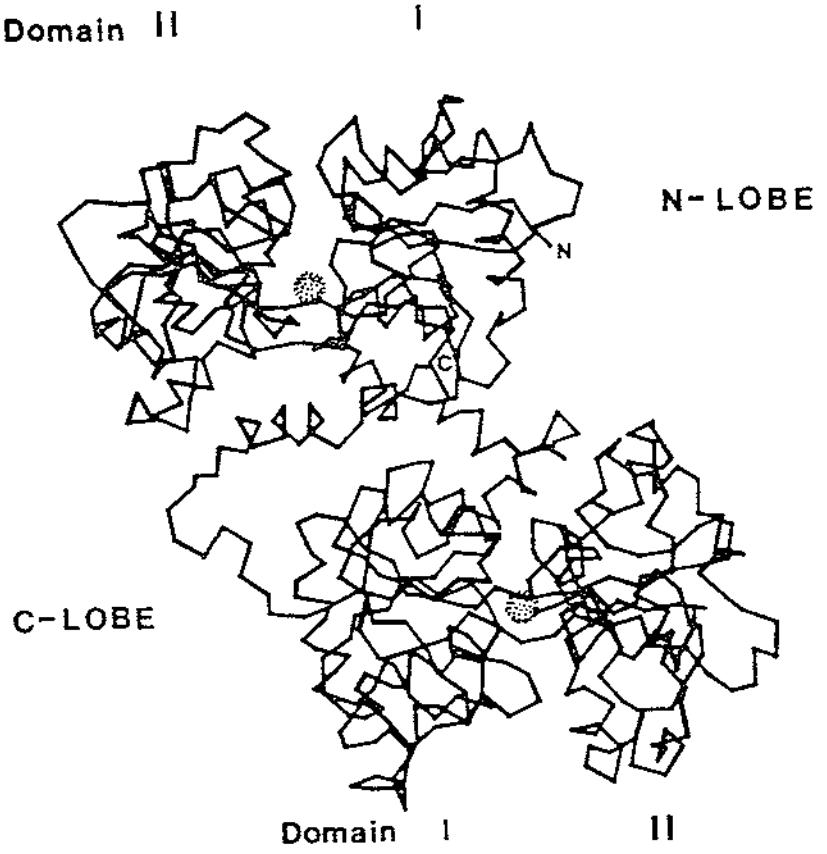

Figure 1.1 The a-carbon chain structure of rabbit transferrin

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.

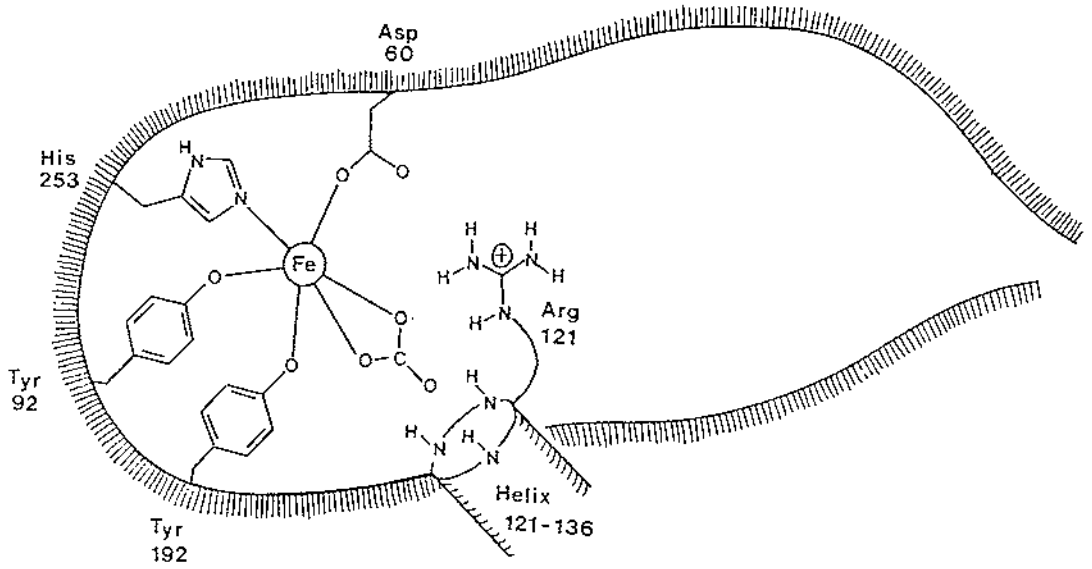

Figure 1.2 The iron-binding site of human lactoferrin

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, α_1-protease inhibitors, vitamin D binding protein, α_1B-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

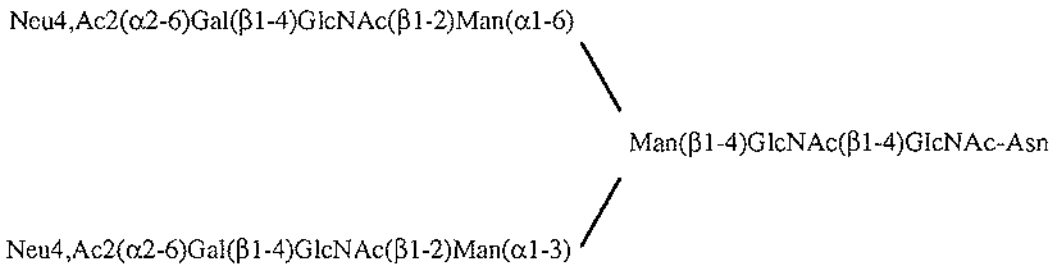

Figure 1.3 General stucture of the biantenary carbohydrate moiety identified in human serum transferrin

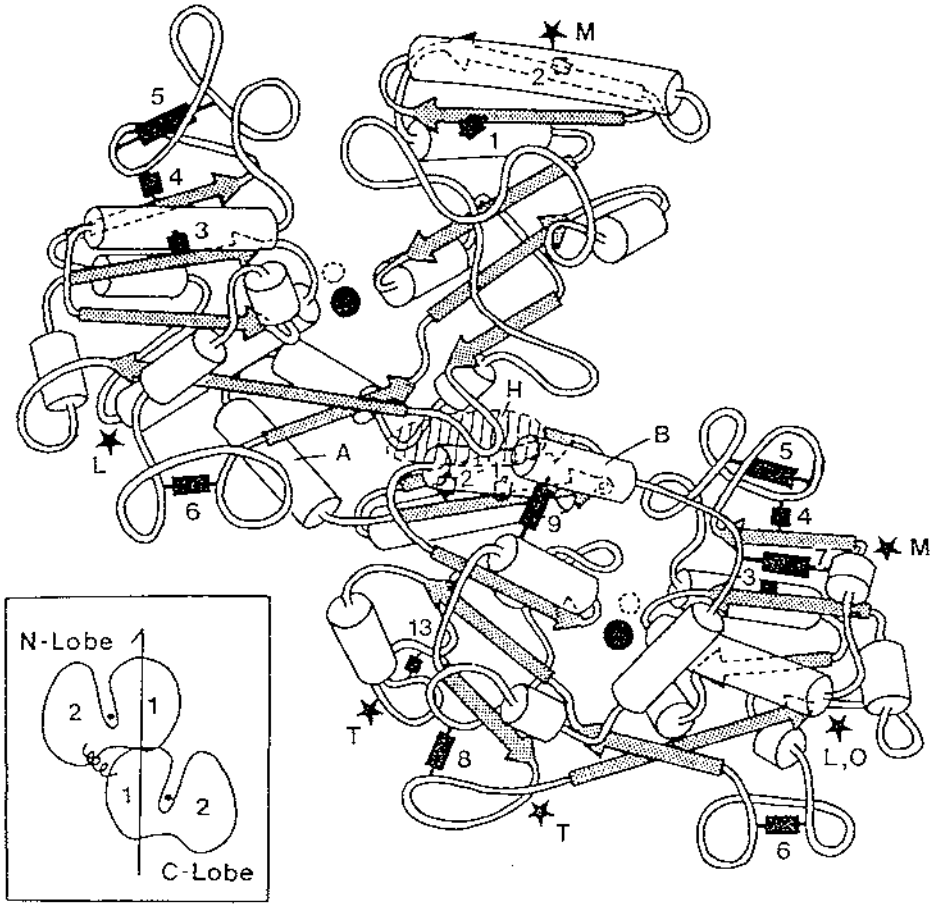

Figure 1.4 Schematic diagram of the complete human lactoferrin molecule

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.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Horse Blood

Horse blood was obtained from the Equine Blood Typing and Research Centre, Massey University, Palmerston North.

2.1.2 Chromatographic resins

Diethylaminoethyl (DEAE) cellulose was obtained from Whatman Biochemicals, England. Q-Sepharose was supplied by Pharmacia, Sweden.

2.1.3 Thoroughbred horse transferrin sequence

An amino acid sequence of thoroughbred horse transferrin was supplied by M. Carpenter (Crown Research Institute, Palmerston North) and may now be obtained through Genbank, number M69020. The amino acid sequence was predicted from cDNA (copy deoxyribonucleic acid) sequences obtained from the transferrin genes of a horse heterozygous for the transferrin variants D and F1. The sequence was pieced together from three separate clones and as a result the sequence presented in the results section (Fig 3.18) may be derived from both D and F1 cDNA sequences.

2.1.4 Miscellaneous Materials

Ethanolamine was supplied by Merck, Germany. Nitrilotriacetic acid was supplied by Aldrich chemical company, USA. Tris-(hydroxymethyl)-aminomethane, dithioerythritol, acrylamide, bis-acrylamide, and TEMED (N,N,N',N'-Tetrametyl-ethylenediamine) were supplied by Serva, New York. Guanidine hydrochloride and ammonium persulphate were supplied by BDH chemicals, England. Trypsin was supplied by Sigma, U.S.A. HPLC (high pressure liquid chromatography) grade acetonitrile was supplied by Mallinckrodt, U.S.A. Isopropanol, also HPLC grade was supplied by May and Baker, England.
All other reagents were, wherever possible, of analytical grade.

2.2 Methods

2.2.1 Spectrophotometric Assays

Fractions from purification columns were routinely scanned from 350nm to 550nm (using a Shimadzu 160 spectrophotometer) to identify the transferrin-containing fractions and the absorbance at the wavelength of maximum absorbance (460nm) was used as an index of the transferrin content. Since this measurement was used only to determine the relative concentration of transferrin during purification, calibration to determine absolute concentration was not attempted. There are no published values of the molar absorption coefficient for horse transferrin

During purification relative protein concentrations were determined by obtaining absorbance readings at 280nm. A Hitachi U-1100 spectrophotometer was employed for this purpose.

2.2.2 Iron Saturation

Iron saturation of the transferrin was accomplished using a solution containing $FeCl_3$ and nitrilotriacetic acid (NTA) prepared as follows.

Iron-NTA stock solution

$FeCl_3.6H_20$	270 mg
NTA	270 mg
distilled water to 1 litre	

Iron-saturating solution

Iron-NTA stock solution	1 ml
tri-sodium citrate (0.1M)	10 ml
sodium bicarbonate (0.1M)	10 ml

2.2.3 Polyacrylamide Gel Electrophoresis (PAGE)

The following stock reagents were prepared and stored at 4°C

30% Acrylamide:

Acrylamide	29.2 g
Bis-acrylamide	0.8 g
Milli Q water up to	100 ml

Running Gel Buffer

Tris	36.3 g
Adjust to pH 8.8 with 1M HCl	
Milli Q water up to	100 ml

Stacking Gel Buffer

Tris	6.0 g
Adjust to pH 6.8 with 1M HCl	
Milli Q water up to	100 ml

Electrode Buffer (pH 8.3)

Tris	15.0 g
Glycine	72.0 g
Milli Q water up to 5L	

Sample Buffer

Stacking gel buffer	1.0 ml
Glycerol	0.8 ml
0.05% (w/v) bromophenol blue	200 µl
2-β-mercaptoethanol	0.4 ml
Milli Q water up to	8.0 ml

Mixtures used for preparation of 10% non-denaturing gels were as follows.

Running gel

30% acrylamide	16.8 ml
Running Gel Buffer	12.5 ml
TEMED	20 µl

Milli Q water up to	50 ml
Ammonium persulphate	25 mg

Stacking gel

30% acrylamide	1.3 ml
Stacking gel buffer	2.5 ml
TEMED	10 µl
Milli Q water up to	10 ml
Ammonium persulphate	5 mg

Solutions for staining were made up as follows:

Stain

1% Coomassie Blue R-250 in water	62.5 ml
Methanol	250 ml
Acetic Acid	50 ml
Milli Q water up to	500 ml

Destain

Methanol	500 ml
Acetic Acid	50 ml
Milli Q water	500 ml

A Mighty Small II slab gel electrophoresis unit (Hoefer Scientific Instruments) was used for rapid polyacrylamide gel electrophoresis. 10% polyacrylamide gels were made and run in accordance with the Mighty Small II instruction manual. The running gels were 6cm deep and 8cm wide with a stacking gel with wells for 10 samples. Appropriately diluted protein samples were mixed with an equal volume of sample buffer then loaded onto the gels in 2-25µl amounts. The gels were run at 10mA until the samples had entered the running gel, when the current was increased to 20mA. Electrophoresis was stopped when the dye-band had reached the end of the gel.

A larger PROTEAN II slab cell apparatus (Bio-Rad) was also used for PAGE when greater resolution was desired. 10% native gels were run on this apparatus in accordance with the manufacturer's specifications. The running gels were 12cm deep and 16cm wide with a stacking gel that could accommodate 15 samples. Equal volumes of sample and sample buffer were mixed then loaded onto the gels in varying amounts between 2 and

60µl. Each gel was run at 13mA until the dye was seen to enter the resolving gel upon which the current was increased to 18mA. The gels were often run for two hours after the dye band had run off the bottom of the gels to obtain greater band separation.

2.2.4 HPLC (High Pressure Liquid Chromatography)

Peptide separation was carried out on a Vydac 218 TP C18 reverse-phase column [30nm pore size with dimensions 250 x 4.6mm] attached to a Spectra-Physics SP8800 Liquid Chromatograph.

Tryptic peptides were weighed out and made up to solutions of 4mg/ml. All samples were filtered through a 0.22µm Millipore filter before injection.

Trifluoroacetic acid (TFA)/acetonitrile buffer systems were used as chromatographic solvents. Loading buffer (buffer A) consisted of redistilled water and 0.1% TFA, whilst the eluting buffer (buffer B) contained 0.08% TFA in acetonitrile. Details of elution gradients used are given in the appropriate section of the results.

All solvents were HPLC grade and filtered through a 0.22µm Millipore filter before use. During use, the solvents were continually degassed with helium.

2.2.5 2-Pyridylethylation

20mg of purified transferrin were made up to 1ml in water and then 0.2ml of 1M Tris/HCl buffer at pH 8 was added. To this solution was added 1.14g guanidine-HCl (to make the final mixture 6M) and 5mg dithioerythritol (DTE) (to give a concentration in molar excess of the cysteine content of the protein) and the mixture was left under nitrogen about an hour to ensure that complete reduction of disulphide groups had occurred. Then 200µl of 4-vinyl pyridine (to give a concentration in molar excess of the cysteine content of the protein) was added to react with thiol groups and left to stand for another hour. After completion 500µl of β–mercaptoethanol was added to couple left over vinyl pyridine and prevent any further reaction from taking place.

2.2.6 Tryptic Digests

Samples for digestion were dialysed against 1% (w/v) ammonium bicarbonate to bring the solution to pH 8. The dialysed sample was treated with 0.2mg trypsin in a round bottomed flask at 37 °C for seven hours, after which the digest was freeze-dried.

2.2.7 Peptide Sequencing

Peptide sequencing was performed on an Applied Biosystems 470A gas-phase protein sequencer, connected to 120A PTH Analyser. All sequencing programs included three cycles of prewash to remove any contaminants followed by the standard peptide sequencing program which was an automation of the Edman method. The phenylthiohydantoin (PTH) derivatives were separated by HPLC (i.e. using the 120A PTH Analyser), and identified by retention time with reference to a standard amino acid mixture.

Samples were loaded in 15-90µl aliquots (containing about 1nmol).

Buffers and other reagents used were in accordance with the manufacturer's specifications.

2.2.8 Mass Spectrometry

The identity of the peptides was confirmed by comparing the mass calculated from the sequence with that determined by mass spectrometry. Mass spectra were determined using a VG 70-250 double focussing magnetic sector mass spectrometer (VG Analytical, Manchester, UK) fitted with a liquid secondary mass spectrometry ion source and caesium ion gun. Peptides were analysed in a matrix of acidified glycerol. Analysis was carried out by John Allen and John Shaw, AgResearch, Palmerston North.

2.2.9 Storage of Plasma Samples

Horse blood samples were centrifuged in a Sorval centrifuge at 8000g for 20 minutes and the supernatant (i.e. plasma) decanted and stored in 100 ml batches at -20° C until required.

Chapter 3 Results

3.1 Purification of transferrin variant F1

The major aim of this project was to carry out a sequence comparison of the putative transferrin variant * and the transferrin isoforms found in the blood of the parent horses i.e. D and F1. The first step was to obtain pure samples of the three different transferrins. The horse transferrin variant D had previously been purified from blood of a horse known to be homozygous for the D allele (C. H. Moore & G. Borrie, unpublished work). Transferrins from the other parent, variant F1, and the putative transferrin variant * still needed to be isolated. Homozygous F1 type blood was obtained and from this the F1 variant was isolated. The purification of the F1 variant will be described first. Minor deviations from this procedure which were involved in the purification of the * variant will be described in a later section.

3.1.1 Ammonium Sulphate Precipitation

Blood from a horse homozygous for the F1 allele was obtained from the Equine Bloodtyping and Research Centre, and plasma obtained as outlined in section 2.2.10.

A sample (100ml) of frozen plasma was thawed and solid ammonium sulphate was added (27.7g per 100ml) with stirring at 4° C to bring it to 45% saturation. The precipitated protein was sedimented by centrifugation at 23 500g for 10 minutes. The supernatant (containing the transferrin) was decanted and dialysed against 3 successive 5L volumes of 0.05 M Tris/HCl (pH 8.0) containing 0.03M NaCl (DEAE loading buffer) over a period of two days.

3.1.2 DEAE Cellulose Chromatography

After pretreatment of DEAE cellulose with 0.5M HCl and 0.5M NaOH and washing as specified in the manufacturers instructions, a 5cm x 38cm column of this anion exchanger was equilibrated with 0.05M Tris/HCl buffer (pH 8.0) containing 0.03M NaCl until the pH and conductivity of the effluent was the same as that of the buffer. Iron-nitrilotriacetic acid (iron-NTA) solution (21ml) was added to the dialysed 45% ammonium sulphate

supernatant (159ml) to iron-saturate the transferrin which was then loaded onto the DEAE column using a Pharmacia P-3 peristaltic pump. A three litre gradient former containing 1.5L of loading buffer in the first compartment and 1.5L of loading buffer with 0.5M KCl in the second compartment was used to generate a linear elution gradient. Elution was carried out at 1ml min^{-1} and 15ml fractions were collected.

The relative transferrin content of the fractions was determined by reading the absorbance at 460nm. Transferrin was eluted between fractions 30 and 65 corresponding to a conductivity range of about 4-4.5 mmho (Figure 3.1). The A_{460} values are shown only up to fraction 60 in Fig 3.1. The A_{460} values increase considerably in the later fractions but this increase is due to the high content of oxyhaemoglobin (absorbance maximum at 414nm). This is evident in Fig 3.2 which shows a native PAGE of every fifth sample across the elution profile after staining with Coomassie Blue R-250. Samples from fractions 65 and above show numerous protein components and particularly a high content of haemoglobin in later fractions. On the other hand fractions 35 to 60 contain mainly transferrin which elutes very early, presumably because of weaker binding to the anion exchanger due to its low isoelectric point. Although the elution of transferrin and its separation from haemoglobin could usually be observed visually the identification of the transferrin-containing fractions was always confirmed by scanning fractions spectrophotometrically from 350nm to 550nm to establish that there was a true transferrin peak at 460nm.

Most of the transferrin elutes in a major peak (fractions 30-45) but following this there is a second smaller peak from fractions 46-60 (see Fig 3.1). The native PAGE confirms that this smaller peak is due to transferrin (Fig 3.2). The presence of two peaks may be due to microheterogeneity of the F1 isoform (see below). Alternatively it may be that the particular blood sample used for the purification of F1 was not from a homozygote. The second possibility seems very unlikely since the relative quantities of the two peaks are very different and gels of plasma samples from the source of this blood, run by the Equine Blood Typing and Research Centre did not indicate the presence of a different transferrin isoform.

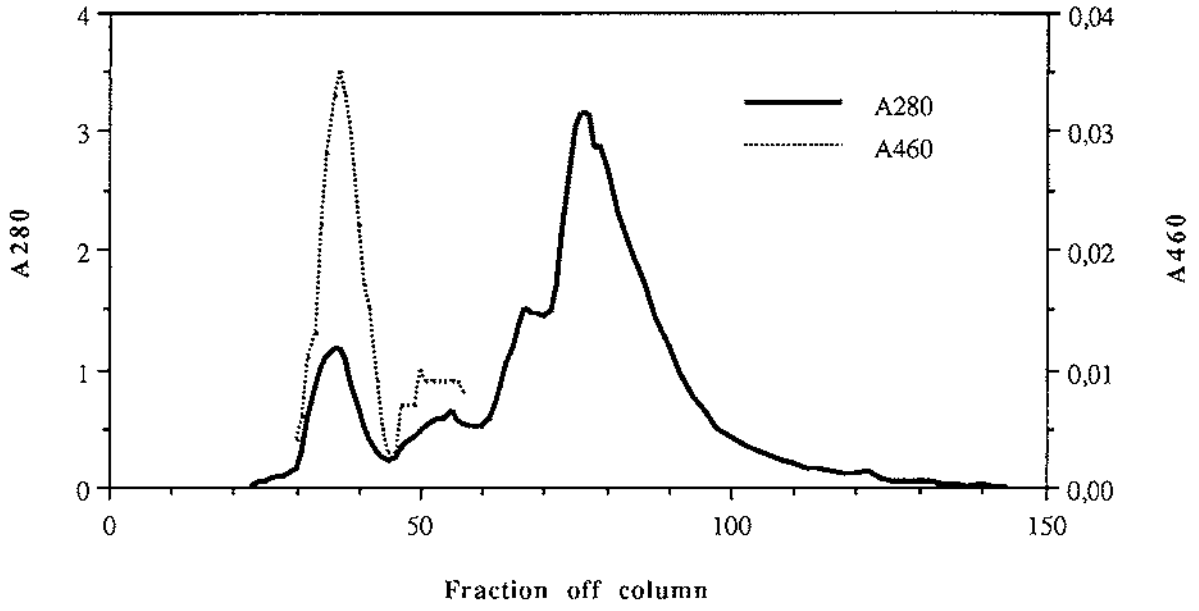

Figure 3.1 Elution profile of transferrin from a DEAE anion exchanger column.

Conditions as described in section 3.1.2. Absorbance readings at a wavelength of 460nm are shown for fractions 30-60. Fractions 30-65 were pooled.

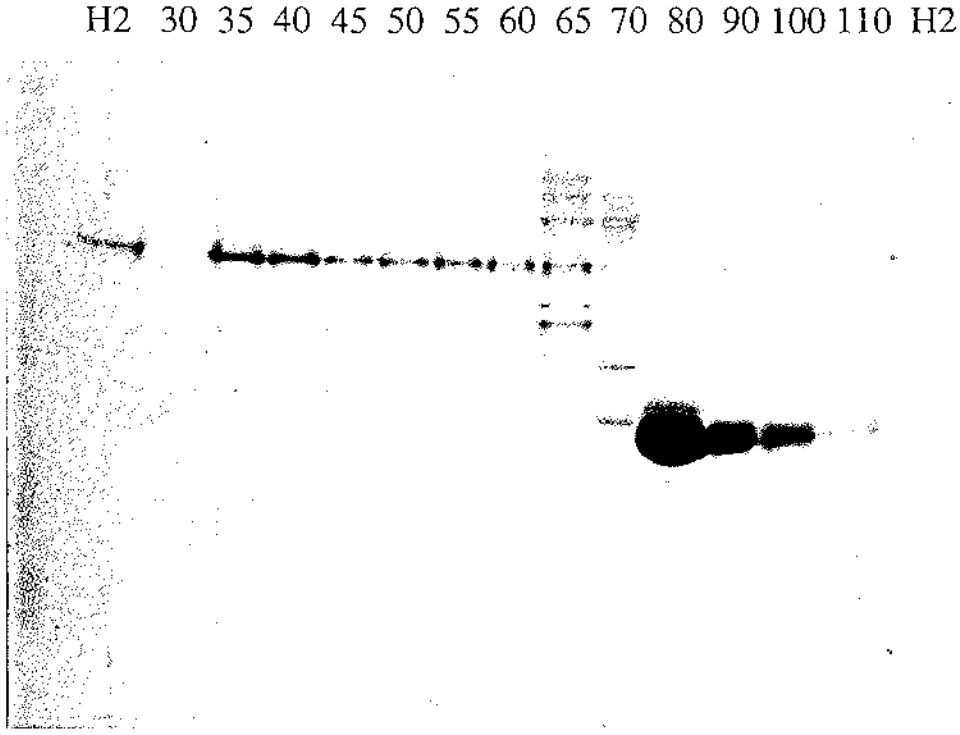

Figure 3.2 Native-PAGE (10% gel) showing the protein content of fractions off a DEAE column.

Samples from every fifth fraction from the DEAE cellulose chromatography described in Fig 3.1 were analysed by PAGE. An equal volume (20µl) from each fraction was loaded onto the gel. The numbers above the lanes refer to the fraction number. Fractions 30-65 were pooled. H2 is a transferrin marker.

All developed gels were stained with Coomassie Blue R-250.

3.1.3 Q-Sepharose Chromatography

The pooled transferrin-containing DEAE fractions were then concentrated (to a volume of 43ml) by ultrafiltration (XM 50 membrane) and loaded onto a Q-Sepharose anion exchange column (11cm x 2.4cm) attached to a Bio-Rad "Econosystem". The Q-Sepharose column was equilibrated with 20 mM triethanolamine buffer (pH 7.5) before loading the sample. Elution was accomplished with a linear gradient of 0-0.5M KCl using a flow rate of 2 ml min^{-1}. The major protein peak eluted at about 33% B (B=0.5M KCl) with smaller peaks at about 45% and 53% B (Fig 3.3). The A_{460} readings indicate that the transferrin was present in the major protein peak. Fractions 28-42 (i.e. 27-41% B) were pooled. Native gel electrophoresis of fractions from this profile (Fig 3.4) showed that the main protein peak eluting at 33% B contained only a single protein with a mobility corresponding to that of the transferrin marker.

This protein was used as the source of the transferrin in subsequent sequence analysis studies. The purification of transferrin by the three steps described above is illustrated in the native PAGE profile shown in Fig 3.5 and in the histogram Fig 3.6 which shows an increasing ratio of A_{460}/A_{280} in successive purification steps.

3.1.4 Heterogeneity of the transferrin variant F1 eluted from the Q-Sepharose column

Closer inspection of the PAGE analysis of the Q-sepharose fractions containing the transferrin (Fig 3.7) showed that the mobility of the protein was slightly faster in the later eluting fractions. In order to investigate the possible basis for this difference in mobility, individual fractions from across the transferrin peak were analysed by ion exchange chromatography on a Mono Q column in 20 mM triethanolamine buffer (pH 7.5) eluting with a linear gradient of 0-0.5M KCl (see Fig 3.8). In fraction 29 from the Q-Sepharose chromatography an apparently single protein eluting from the Mono Q column at 23% B was present (B=0.5M KCl). In fraction 31 two peaks (designated I and II) were present; the position of peak I corresponded to the single peak in fraction 29, whereas peak II eluted at about 26% B. The presence of two components in fraction 31, clearly separable by FPLC, was surprising since PAGE analysis of this fraction showed only a single component (Fig 3.7 lane 3). However, a more heavily loaded gel (Fig 3.9) does show the presence of two bands in this fraction. In fraction 34 peak II was the major component present. Then in later fractions decreasing amounts of peak II and a new peak (designated III and eluting at 29% B) appeared. The peaks (I,II, and III) were collected

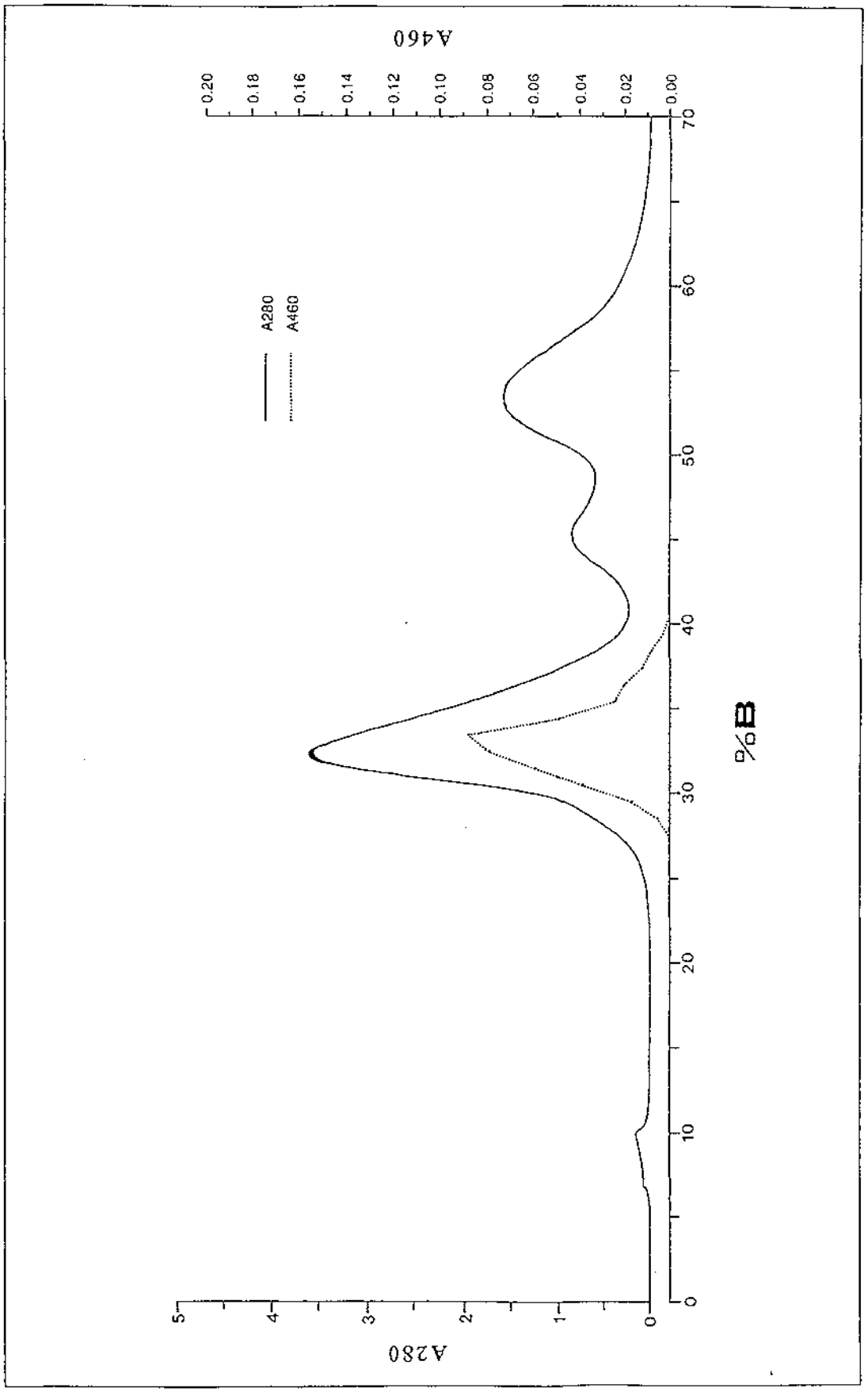

Figure 3.3 FPLC elution profile of the transferrin from a Q-Sepharose column.

Fractions between 27% B and 41% B were pooled. Conditions as described in section 3.1.3. B = 0.5M KCl

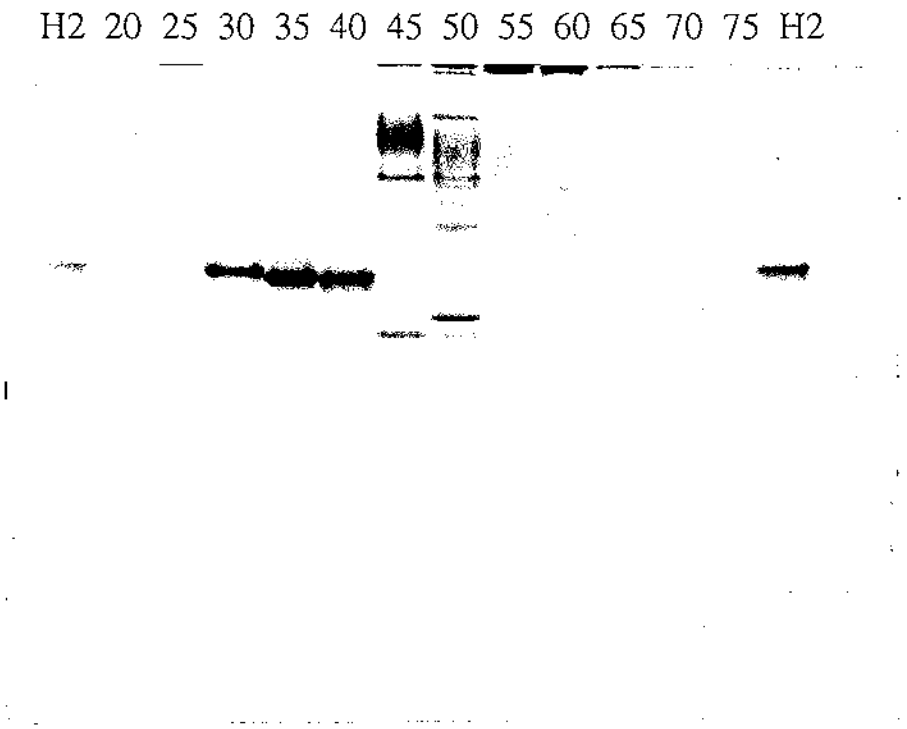

Figure 3.4 Native-PAGE (10% gel) showing the protein composition of fractions off a Q-Sepharose column.

A Q-Sepharose column was eluted at 2ml/min and 8ml fractions were collected Samples from every fifth fraction were analysed by PAGE. 20µl of each fraction was loaded onto the gel except for fractions 30 and 40 of which only 10µl of each fraction was loaded, and fraction 35 of which only 5µl was loaded. The numbers above the lanes refer to the fraction number. Fractions 25-42 were pooled. H2 is a transferrin marker.

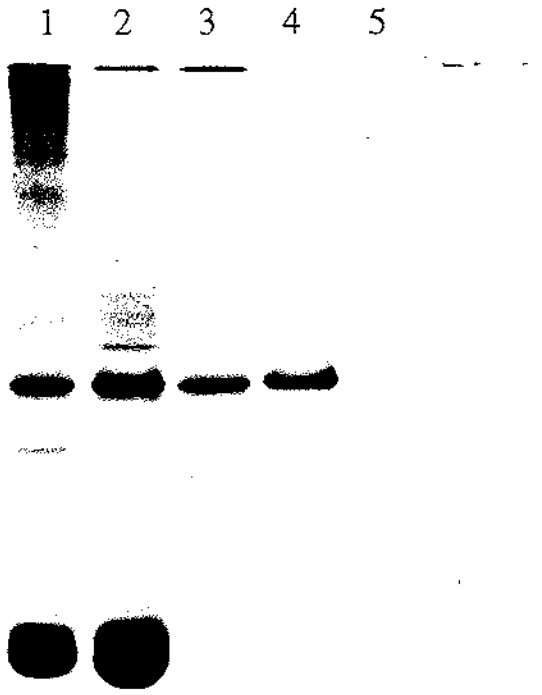

Figure 3.5 Photo of 10% native-PAGE showing the composition of samples taken at each step from the purification of transferrin F1.

Lane 1 plasma sample
Lane 2 45% ammonium sulphate supernatant
Lane 3 DEAE sample (pooled fractions)
Lane 4 Q-sepharose sample (pooled fractions)
Lane 5 transferrin marker H2

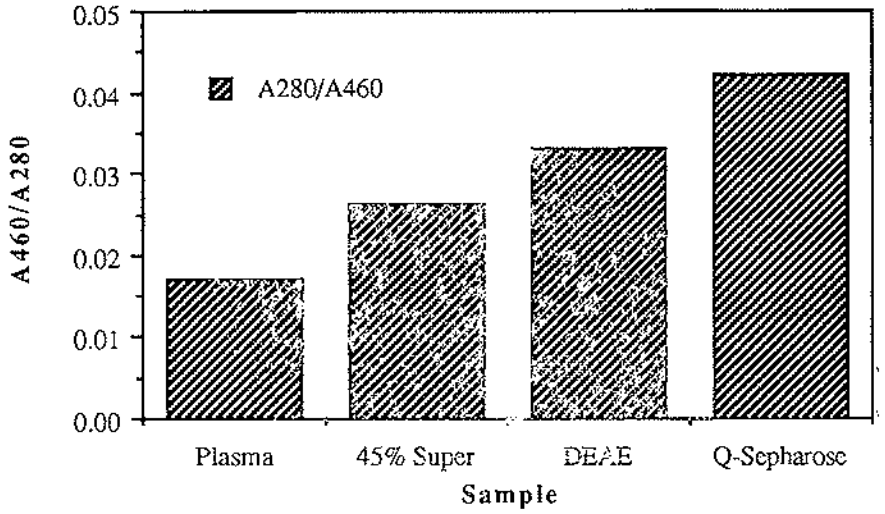

Figure 3.6 Histogram showing the absorbance at 460nm (transferrin) relative to the absorbance at 280nm (protein) in each step of the purification procedure.

(Average data from two purification runs.)

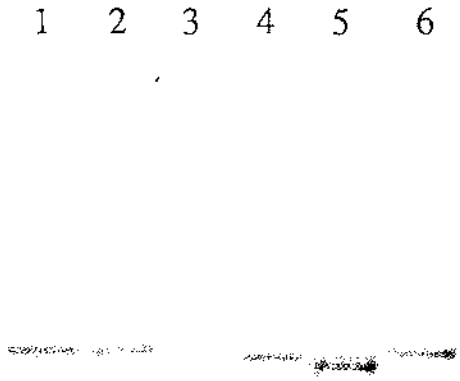

Figure 3.7 10% native-PAGE stained with Coomassie Blue showing the variation in mobility of the transferrin in fractions off a Q-Sepharose column.

Lanes 1 and 6 transferrin marker H2
Lane 2 fraction 29
Lane 3 fraction 31
Lane 4 fraction 34
Lane 5 fraction 38

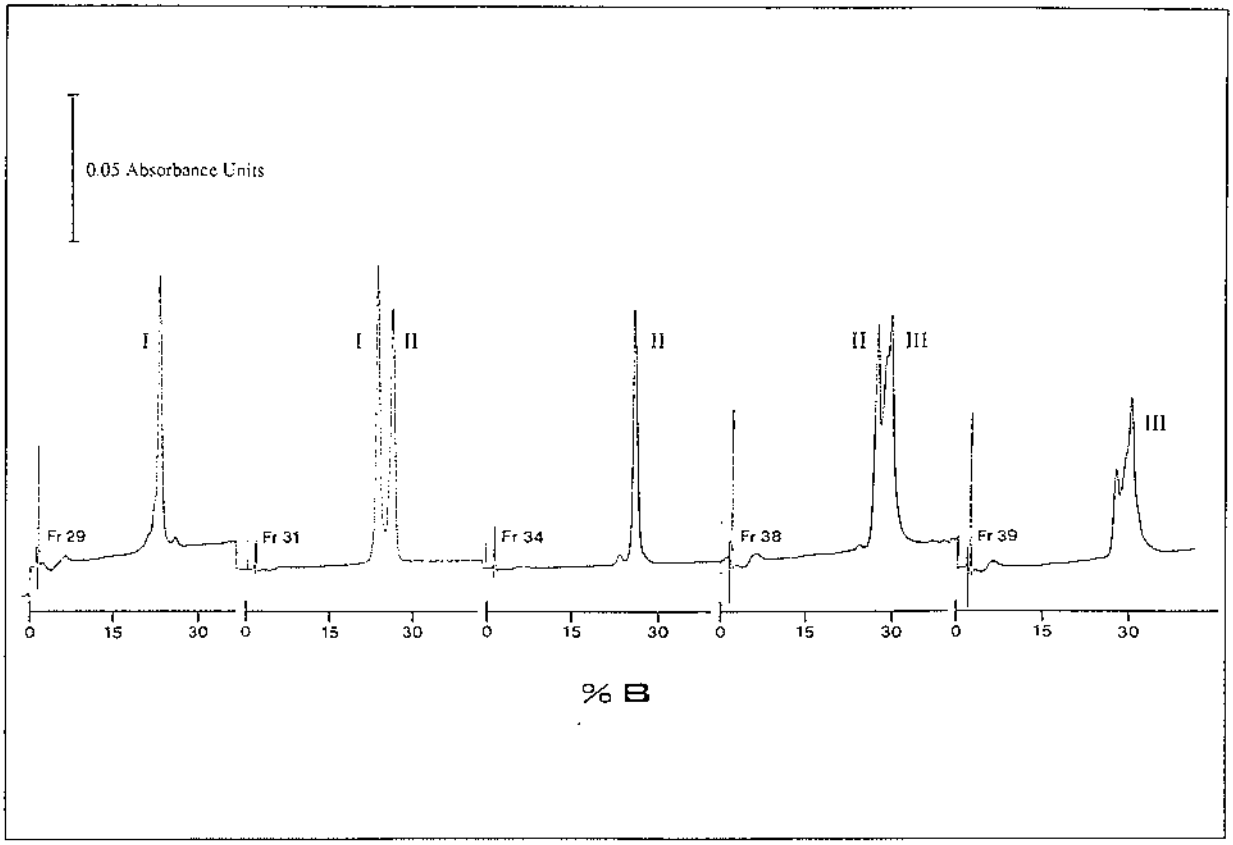

Figure 3.8 Absorbance profiles (280nm) obtained by FPLC on a Mono Q anion exchanger of transferrin-containing fractions from a Q-Sepharose column.

All five FPLC traces were run at an AUFS of 0.2, however the amounts injected were varied. The amount injected is shown after the fraction number. B = 0.5M KCl.

Fraction 29 (100ml) shows a single peak eluting at 23% B.

Fraction 31 (100ml) shows two peaks, I and II. Peak I elutes at 23% B (corresponding to the peak in fraction 29) and peak II elutes at 26% B.

Fraction 34 (20ml) shows a peak eluting at 26% B (corresponding to peak II in fraction 31).

Fraction 38 (100ml) shows the presence of peak II and a new peak eluting at 29% B (peak III).

Fraction 39 (200ml) shows the reduction of in the concentration of peaks II and III.

and the sequence of the first ten residues from the N-terminus was determined. The N-terminal sequences obtained are given below:

Peak I :	EQT(V)RW(C)TVS
Peak II :	EQTVRW(C)TVS
Peak III :	EQTVRW(C)TVS

These sequences were obtained by N-terminal sequencing of transferrin that had not undergone tryptic digestion or treatment with 4–vinyl pyridine. Therefore the cysteine residues (which are probably present in the actual protein sequences (see Fig 3.20)) were unable to be identified by HPLC. The bracketed valine residue is probably present in the actual protein sequence (see Fig 3.20) but it was not identified by protein sequencing.

Together these results suggest that the I, II, and III peaks all contain transferrin. Since the blood sample used was confirmed to be from an F1 variant by the Equine Blood Typing and Research Centre, the multiple peaks revealed by Mono Q chromatography could be due to transferrin microheterogeneity (i.e. slightly different forms of the same variant).

Horse blood transferrin microheterogeneity has been reported previously (Coddeville et al., 1989) to be related to variation in the number of glycans and the presence of different types of sialic acid in these glycoproteins. To investigate whether the differences in transferrin mobility on native PAGE (Fig 3.7) and the multiple transferrin peaks from the Mono Q column (Fig 3.8) were due to differences in glycosylation of the transferrin an attempt was made to remove the carbohydrate moieties by the use of deglycosylating enzymes. Two such enzymes were available in the department. The enzyme Endo-F catalyses hydrolysis of the second glycoside bond in glycan side chains so that a single β-D-N-acetylglucosamine (GlcNac) monosaccharide remains attached to an asparagine. The enzyme PNGase cleaves the entire carbohydrate group from the protein forming an aspartic acid residue from the glycan-linked asparagine.

Fractions 29, 31, 34, and 38, from the Q-Sepharose column were dialysed overnight in distilled water then freeze-dried. 2mg of residue from each fraction was then taken up in 40µl of 0.03M Tris/HCl pH 7.0 containing 5mM EDTA. 10µl of an Endo-F/PNGase mixture [crude extract prepared by Dr G. E. Norris (1989) based on the method of Elder and Alexander (1982)] was added to each reaction mixture which was left to react at 37° C for 1 hour. A Mono Q column (using the same buffer system as described above) was then employed to analyse the Endo-F/PNGase-treated and untreated samples. PAGE

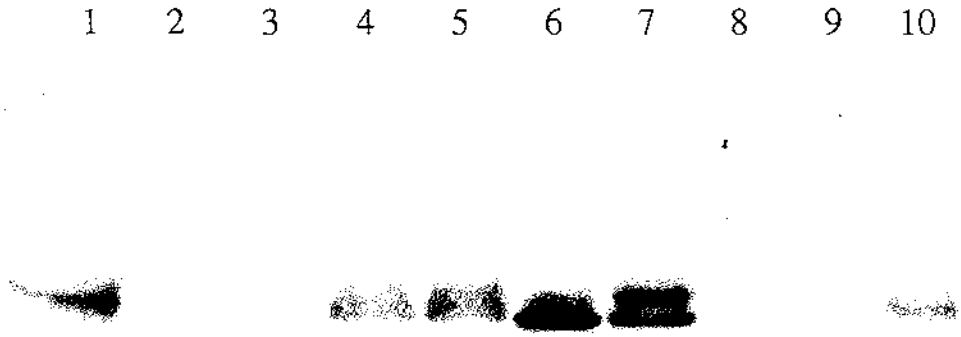

Figure 3.9 10% native-PAGE stained with coomassie blue showing the effect of the EndoF/PNGase digestion of transferrin.

Lanes 1 and 10 transferrin marker H2
Lane 2 Fraction 29
Lane 3 Fraction 29 treated with EndoF/PNGase
Lane 4 Fraction 31
Lane 5 Fraction 31 treated with EndoF/PNGase
Lane 6 Fraction 34
Lane 7 Fraction 34 treated with EndoF/PNGase
Lane 8 Fraction 38
Lane 9 Fraction 38 treated with EndoF/PNGase

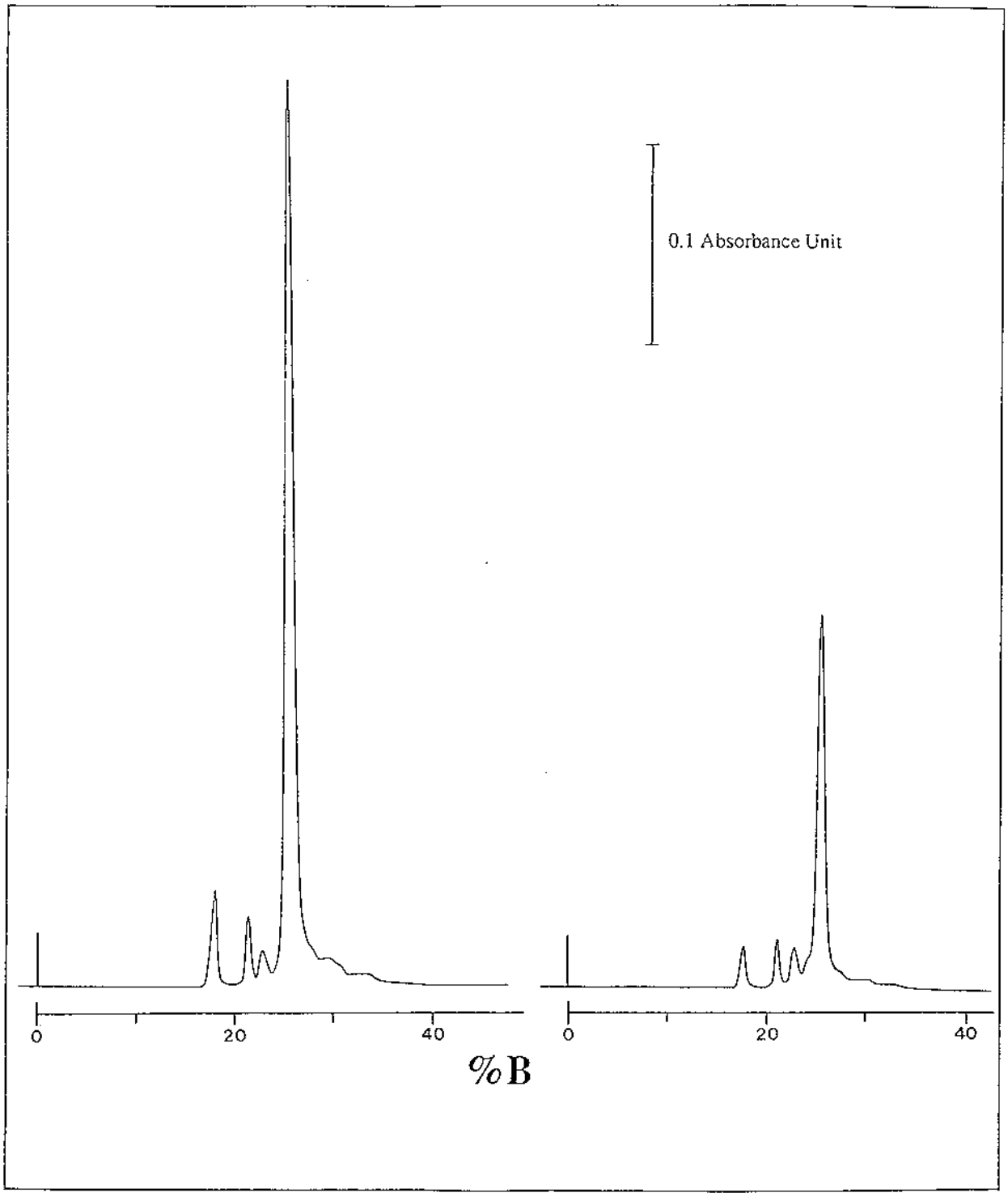

Figure 3.10 Absorbance profiles (280nm) obtained by FPLC on a Mono Q anion exchanger of transferrin-containing fractions from a Q-Sepharose column showing the effect of Endo-F/PNGase treatment.

The left trace shows fraction 34 without Endo-F/PNGase treatment, and the right trace shows fraction 34 with Endo-F/PNGase treatment.

analysis (Fig 3.9) showed no significant change in movement between the treated and untreated transferrins. In fraction 34 a slight difference in the relative quantity of each band can be seen. In the untreated sample the faster moving band appears to be in greater quantity than the slower moving band but in the treated sample the inverse is true. However, in the FPLC profile of fraction 34 before and after treatment with Endo-F/PNGase (Fig 3.10) no significant difference in peak position can be seen.

Apart from the number of glycan moieties on the protein the type of sialic acid residues present on the carbohydrate chains have also been reported to be a cause of transferrin microheterogeneity (Coddeville et al., 1989). Attempts to remove the entire glycan moieties appeared to be unsuccessful so removal of the sialic acids using the glycosidic enzyme neuraminidase was attempted instead. 2mg of freeze-dried material from each of the Q-Sepharose fractions 31, 34, and 38, was taken up in 40µl of 0.1M sodium acetate with 5mM EDTA. Half of the reaction mixture was treated with 1µl of neuraminidase (specific activity 25 U/mg of total protein) and both halves were left at room temperature for a day. FPLC traces showing the effect of neuraminidase treatment are shown in Fig 3.11. The profiles do not show significant peak shifting as would be expected with a change in the overall charge of the protein following removal of sialic acid residues. Instead the profiles show formation of many incompletely separated peaks indicating that protein degradation may have occurred.

3.2 Purification of the transferrin variant *

Obtaining a pure sample of the * variant of transferrin initially presented a problem. Because of the recent origin of the * allele, blood homozygous for this allele does not exist. It was therefore only possible to use heterozygous blood as a source of the * transferrin. The transferrin from a D* heterozygote had previously been isolated free from other contaminating proteins but it proved difficult to separate the D and * isoforms. It was known from previous work at the Equine Blood Typing and Research Centre that D and * ran very close to each other on PAGE. In the early stage of the present study, attempts were made to separate the D and * isoforms by FPLC on a Mono-Q anion exchange column but these proved unsuccessful.

Efforts were subsequently focussed on blood from a different heterozygote containing the O and * alleles. It was known that the O isoform was well separated from the D isoform by PAGE. It therefore seemed possible that the * variant could be obtained free from contamination by the O isoform.

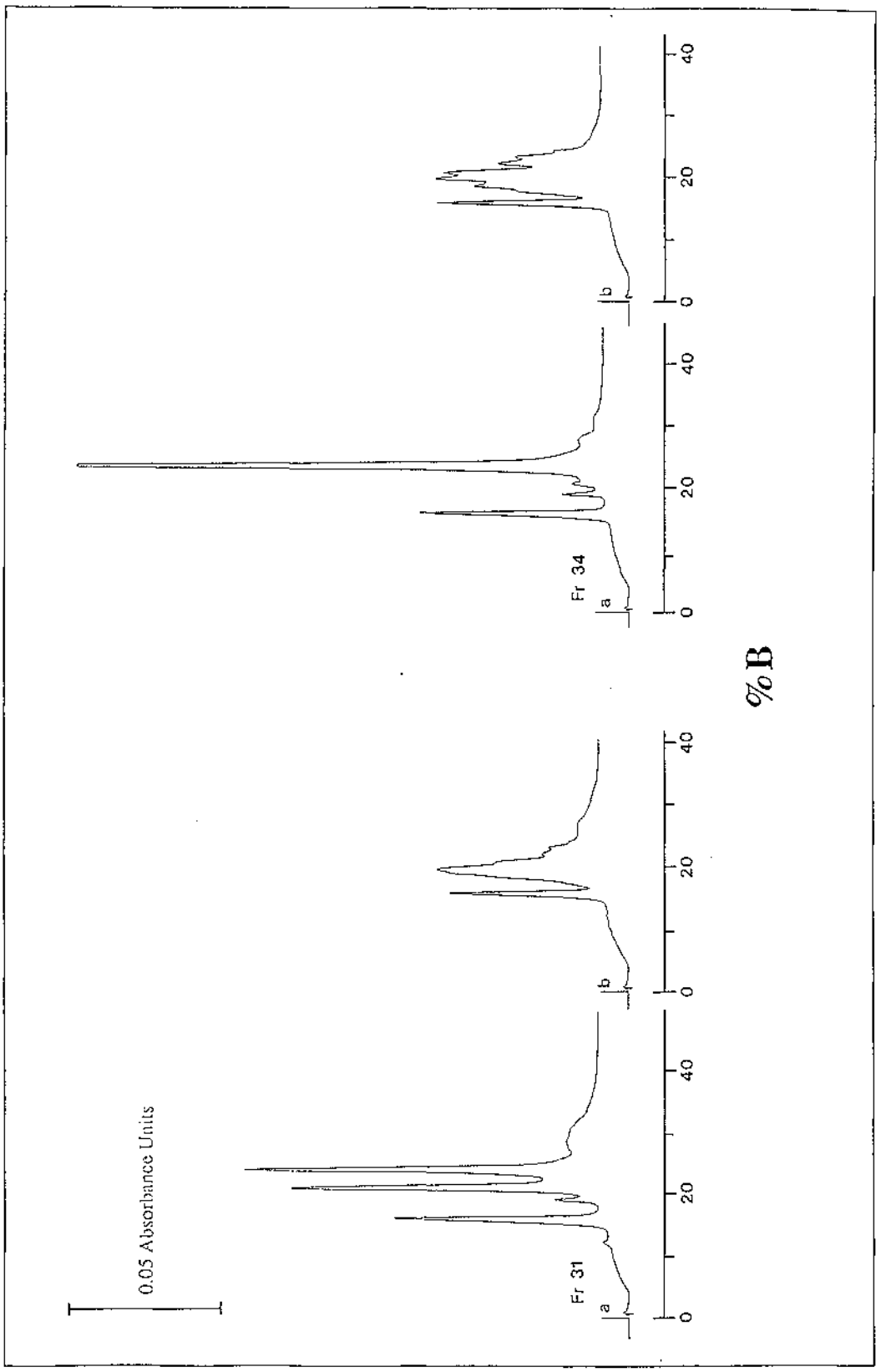

Figure 3.11 Absorbance profiles (280nm) obtained by FPLC on a Mono Q anion exchanger of transferrin-containing fractions from a Q-Sepharose column showing the effect of neuraminidase digestion..

The fractions are shown in pairs: (a) untreated sample
(b) neuraminidase treated sample

This proved to be the case. Blood from the O* heterozygote was subjected to the same purification procedure as that described for transferrin F1 in sections 3.1.1 to 3.1.3. i.e. using ammonium sulphate fractionation and successive passage through DEAE cellulose and Q-Sepharose columns.

3.2.1 Chromatography on a DEAE ion exchange column

During passage through the DEAE cellulose column two distinctive red bands were seen to elute off the DEAE cellulose column instead of the usual single band. Spectrophotometric analysis indicated that both were transferrin bands. Fractions containing these two transferrins were pooled separately and subjected to PAGE analysis. The fractions from the earlier eluting transferrins revealed two major protein components running more cathodal than the DH2 marker transferrin and with mobilities characteristic of the O variant (Fig 3.12 lane 6). Two distinct minor variant forms (a and b) of the R transferrin group (containing the O and R isoforms) exist with an a to b ratio of about 2 in horse serum (Watanabe et al., 1989). Chung and McKenzie (1984) had previously shown that the minor variant forms of the transferrin isoform R appeared with similar intensity on PAGE. Therefore, it is probable that two bands with similar intensities would also be seen for the O isoform. Pooled fractions from the second transferrin band indicated a major protein running very close to the DH2 marker, presumably the * variant (see Fig 3.12 lane 3). The earlier elution of the O variant is consistent with its lower negative charge which would result in weaker binding to the anion exchanger.

Data were not recorded for the fractions from the DEAE cellulose column used in the major purification reported above. The elution of the transferrin was simply observed visually. However, to illustrate the separation of the O and * transferrin isoforms on a DEAE ion exchanger a further small scale separation of O* plasma samples was carried out on a DEAE Sepharose column (with dimensions 1.6cm x 11.5cm). This column provided an elution profile that was very similar to that obtained with the preparative DEAE cellulose column.

10ml of plasma from the O* heterozygote was subjected to ammonium sulphate fractionation. The DEAE Sepharose column was equilibrated with 0.05M Tris/HCl buffer (pH 8.0) containing 0.03M NaCl until the pH and conductivity of the effluent was the same as that of the running buffer. Elution was accomplished with a linear gradient of 0.03-0.33M NaCl (0-60% B) in 0.05M Tris/HCl buffer (pH 8.0) in 120 min using a flow rate of 1 ml min^{-1} (B=0.5M NaCl). 4ml fractions were collected.

The relative transferrin content of the fractions was determined by reading the absorbance at 460nm. Transferrin was eluted between 12-28% B (Figure 3.13). The A_{460} values are shown only up to 40% B in Fig 3.13. The A_{460} values increase considerably in the later fractions but this increase is due to the high content of oxyhaemoglobin (wavelength maximum 414nm). This is evident in Fig 3.14 which shows a native PAGE of samples across the elution profile. Samples from fractions 16 and above show numerous protein components.

Two protein peaks can be seen to elute off the DEAE Sepharose before the major protein peak in the FPLC profile (Fig 3.13). Both of these peaks contain protein that absorbs at 460nm. Fig 3.14 shows that the second peak (fractions 11-14) consisted of a major protein running very close to the H2 marker, presumably the * variant. The fractions from the earlier eluting peak revealed a major protein component running more cathodal than the H2 marker transferrin i.e. presumably the O variant.

The separation of the two variants O and * by the DEAE anion exchanger can be clearly seen between fractions 10 and 11 on the native PAGE (Fig 3.14).

3.2.2 Chromatography on a Q-sepharose ion exchange column

The pooled fractions from the later-eluting red band from the DEAE ion exchange column which were presumed to contain the * variant, were further purified by passage through Q-Sepharose (as described in section 3.1.3). Since the pooled transferrin-containing fractions from the Q-Sepharose still contained some contaminating proteins (Fig 3.12, lane 4), pooled fractions were put through a Sephacryl S-300 column in an attempt to remove these contaminants. A Sephacryl S-300 gel filtration column (750mm x 25mm) was equilibrated with 0.05M Tris/HCl buffer (pH 8.0) containing 0.03M NaCl. 2-3 ml of highly concentrated protein solution was loaded onto the column and the protein was eluted at a flow rate of 0.3 ml min^{-1}. PAGE analysis of the transferrin-containing peak from the gel filtration column (Fig 3.12) appears to show that some contaminants have been removed.

A comparison of the PAGE mobilities of the purified * variant and that of the purified forms of the other transferrin variants is shown in Fig 3.15.

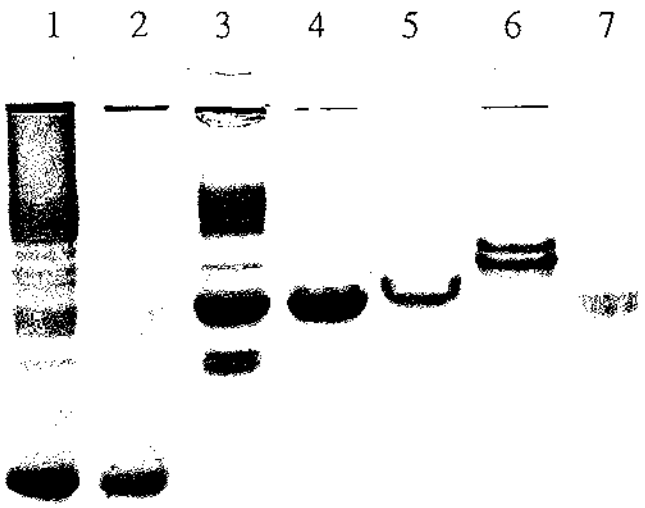

Figure 3.12 Photo of 10% native-PAGE showing the protein at each step from the purification of transferrin *.

Lane 1 plasma sample
Lane 2 45% ammonium sulphate supernatant
Lane 3 DEAE column pooled fractions of later eluting red band
Lane 4 Q-Sepharose column pooled fractions
Lane 5 Sephacryl S-300 pooled fractions
Lane 6 pooled fractions of early eluting red band from DEAE column
Lane 7 marker transferrin H2

Lanes 3, 4 and 5 represent successive stages in the purification of the later eluting red bands eluted from the DEAE cellulose column.

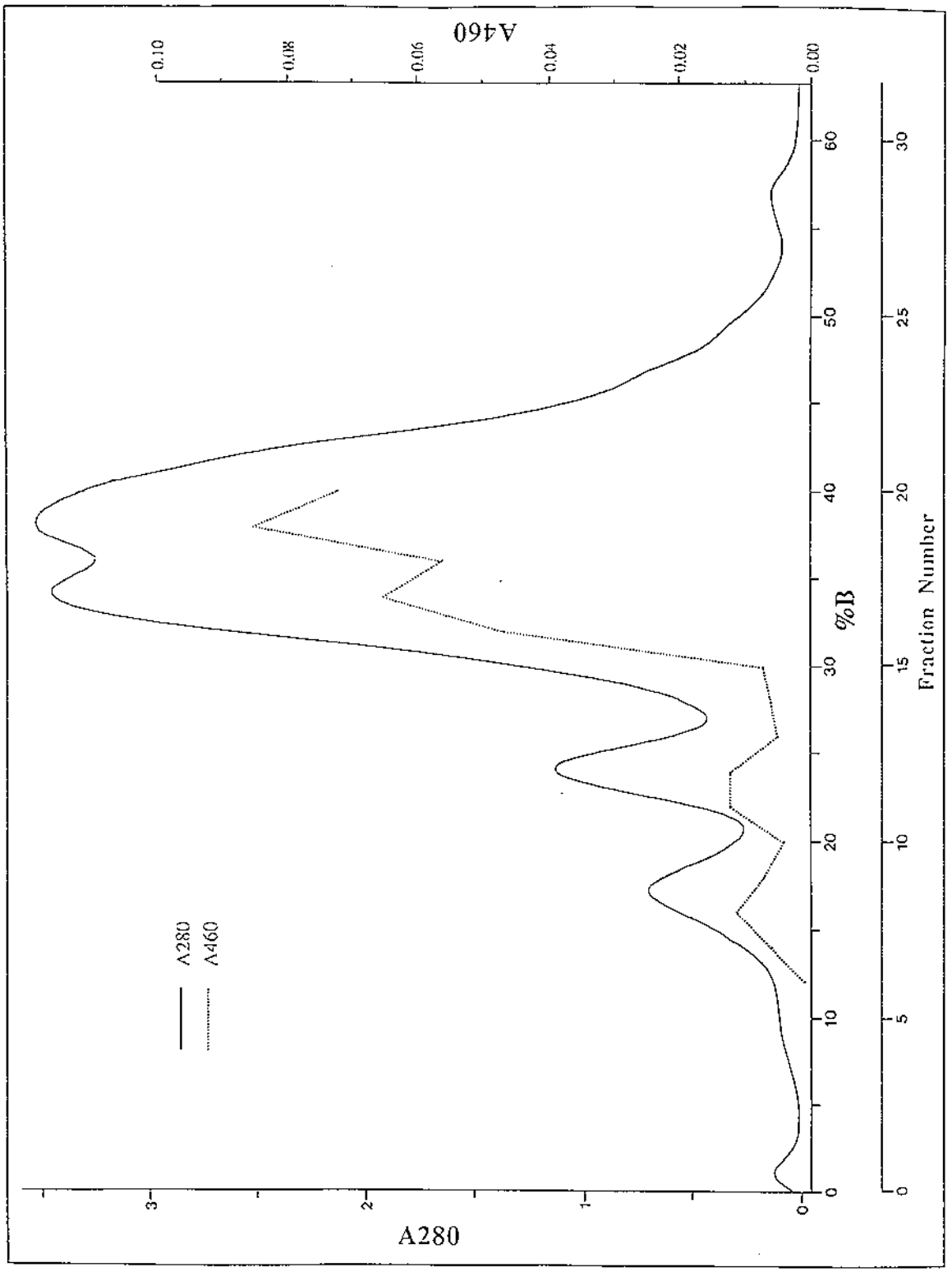

Figure 3.13 Elution profile of transferrin O* from a DEAE anion exchanger column.

Elution was accomplished with a gradient of 0-60% B in 120 min using a flow rate of 1 ml min^{-1}. 4ml fractions were collected.

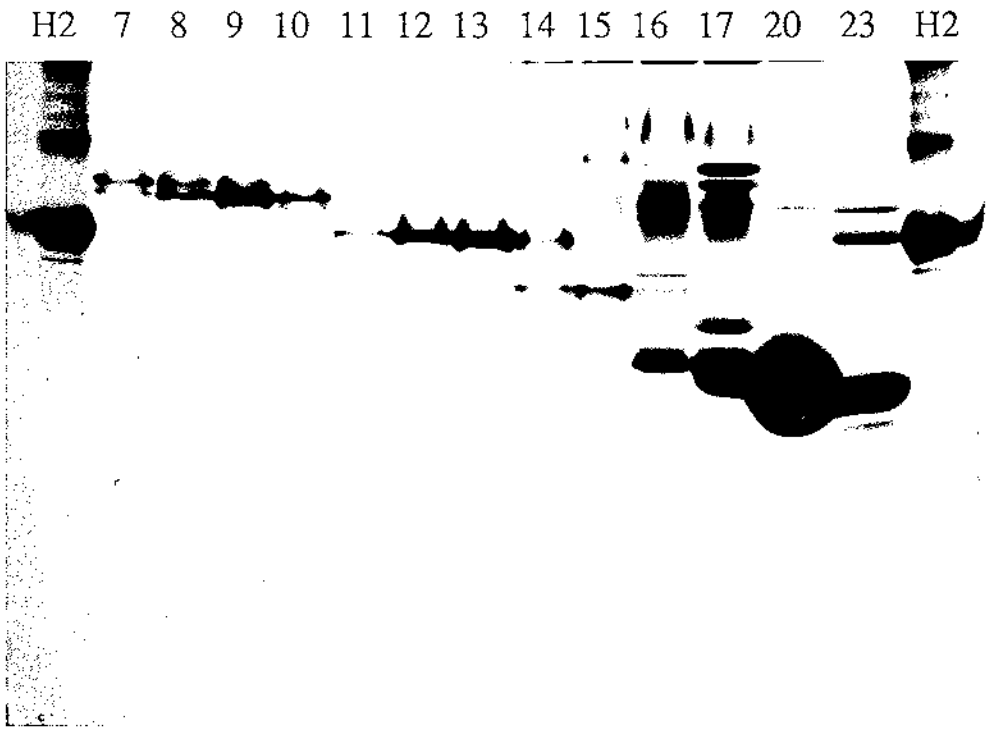

Figure 3.14 Native-PAGE (10% gel) showing the protein fractions from a DEAE sepharose column for blood from a horse heterozygous for the O and * transferrin alleles.

The DEAE sepharose column was eluted at 1ml/min and 4ml fractions were collected. An equal volume (40µl) from each fraction was loaded onto the gel. H2 is a transferrin marker. The numbers above the lanes refer to the fraction number.

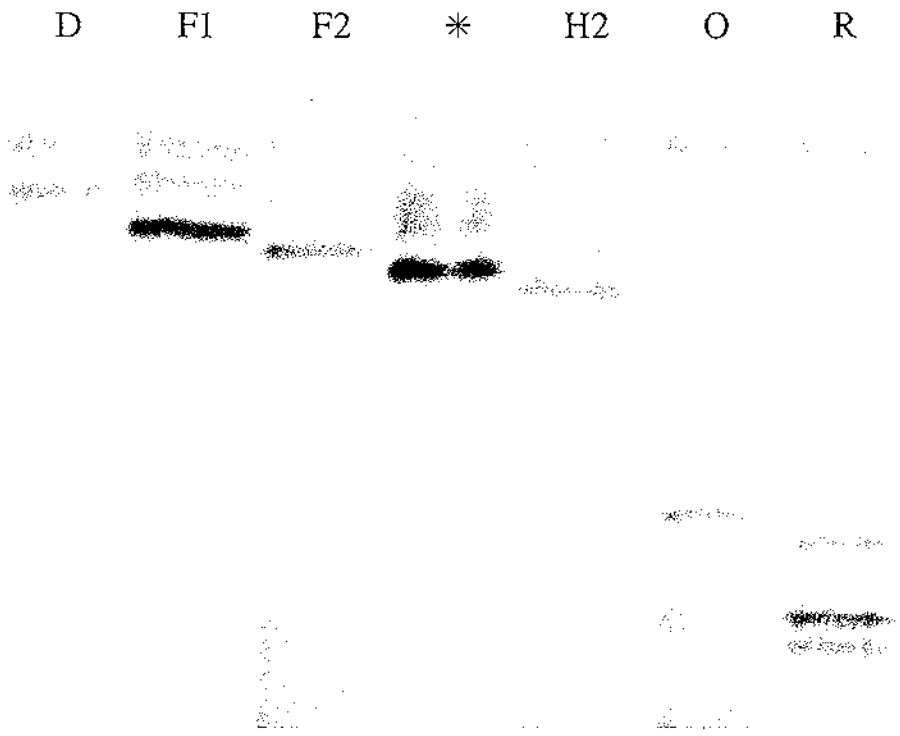

Figure 3.15 Native-PAGE (10% gel) showing the transferrin variants previously found in thoroughbred horses and the new variant *.

All the transferrin variant samples were taken from horse blood homozygous for each particular variant. The transferrin variant * was purified to the stage of chromatography through Q-sepharose.

This gel was run by the Equine Blood Typing and Research Centre, Massey University.

3.3 Primary structure comparisons between three transferrin variants: D, F1 and *.

In order to identify which of the two parental genotypes was the source of the * transferrin variant it was necessary to make amino acid sequence comparisons between the three variants. This was done by making tryptic digests of each of the three purified transferrins, separating the peptide products by HPLC and carrying out sequence determination of selected D, F1 and * peptides. Sequence determination was carried out in two ways.

(1) By automated Edman N-terminal sequencing. In view of the complexity of the HPLC profiles it was not practicable to sequence all of the tryptic peptides resolved by HPLC. By careful comparison of the HPLC profiles of the different transferrins, those peaks which showed marked differences in peak height between the variants were selected for sequencing.

(2) By mass spectrometry. Mass values for the peptides present in the HPLC peaks were determined by mass spectrometry. Using the amino acid sequence predicted from the D/F1 cDNA sequence (determined by M. Carpenter), the theoretical mass data were calculated for all the possible tryptic peptides using the Macintosh MacProMass software. Comparison of the experimentally determined mass values with these calculated values allowed the identification of many peptides in the digest.

3.3.1 Tryptic Digests

Samples (20mg) of the purified transferrin variants D, F1 and * were digested with trypsin as described in Section 2.2.7. After freeze-drying, small samples (approx. 0.08mg) were dissolved in 0.1% TFA (HPLC buffer A) and injected into a C18 reverse phase HPLC column to separate the tryptic peptides. The HPLC parameters and elution conditions are specified in the legend to Fig 3.16. The HPLC elution profiles of the digests for each of the three transferrin variants are shown in Figs 3.16 - 3.18 with the numbers used to refer to each peak in subsequent sections indicated.

Visual comparison of the HPLC profiles of the transferrin D, F1 and * tryptic digests revealed several peaks which showed significant differences between the three profiles. These are shown by shading in the individual profiles in Figs 3.16 - 3.18. For ease of

comparison the three figures have been aligned in Fig 3.19 and the position of the main peak differences indicated by vertical dashed lines.

Table 3.1 lists those peaks showing major differences and indicates whether the particular peak is present, absent or present at an intermediate level in each of the three variants. It was expected that sequence analysis of these particular peaks would reveal differences in sequence between the variant forms.

Table 3.1 Table shows the presence or absence of selected peaks in the HPLC profiles of tryptic digests of transferrins from the three variants F1, D, and *.

The state of being of each peak was determined visually.
A + means that the peak is present in the variant.
A – means that the peak is not present in the variant.
A ± means that the peak is present in the variant but at a relatively low level

Peak Number	Variant		
	F1	D	*
5a	±	–	+
9	+	–	+
11	–	+	–
23a	+	–	–
27	±	+	+
28a	+	–	–
38	–	+	+
51	±	+	±

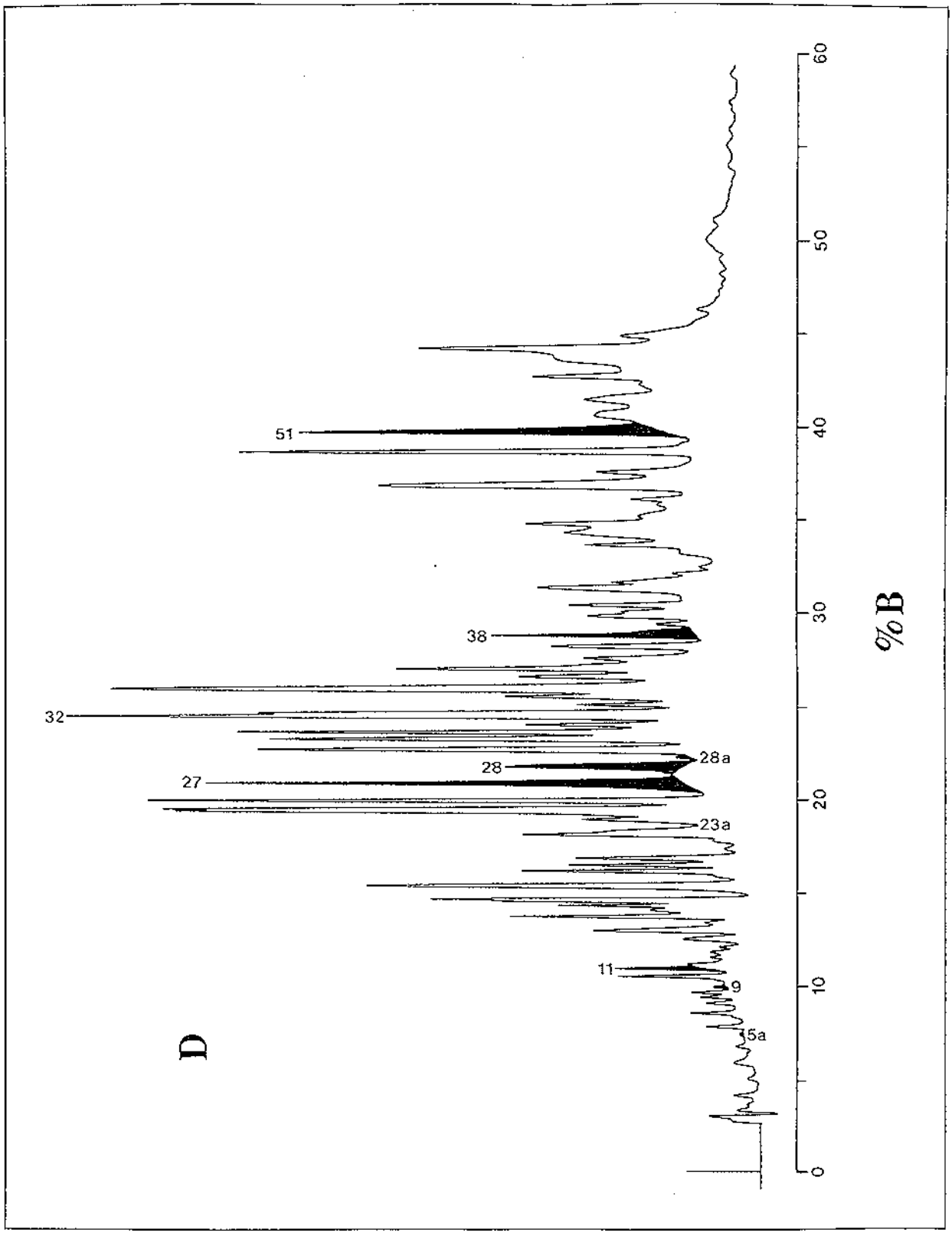

Figure 3.16 HPLC profile of the D tryptic digest.

0.08mg of the digested protein was loaded onto the C18 column to provide this analytical profile. Elution was accomplished with a linear gradient of 0-60% B in 60 min using a flow rate of 1 ml min^{-1}. The spectrophotometric detector was set to a wavelength of 220nm and at an AUFS of 0.2 absorbance units.

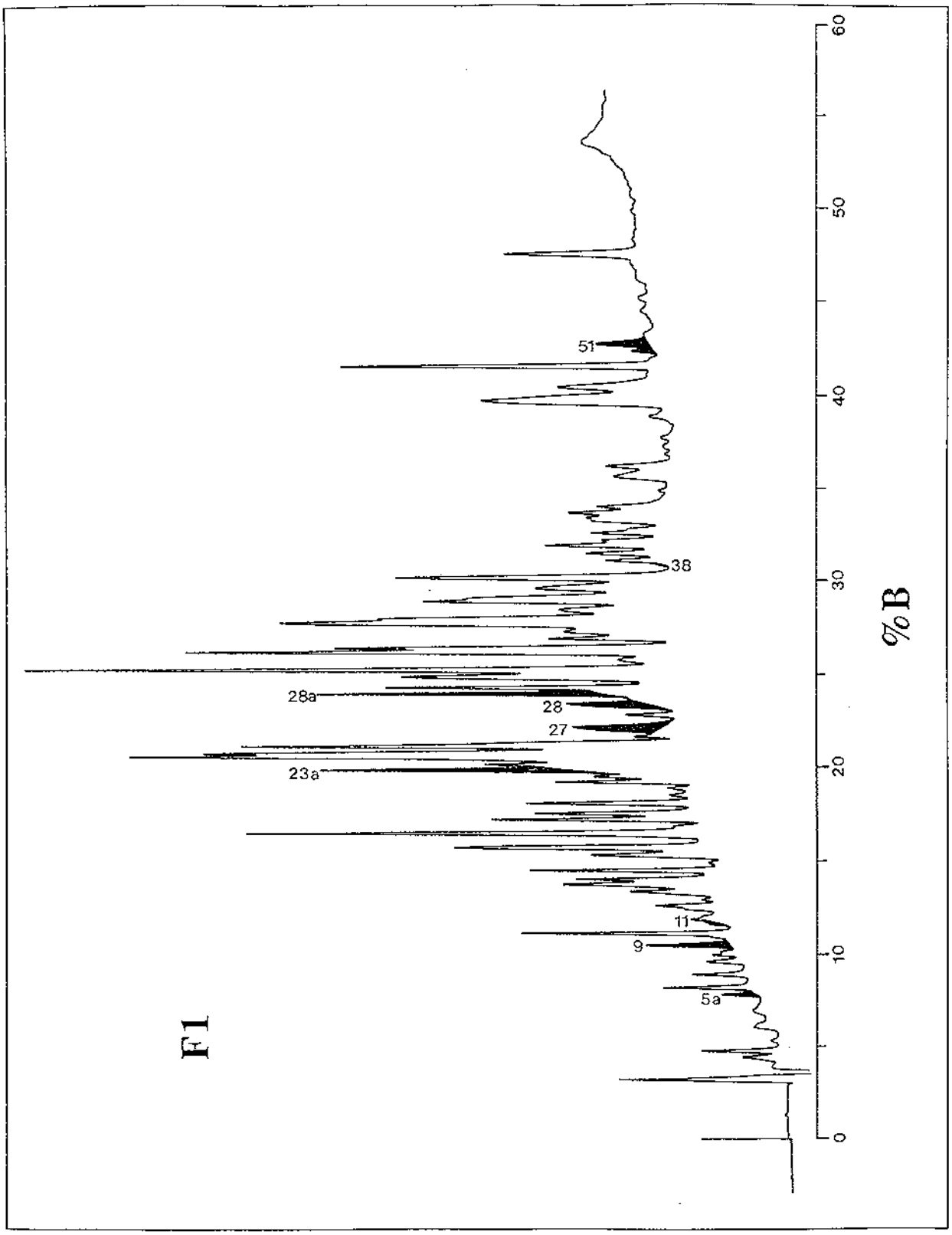

Figure 3.17 HPLC profile of the F1 tryptic digest.

Conditions identical to those described for the D variant tryptic digest (see legend for Fig 3.15).

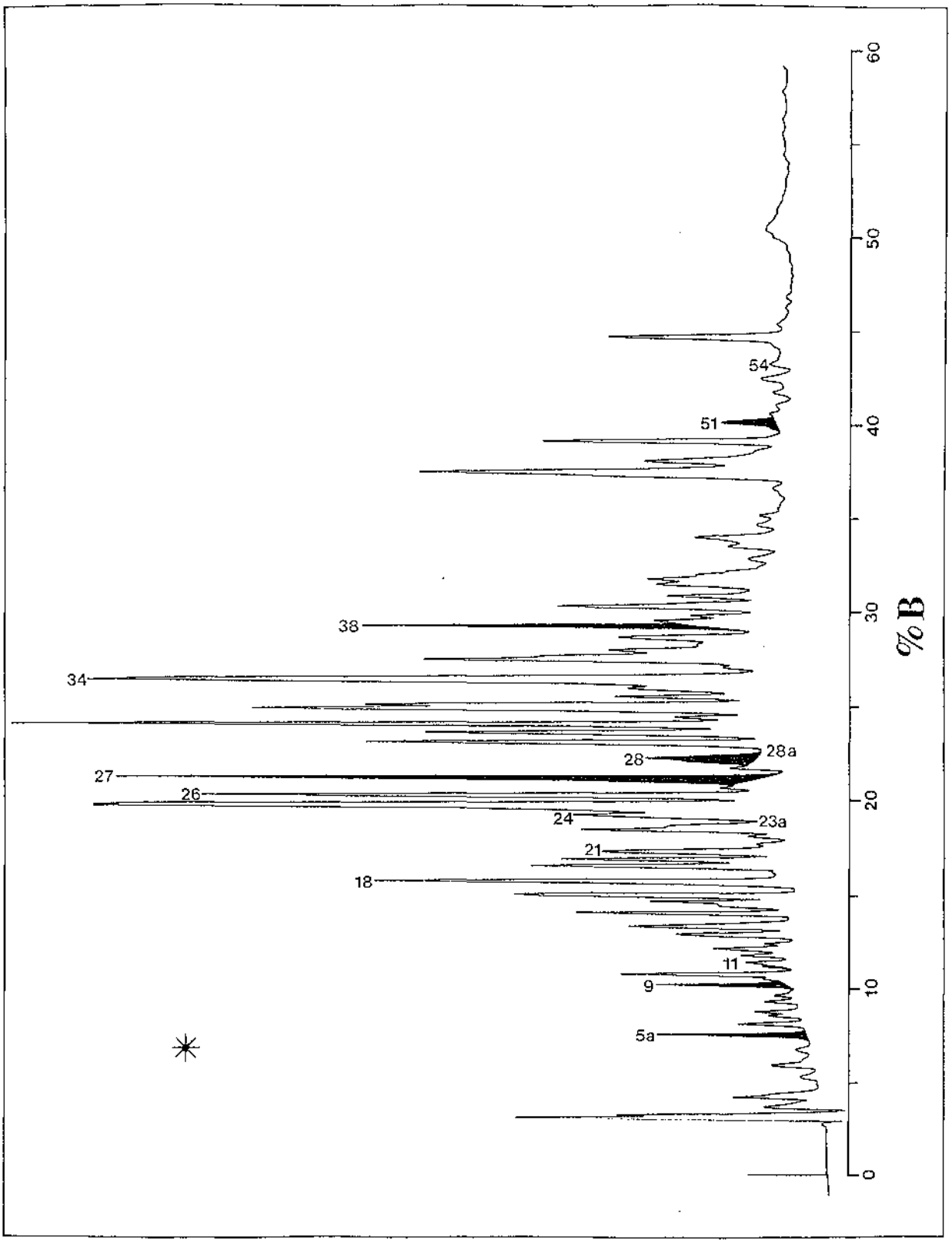

Figure 3.18 HPLC profile of the * tryptic digest.

Conditions identical to those described for the D variant tryptic digest (see legend for Fig 3.15).

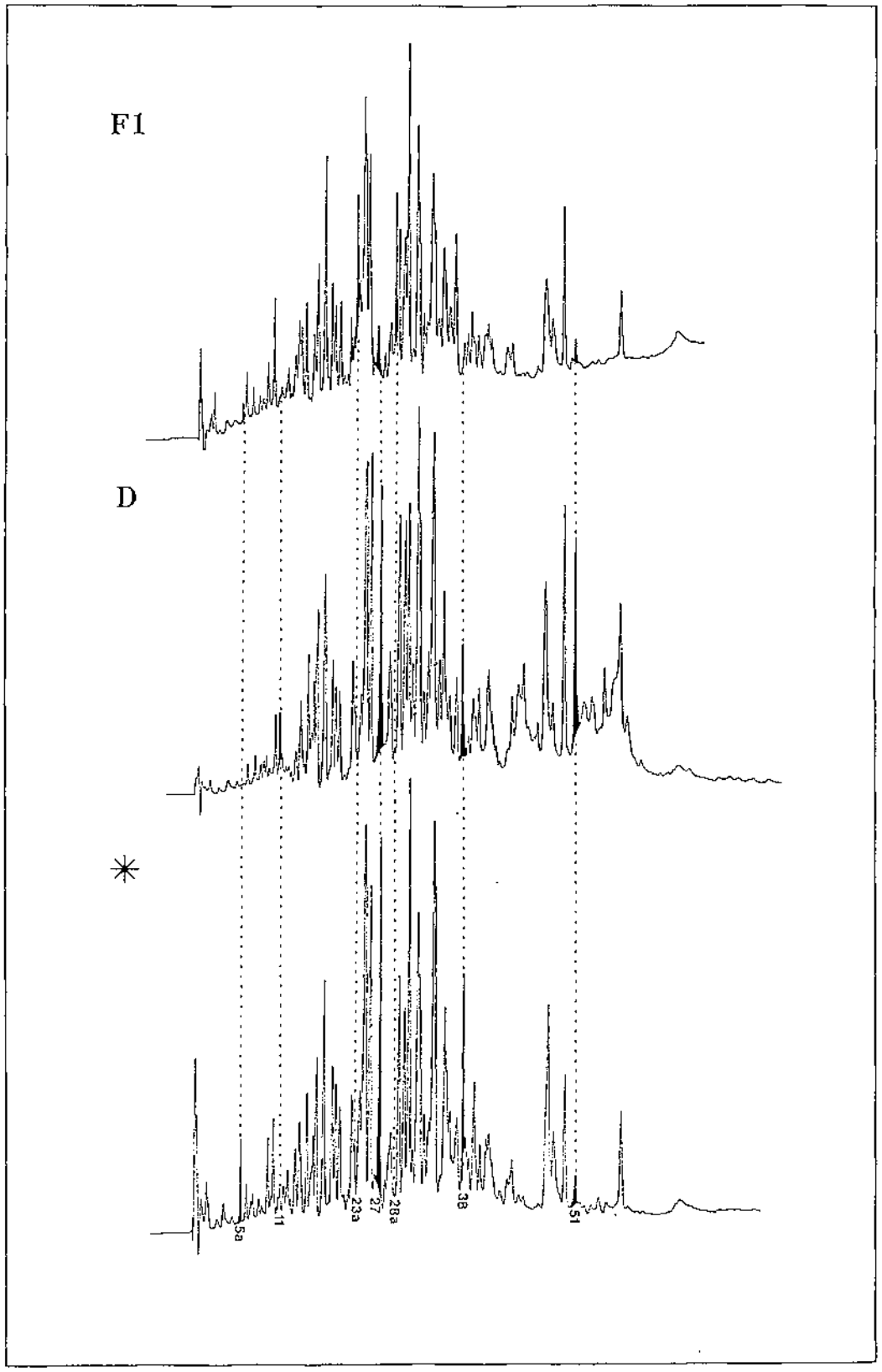

Figure 3.19 HPLC profiles of the D, F1, and * tryptic digests.

The HPLC profiles of the tryptic digests have been aligned vertically for ease of comparison. Important peaks have been shaded and their comparative positions in the profiles of the other digests are shown with dashed lines. Peak numbers are indicated below the * profile.

3.3.2 Peptide identification by N-terminal sequencing.

(1) Peptides from the transferrin variant D:

Comparison of the HPLC profiles from the D and F variants revealed the presence of four peaks which are prominent in the D transferrin profile but absent or present as only small peaks in the F1 transferrin profile. These peaks (11, 27, 38, and 51) are shown shaded in the HPLC profile of the D variant transferrin tryptic peptides (Fig 3.16) and the corresponding positions in the HPLC profile of the F1 variant are marked (Fig 3.17). Two other peaks (28 and 32) for which sequence information was obtained at an early stage of this study are also shown. The sequences found for the peptides in these peaks are shown in Table 3.2a.

In four of these peaks there were three different peptides present and the assignment of particular residues in each sequencing cycle to a particular peptide was only possible after alignment of the possible sequences with that deduced from the cDNA sequence. This alignment is presented in section 3.3.3. Certain residues which could not be clearly identified from the sequence data but which are identifiable by alignment of the rest of the sequence with that predicted from the cDNA are shown in parenthesis. Also some peptides have clearly not been sequenced to the C-terminus since the C-terminal residue should be either a lysine or an arginine residue. In these cases the C-terminal sequence to the nearest K or R in the predicted sequence is shown in parenthesis.

(2) Peptides from the transferrin variant F1:

Comparison of the HPLC profile from transferrin variant F1 with that of variant D reveals three peaks which are prominent in the F1 variant profile but not in D. These are peaks 9, 23a and 28a. The sequences of peptides in peaks 23a and 28a are shown in Table 3.2b. The sequence for peptide 9 was obtained from the corresponding peak in the * variant (see below).

(3) Peptides from the transferrin variant *:

The peptides from several peaks from the HPLC profile of the transferrin variant * which were prominent in either the D variant profile or in the F1 variant profile (but not prominent in both) were also sequenced. These are presented in Table 3.2c. Peaks 10, 18, 21, 24 and 27, from the * variant were also sequenced. These peptides were sequenced as no mass spectrometric data were obtained for these peaks and they fell

within a region of interest within which all peaks were being analysed (see section 3.3.4).

Table 3.2a Sequences found for the peptides in selected peaks from the D variant HPLC profile.

Peak	Sequence
11	IVK
	EQTVR
	(N)GSHCPDK
27	HQTVEQNTDGCNPDD(W)AK
	GDVAFVK
	?LCQLCVG(K)
28	(N)STLCNLCIGSASGPG(R)
	TEPQ(T)HYYAVA(VVK)
	?CIGSAS(GPGR)
32	DFHLF(SSPHGK)
38	CGLVPVLAENYETRPGSACVDTPEEGYHAV(AVVK)
	SLDGGFIYIAG(K)
	(T)AVPNL
51	SAG(W)NIPIGLLY(WQLPEPR)

Table 3.2b Sequences found for the peptides in selected peaks from the F1 variant HPLC profile.

Peak	Sequence
23a	NPDDWAK
	LLCPDGTR
	EDIRPEVPK
	GDVAFV(K)
28a	SGSACVDTPEEGYHAVAVV(K)

Table 3.2c Sequences found for the peptides in selected peaks from the * variant HPLC profile.

Peak	Sequence
5a	VPK
9	DLK
10	DLK QCS(T)SR
11	IVK EQTVR (N)GSHCPDK
18	EGCAPGYR SCHTGLGR
21	(N)GSHCPDK TYLGEK
24	EDIRPVP(K) LLCPDGT(R) TEPQTHY(YAVAVVK)
26	WCAIGHHE(K)
27	GDVAFVK HQTVEQNTDGCNPDD(WAK) AACVCQELHNQQASY(GK) ?LCQLCVGK
28	(N)STLCNLCIGSASGPG(R) ?CLFQSAT(K)

continued next page ...

continued from previous page ...

29	DDTQCLANLQPTTTYK
32	CLADGAGDVAFVK
34	SSSDPDL(TW)N(SLK) CACSNHEPYFGYS(G)AF(K) AVSNFFAGSCVPCADR LLEACT(FHR)
38	CGLVPVLAENYETRPGSACVDTPEEGYHAV(AVVK)
51	SAGWNIPIGLLYWQLPEP(R) AIADNEADAVTLDAGLVF(E)AG(LSPYNLKPVVAEFYGSK)
54	AIADNEADAV(TLDAGLVFEAGLSPYNLKPVVAEFYGSK) SIVPAPPLVACV(K)

3.3.3 Alignment of peptide sequences with the sequence deduced from the cDNA sequence.

The sequences determined for the various peptides from the HPLC profiles of the D, F1 and * transferrin variants are shown aligned with the amino acid sequence deduced from the transferrin cDNA sequence obtained from clones from a D/F1 heterozygote in Fig 3.20. D peptides are represented by blue lines and * and F1 peptides by red and green lines respectively.

The interpretation of the relationship between the sequence information obtained, the HPLC profiles shown in Figs 3.16 to 3.19 and possible differences between the transferrin variants will be discussed in section 4.

3.3.4 Identification of peptides by Mass Spectrometry

The amino acid sequence of horse transferrin predicted from cDNA sequences obtained from the transferrin gene clones of a horse heterozygous for the transferrin variants D and F1 (M. Carpenter) was used to assign probable sequences to mass data obtained for tryptic peptides by mass spectrometry. Macintosh MacProMass software was employed to model a tryptic digestion of the predicted amino acid D/F1 transferrin sequence. This provided a complete list of all the tryptic peptides assuming cleavage at all trypsin-sensitive sites. The list is presented in Table 3.4 where the predicted tryptic peptides are listed in order of their position in the transferrin primary sequence (obtained from the cDNA information) together with the calculated mass values.

Mass spectra were determined for a number of the peptide peaks resolved in the HPLC profiles shown in Figs 3.16-3.18. Peaks were selected for analysis by mass spectrometry for three main reasons:

(a) to confirm the identities of those peptides characterised by N-terminal sequencing, especially in those cases where the sequence appeared to be incomplete or where certain residues were not clearly identified,

(b) to identify other tryptic peptides which were not characterized by N-terminal sequence analysis

Figure 3.20 Map showing position of sequenced peptides relative to the whole transferrin amino acid sequence.

The amino acid sequence (described in section 2.1.3) was predicted from a D/F1 cDNA sequence (M. Carpenter, 1992). The blue, red, and green lines beneath the amino acid sequence represent the regions that have been sequenced from the D, *, and F1 variants respectively. All the peptides sequenced may be seen in tabulated form in Tables 3.2a, 3.2b and 3.2c.

(c) to search for peptides in the tryptic digest of the * variant which do not have masses corresponding to any of the predicted tryptic peptide mass values and which might therefore indicate an amino acid sequence difference in the * transferrin sequence.

The mass spectrometry profiles were examined for prominent molecular ions (MH^+) and the mass of each MH^+ was compared to the list of predicted masses. If an MH^+ value was found to agree closely with a value in the predicted mass list then it was assumed that the HPLC peptide peak from which the mass spectrometry profile was obtained contained that particular peptide.

Prominent molecular ions were defined as those lines in the mass spectrum that showed up as significantly larger than the background (noise) values and which only appeared in the mass spectra of one or two of the profiles. Fragments which appeared in many or most of the profiles were probably due to the presence of some common contaminating constituents (see Fig 3.21).

MH^+ values and predicted mass values were assumed to be the same if the two values were within 1 mass unit of each other. A discrepancy of one mass unit arose when the mass values were rounded up or down. In some cases the MH^+ values corresponded to more than one predicted sequence within this range. All the possible sequences corresponding to a particular MH^+ value were therefore listed as possibilities. Some peptides have been listed that lie outside this range, reasons for this are discussed later in section 3.3.5.

The MH^+ values obtained in this way are presented in Table 3.4 which also indicates the particular peak in which the peptide was found. Tryptic peptide sequences and their corresponding calculated mass values are also shown.

Sequences predicted from the mass spectrometry data are shown aligned with the horse transferrin sequence in figure 3.22.

The relationship between the information on tryptic peptide sequences and possible differences between the transferrin variants will be discussed in section 4.

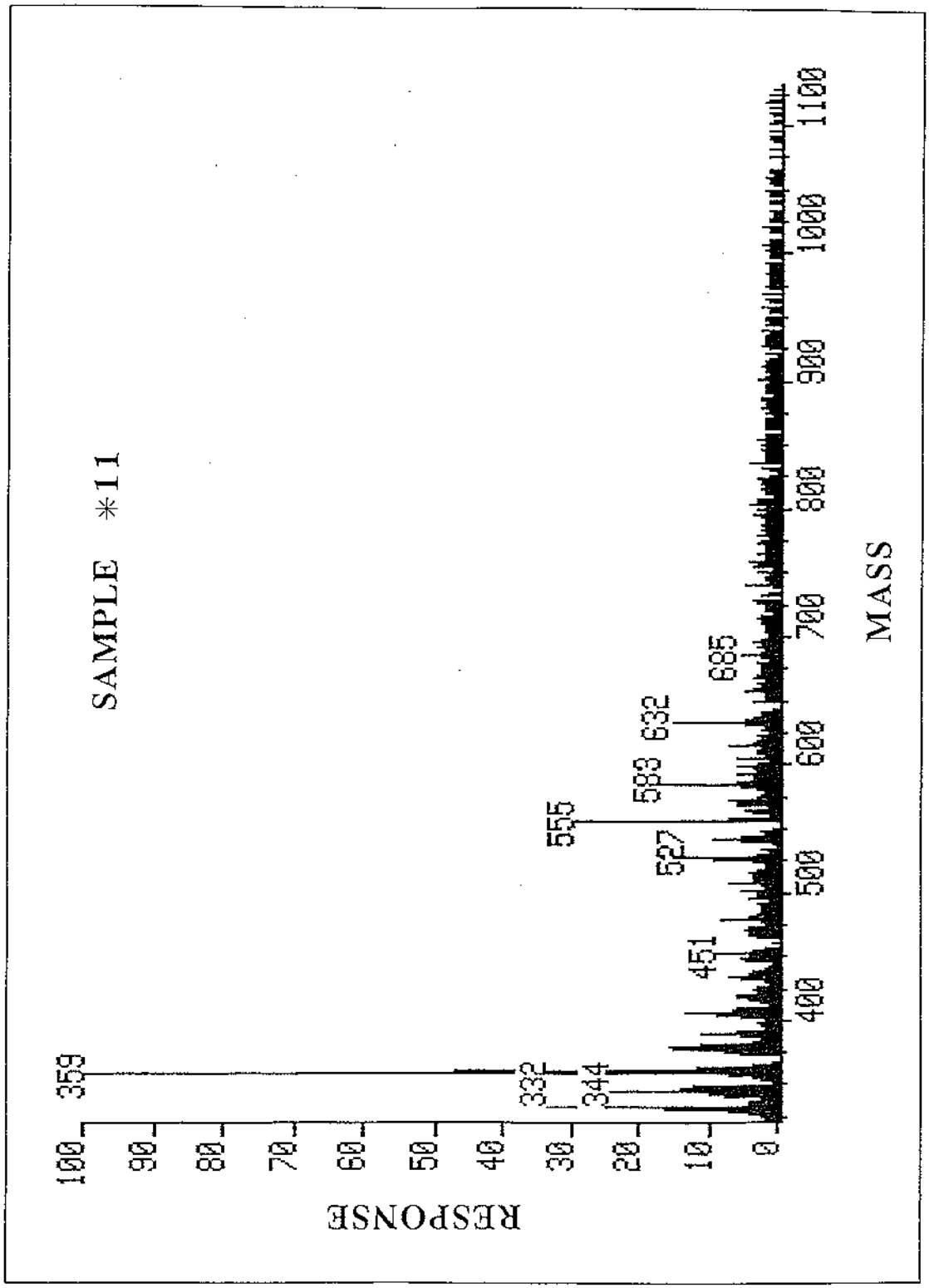

Figure 3.21 A typical mass spectra obtained for a peak from an HPLC profile of a tryptic digest.

The mass spectra shown was obtained for the peptides present in peak 11 of the HPLC profile of the * tryptic digest. The molecular ion with a mass of 359 corresponds to the tripeptide IVK (residues 383-385). The same peptide was also found in peak 11 of the HPLC profile of the * tryptic digest by peptide sequencing (see Table 3.2c). The molecular ions with masses 527, 555, and 583 were present in many mass spectra obtained for peptides of the * variant and as such were regarded as a contaminant.

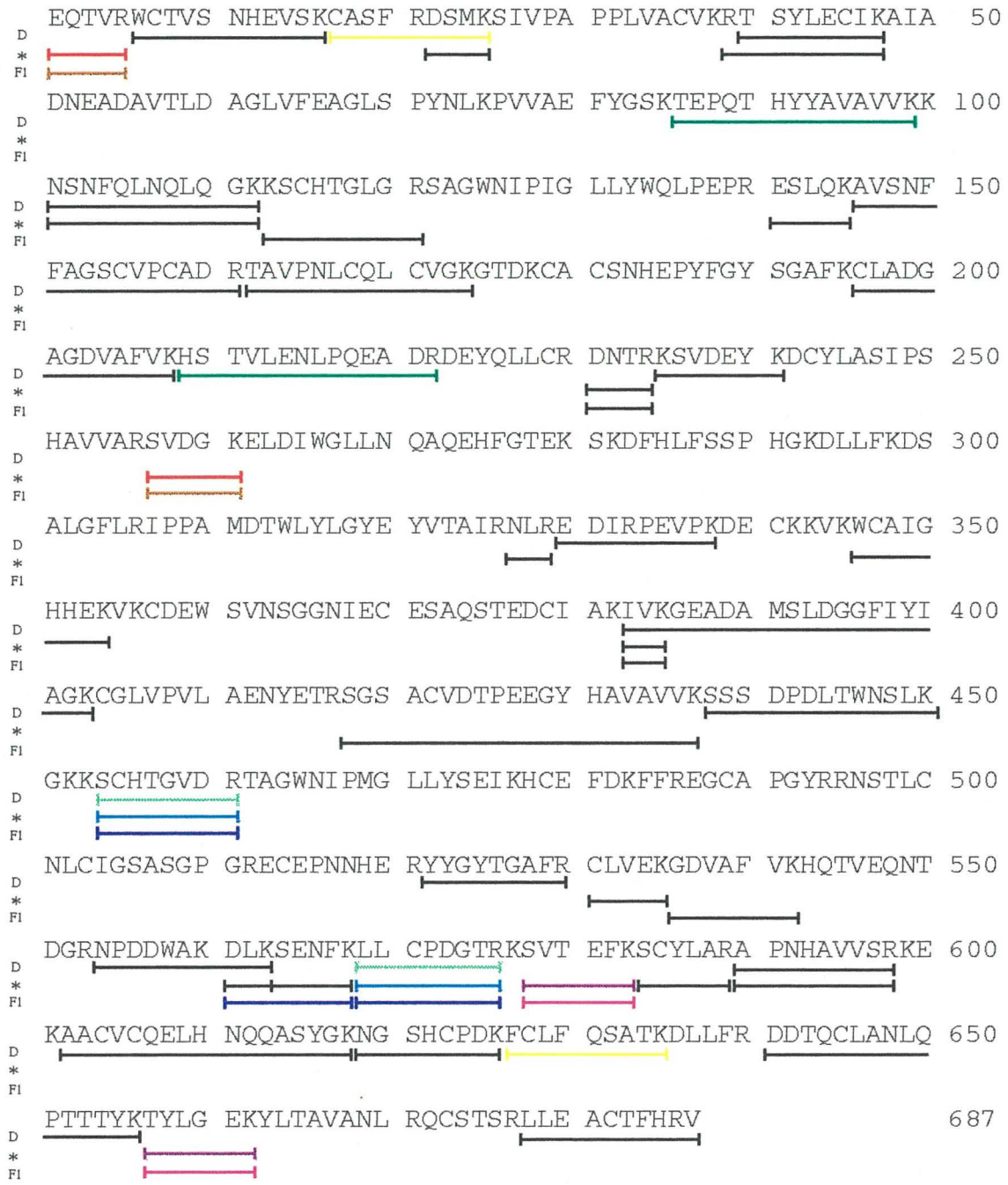

Figure 3.22 Map showing position of peptides identified by mass spectrometry relative to the whole transferrin amino acid sequence.

The amino acid sequence (described in section 2.1.3) was predicted from a D/F1 cDNA sequence (M. Carpenter, 1992). The lines drawn beneath the rows of amino acid sequence indicate the peptides that have been identified in the variants. The specific variant in which the peptide occurs is indicated by the position of the line below the amino acid sequence (as shown to the left of the map). Most molecular ions correspond to only one possible peptide, these peptides are indicated by a black line. Where two or more peptides have identical masses the lines are drawn with the same colour. All the peptides sequenced may be seen in tabulated form in table 3.4.

3.3.5 Validity of peptide identifications from molar mass values.

It is evident from Table 3.4 and Fig 3.21, that the data obtained from the use of mass spectrometry to determine the mass values of the molecular ion species found in the peptide peaks separated by HPLC, when combined with the mass values predicted from the cDNA sequence for tryptic peptides, have yielded considerable information on peptide identification. In many cases where there is a close agreement between the observed and predicted mass values and where only one possible predicted mass value corresponds to a value actually found, the identifications are probably unambiguous and accurate. In a number of cases these identifications have been confirmed by N-terminal sequence information. In several of these cases the mass value has served to clarify uncertainty concerning the C-terminal end of the peptide sequence.

For example the sequence WCAIGHHE(K) (of theoretical mass 1186), was found in peak 26 of the * HPLC profile. Upon mass spectrometry peak * 26 was found to contain a molecular ion of mass 1186. Another sequence, GDVAFV(K) (of theoretical mass 735), was found in peak 27 of the F1 HPLC profile and mass spectrometry of peak F1 27 revealed that a molecular ion of mass 736 was present. In neither case was the C-terminal lysine residue detected by N-terminal sequencing but mass determination confirms that it is, as expected, present in both these peptides.

However, there are also several instances in the list of peptides present in Table 3.4 where identification is ambiguous or open to question, even in cases where there is quite close agreement between predicted and actual mass values. This is particularly the case where the predicted mass values of several different tryptic peptides are closely similar to the mass value actually found. Some examples of this are discussed below to illustrate some of the difficulties and limitations of reliance on mass spectrometry alone for peptide identification.

(1) Peptides present in peak D 25 (see Fig 3.16).

The material in peak D25 was not subjected to N-terminal sequence analysis since this peak did not appear to differ greatly in size between the three transferrin variants. Consequently the only information on peptide identification was that obtained from mass spectrometry. Mass spectrometry of the peptides present in this peak revealed several molecular ions values which could be equalled with the calculated mass values of tryptic peptides.

Table 3.3 Molecular ions and corresponding sequences found in peak 25 of the D HPLC profile. Both the molecular ions found and the predicted mass values are monoisotopic values in the form MH^+.

Mass			Peptide	Residues	Sequence
Found		Predicted			
979		979	49	454-461	SCHTGVDR
	or	980	64	569-576	LLCPDGTR
1083		1083	35	330-338	EDIRPEVPK
1396		1395	2	6-16	WCTVSNHEVSK
1557		1556	17	162-174	TAVPNLCQLCVGK
1855		1854	16	146-161	AVSNFFAGSCVPCADR
2061		2060	71	602-618	AACVCQELHNQQASYGK

(a) Two possible tryptic peptide sequences may correspond to the mass value of 979 found in the mass spectrum of this peak. These are LLCPDGTR (mass 980) and SCHTGVDR (mass 979). The former peptide was identified by N-terminal sequence analysis of the peptides present in the adjacent peak 24 of the * variant HPLC profile (see Fig 3.18). So this peptide could occur in * peak 25 by overlap of peak 24. Indeed a peptide with mass value 980 was found in * peak 25. So it is probable that LLCPDGTR is the correct identification of the peptide in peak D 25.

(b) Only a single peptide, EDIRPEVPK, corresponds to the mass value 1083. This peptide has been identified in * 23a by sequencing (see Table 3.2c) so its detection in peak D 25 could well be due to peak overlapping.

(c) Peak D 25 contains a molecular ion of 1855 mass units, which corresponds to the mass of the peptide AVSNFFAGSCVPCADR. However the same peptide was found in peak * 34 by N-terminal sequencing and so it does not seem likely that this peptide would actually be present in peak D 25. The identity of the peptide giving the molecular ion of 1855 mass units is therefore not known.

(d) A peptide of 2061 mass units was found in peak D 25. The corresponding peptide AACVCQELHNQQASYGK (predicted mass 2060) was sequenced in peak 27 of the * HPLC profile (see Fig 3.18 and Table 3.2c) and its detection in D 25 could be due to peak overlapping.

(2) Peptide present in peak D 27 (see Fig 3.16).

Peak D 27 contains a molecular ion of 867 mass units and peak D 26 contains a molecular ion of 869 mass units. Both of these molecular ions correspond to the peptide KSVDEYK which has a calculated mass of 868 mass units. It is possible that some of this peptide eluted in peak 26 was also collected in peak 27.

(3) Peptides present in peak * 22 (see Fig 3.18).

Peak * 22 contains a molecular ion of 710 mass units that could correspond to either one of the two peptides SVTEFK (mass 710) or TYLGEK (mass 710). Sequencing information shows that the TYLGEK peptide is present in * 21. Overlapping of peaks 21 and 22 could explain the presence of TYLGEK in this peak .

(4) Peptides present in peak F1 23a (see Fig 3.17).

Peak F1 23a contains a molecular ion with a mass of 980 mass units, for which there are three sequence options, SCHTGDVR (mass 979) or LLCPDGTR (mass 980) or DLKSENFK (mass 981). Comparison of sequence information for F1 23a (see Table 3.2b) and * 24 (see Table 3.2c) would indicate that LLCPDGTR is most likely to be the peptide present in this peak.

(5) Peptides present in peak F1 27 (see Fig 3.17).

The molecular ion of 736 mass units found in F1 27 corresponds to the peptide GDVAFVK (mass 735) confirming the sequence information obtained for peaks F1 27, D 27 and * 27 (see Tables 3.2a, 3.2b and 3.2c); i.e. all three peaks contain the peptide GDVAFVK.

(6) Peptides present in peak F1 28a (see Fig 3.17).

A molecular ion of 2128 mass units was found in peak F1 28a. The peptide found on the list of predicted tryptic peptides with the mass closest to this value is SGSACVDTPEEGYHAVAVVK (mass 2124). While the agreement between the actual and predicted mass values would not, by itself, be sufficient to indicate identity, in this case independent N-terminal sequence information confirms that the sequence S_{418}-K_{437} is correct and that this peptide is present in peak F1 28a.

3.3.5 Peptide sequences deduced from mass measurements

Table 3.4 List of expected peptides from a tryptic digest and their calculated mass values. Actual mass values and HPLC peaks in which peptides corresponding to these values were found are listed to the right of the corresponding calculated mass values.

C=Pyridylethyl Cysteine

Residues	Expected Tryptic Peptide	Calculated Mass	Molecular Ion Mass	Peak
1-5	EQTVR	632	633	F1 11
			632	* 11
6-16	WCTVSNHEVSK	1395	1396	D 25
17-21	CASFR	688		
22-25	DSMK	480		
26-38	SIVPAPPLVACVK	1399		
39-39	R	175		
40-47	TSYLECIK	1062	1061	D 31
48-85	AIADNEADAVTLDAGLVFEA–GLSPYNLKPVVAEFYGSK	3955		
86-99	TEPQTHYYAVAVVK	1606	1607	D 30
100-100	K	147		
101-112	NSNFQLNQLQGK	1391	1392	D 30
			1391	* 30
113-113	K	147		
114-121	SCHTGLGR	935	935	D 18
122-140	SAGWNIPIGLLYWQLPEPR	2210		
141-145	ESLQK	604	604	* 14
146-161	AVSNFFAGSCVPCADR	1854	1855	D 25
162-174	TAVPNLCQLCVGK	1556	1557	D 25
175-178	GTDK	420		
179-195	CACSNHEPYFGYSGAFK	2091		
196-208	CLADGAGDVAFVK	1371	1371	D 32
209-222	HSTVLENLPQEADR	1609	1608	D 31
223-230	DEYQLLCR	1145		
231-234	DNTR	505	505	* 8

235-235	K	147		
236-241	SVDEYK	740		
242-256	DCYLASIPSHAVVAR	1707		
257-261	SVDGK	505	505	* 8
262-280	ELDIWGLLNQAQEHFGTEK	2228		
281-282	SK	234		
283-293	DFHLFSSPHGK	1272		
294-298	DLLFK	635		
299-306	DSALGFLR	878		
307-326	IPPAMDTWLYLGYEYVTAIR	2372		
327-329	NLR	402	402	* 12a,b
330-338	EDIRPEVPK	1083	1083	D 25,26
339-342	DECK	599		
343-343	K	147		
344-345	VK	246		
346-354	WCAIGHHEK	1186	1186	* 26
355-356	VK	246		
357-382	CDEWSVNSGGNIECESAQST--			
	EDCIAK	3090		
383-385	IVK	359	359	* 11
			359	* 12
386-403	GEADAMSLDGGFIYIAGK	1815		
404-417	CGLVPVLAENYETR	1669		
418-437	SGSACVDTPEEGYHAVAVVK	2124	2128#	F1 28a
438-450	SSSDPDLTWNSLK	1450	1451	D 35
451-452	GK	204		
453-453	K	147		
454-461	SCHTGVDR	979	979	D 25
			980	* 25
			980	F1 23a
462-477	TAGWNIPMGLLYSEIK	1794		
478-483	HCEFDK	883	884	* 20
484-486	FFR	469		
487-494	EGCAPGYR	957		
495-495	R	175		
496-512	NSTLCNLCIGSASGPGR	1860		
513-521	ECEPNNHER	1233		
522-530	YYGYTGAFR	1098	1097	D 32

531-535	CLVEK	696	697	* 20
536-542	GDVAFVK	735	736	F1 27
543-553	HQTVEQNTDGR	1285		
554-560	NPDDWAK	845		
561-563	DLK	375	375	* 10
564-568	SENFK	624		
569-576	LLCPDGTR	980	979	D 25
			980	* 25
			980	F1 23a
577-577	K	147		
578-583	SVTEFK	710	710	* 22
			711	F1.23a
584-589	SCYLAR	817	817	* 22
590-598	APNHAVVSR	951	951	D 17
			951	* 17
599-599	K	147		
600-601	EK	276		
602-618	AACVCQELNQQASYGK	2060	2061	D 25
619-626	NGSHCPDK	962	961	D 26
627-635	FCLFQSATK	1150	1150	D 35
636-640	DLLFR	663		
641-656	DDTQCLANLQPTTTYK	1917	1918	D 29
657-662	TYLGEK	710	710	* 22
			711	F123a
663-671	YLTAVANLR	1021		
672-677	QCSTSR	786		
678-686	LLEACTFHR	1195	1195	D 29
687-687	V	118		

The hash mark (#) indicates a molecular ion found that is more than 1 mass unit away from a corresponding calculated mass value but there is supporting sequence information to confirm the presence of this molecular ion in this particular peak of the HPLC profile (see section 3.3.5).

Chapter 4 Discussion

4.1 Introduction

This study arose from the discovery of a new transferrin variant at the Equine Blood Typing and Research Centre at Massey University during routine blood-typing by starch gel electrophoresis of horse serum. This variant, which was given the designation *, was found in the progeny of a mating between two horses, both of which were known to carry the transferrin alleles D and F1. The particular horse which contained the new * variant (along with the D allele) was of considerable value as a sire and so it was of interest to characterize this new variant transferrin.

The aim of this project was to characterize the * transferrin variant and the variants D and F1, from one of which * must have arisen, by classical amino acid sequence analysis. A parallel study had also been initiated by M. Carpenter, in the the Department of Veterinary Pathology and Public Health, Massey University, to obtain the sequence of a cDNA of the horse transferrin gene. As described earlier, a sequence was obtained from clones originating from a D/F1 heterozygote. This sequence has been used in the present study to assign the amino acid sequences of tryptic peptides and to generate the theoretical mass values for the tryptic peptides. In this D/F1 cDNA sequence it was not originally known which regions of the sequence arose from the D allele and which from the F1. Therefore a further aim of the present study was to provide confirmation of the cDNA sequence (which may have included errors arising from use of PCR [polymerase chain reaction] techniques) and to identify any sequence differences between the D and F1 alleles, as well as to ascertain which of the two parental alleles might have given rise to the * allele.

[At a late stage in this project, data became available from M. Carpenter's work on the nucleotide sequence from clones from a F1 homozygote. This has provided information complementary to that obtained in the present investigation].

4.2 Purification of horse transferrin

Purification of the three horse transferrin variants F1, D, and * was achieved by a three step purification procedure. Firstly, after centrifugation to remove fibre and red blood cells from the blood, the supernatant was brought to 45% ammonium sulphate saturation. This precipitated some contaminating protein (which mainly consisted of globulin).

Secondly, a DEAE anion exchanger column was used to separate the protein components of the ammonium sulphate supernatant (which consisted mostly of albumin). Thirdly, a Q-sepharose anion exchanger was employed to remove the remaining protein contaminants.

The reason that this very short and simple purification procedure worked is probably due to the relatively high isoelectric point (pI) of transferrin. Penhallow (1991) determined the pI of diferric equine transferrin to be 5.2. The DEAE column and the Q-Sepharose column were run at pH values of 8 and 7.5 respectively. At these pH values transferrin would be in a less negative form than the proteins present with lower pI values, and would therefore elute earlier than the other proteins.

The transferrin purified by this method was examined on PAGE and (after staining with Coomassie Blue) single transferrin bands were seen with only one other very faint band. There appears to be two possible explanations for the presence of the faint band.

(1) The first explanation is that the transferrin has not been purified to homogeneity and the faint band is due to the presence of a small amount of a non-transferrin protein.

(2) A second explanation arises from a report of Makey and Seal (1976) that PAGE of purified human serum transferrin revealed additional bands which they identified as apotransferrins. Although the starting material used here was treated with iron-NTA to convert transferrins to the iron saturated form, it was possible that a detectable, small portion was not iron saturated, or released iron in the course of purification, giving rise to an additional band upon electrophoresis.

The DEAE-Sepharose chromatography elution profile (Fig 3.1) showed that transferrin fractions eluted from the DEAE anion exchanger as two separate peaks. The first peak was much larger than the second peak.

In contrast to the double transferrin peak found in the DEAE elution profile only a single transferrin peak was found in the Q-sepharose chromatography profile (see Fig 3.3). However, the 460nm absorbance of the Q-sepharose peak, showed significant tailing. PAGE analysis of fractions across the Q-sepharose peak showed that the mobility of the protein was faster in the later eluting fractions. Individual fractions from across the Q-sepharose transferrin peak were analysed by Mono Q column chromatography and three separate peaks (I, II, and III) were separated from the single Q-sepharose peak. N-terminal sequencing indicated that all three Mono Q peaks contained transferrin. The

presence of multiple peaks may be explained as arising from transferrin microheterogeneity (i.e. slightly different forms of the same transferrin variant). Interestingly, while Mono Q analysis showed the presence of discrete forms of transferrin this was not clearly evident in the PAGE analyses. For example, Fraction 31 (Fig 3.8) shows the presence of two distinct protein peaks in a fraction off the Mono Q column, but, when this sample was run on PAGE only a single band was detected (Fig 3.7). This may have been due to the inability of the gel system used to resolve very closely similar forms of protein. However, in a subsequent PAGE analysis (Fig 3.9) two bands have been resolved from this same fraction, perhaps due to a higher loading concentration.

Microheterogeneity of horse serum transferrin has been found to be related to variation in the number of glycans present on these glycoproteins and the nature of the sialic acid residues included in these glycans (Coddeville et al., 1989). The variant 2a contains only one glycan whilst 4b and 5b both contain two glycans. Fig 4.1 shows the structure of glycan type I, which is found in all three minor variants studied so far (i.e. 2a, 4b, and 5b). Glycan types II A, II B, and III are only present in the 2a and 4b variants. The structures of the four glycans so far identified have the same number of sugar residues and the same branch pattern, differing only in the type of sialic acid present. The sialic acids present in type I are both *N*-acetylneuraminic acid (Neu5Ac), and in type III they are both *N*-acetyl-4-*O*-acetylneuraminic acid [Neu(4,5)Ac_2]. Types II A and II B have one Neu5Ac residue and one Neu(4,5)Ac_2 residue each. In type II A the Neu5Ac residue is present on the branch joined to the core residues by an α1-6 glycosidic bond, and in type II B it is present on the branch joined to the core residues by an α1-3 glycosidic bond (Coddeville et al., 1989).

Also, microheterogeneity may be due to the loss of carbohydrate groups during circulation in the blood which may be related to the age of the protein.

Fig 4.1 also shows the location of the bonds where Endo-F and PNGase are known to cleave. Endo-F catalyses hydrolysis of the second glycoside bond in the glycan side chain leaving a single β-D-*N*-acetylglucosamine (GlcNac) monosaccharide attached to asparagine. PNGase cleaves the entire carbohydrate group from the protein forming an aspartic acid residue from the previously glycan-linked asparagine.

To determine whether peaks I, II and III resolved by Mono Q chromatography arose from variation of the pattern of glycosylation of transferrin an attempt was made to remove the carbohydrate moieties from the transferrin protein. Transferrin was treated

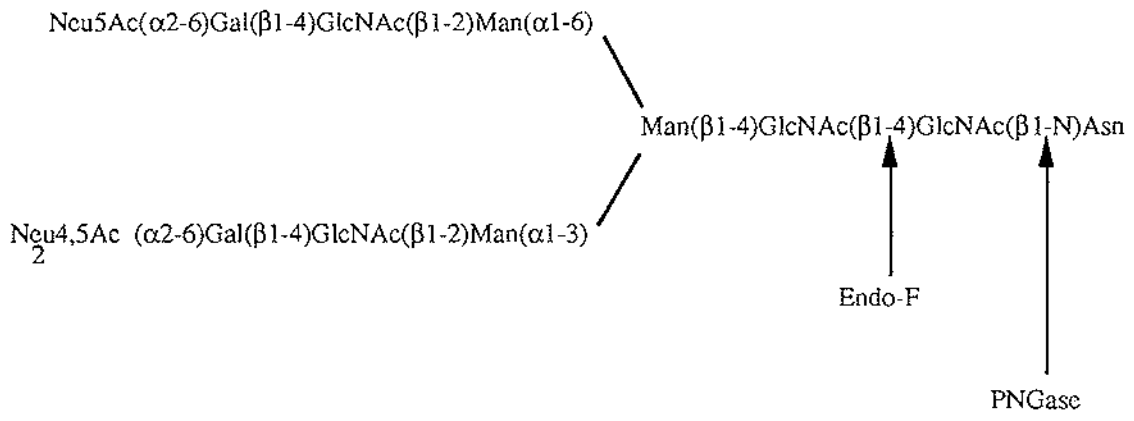

Figure 4.1 General structure of the carbohydrate moiety identified in horse serum transferrin

Four different carbohydrate moieties have been identified in horse transferrin (Coddeville et al., 1989). All the forms have the same structure as that shown above, differing only in the nature of the sialic acids present. The carbohydrate shown above is called type I and is present in Tf 2a, Tf 2b, and Tf 5b (see section 4.2).

with the deglycosylating enzymes Endo-F and PNGase and the treated and un-treated samples were examined by native PAGE (see Fig 3.9) and by chromatography on a Mono Q column (see Fig 3.8). PAGE analysis showed no significant difference in transferrin mobility between the treated and untreated samples. This indicated that no significant change in the overall charge of the protein had occurred. Therefore either the deglycosylating enzymes had probably not removed the carbohydrate moieties from the protein or the microheterogeneity was not due to glycosylation differences. It is possible that the Endo-F and PNGase enzymes were prevented from acting on transferrin due to the conformation of transferrin rendering the carbohydrate chain attachment sites inaccessible to the enzymes. Research performed by Dr G. E. Norris on rabbit transferrin has shown that rabbit transferrin is also not susceptible to the activities of Endo-F and PNGase (personal communication). Rabbit and horse transferrin have an amino acid sequence identity of 72.8%.

An attempt was made to remove sialic acids from the glycans. Fig 3.11 shows the FPLC profiles of transferrin before and after digestion with neuraminidase. There is a definite difference in the profile of each peak after digestion. However the profiles show more complexity than would be expected had just the sialic acid residues been removed. Instead many small peaks are seen indicating that protein degradation may have occurred. This is possibly the result of using neuraminidase contaminated with protease.

This work was done at a late stage in the present study so no further attempts have been made to remove the sialic acids using neuraminidase free of protease. However, it is known that neuraminidase does (to a degree) remove sialic acids from transferrin (Chung et al, 1985), and therefore this area of research does merit further study.

4.3 Comparison of the amino acid sequences of the horse transferrin variants D and F1

Comparison of the HPLC profiles and the partial amino acid sequences of the transferrin variants D, F1, and * enables some conclusions to be drawn concerning the nature and location of differences in the amino acid sequence between the D and F1 variants and the origin of the * variant. Two clear sequence differences between the D and F1 variant have been identified.

(1) The amino acid residue 553.

One major difference between the HPLC profiles of the D and F1 variants was the presence of a large peak (numbered 27) from D (Fig 3.16), for which there was only a small peak in the corresponding position from F1. Sequencing of peptides present in peak 27 of the D HPLC profile revealed that the peptide H_{543}-K_{559} (HQTVEQNTDGCNPDD(W)AK - see Fig 4.2) was a major constituent of this peak. This peptide was not present in F1 peak 27.

The presence of a reduced peak 27 in the F1 profile is due to the peptide GDVAFVK common to both variants.

The absence of peptide H_{543}-K_{559} in peak F1 27 suggests that there is a sequence difference between the D and F1 variants within this peptide which either changes the mobility of the peptide (i.e. a peak shift) or generates a tryptic cleavage site within the peptide resulting in the formation of two new peptides which elute at different positions in the profile.

Examination of the amino acid sequence predicted from the D/F1 cDNA (Fig 4.2) provides a probable explanation for the difference. The predicted sequence of this region contains an arginine residue at amino acid position 553 whereas the tryptic peptide H_{543}-K_{559} from the D variant contains a cysteine residue at this position.

The presence of arginine in the sequence predicted from cDNA could either be due to a sequencing error or that this part of the sequence was derived from an F1 genomic clone. If the cDNA sequence was correct and arginine was present in this position in F1 then trypsin-catalysed cleavage would be expected between the arginine and asparagine residues and the peptide H_{543}-K_{559} would not be found in the F1 elution profile but two separate peptides would be found elsewhere in the HPLC profile. Thus it should be

Figure 4.2 Map showing position of important sequences and amino acid residues relative to the whole transferrin amino acid sequence.

All major sequences mentioned in the discussion have been underlined here for ease of reference. Peptides sequenced from the D variant are underlined in blue, peptides sequenced from the * variant are underlined in red, and peptides sequenced from the F1 variant are underlined in green.

Residue 418 is a serine in the F1 variant but a proline in the * and D variants. Residue 553 is an arginine in the F1 variant but a cysteine in the * and D variants. Asparagine residues 496 and 619 are most probably sites of carbohydrate attachment.

possible to find the two new peptides HQTVEQNTDGR and NPDDWAK in the F1 digest.

The peptide HQTVEQNTDGR has not been detected but the peptide NPDDWAK has been found in the F1 HPLC profile in peak 23a for which there is no counterpart in the D HPLC profile.

From these results it can be concluded that the cDNA for this region of the transferrin gene has been cloned from an F1 allele. It can also be concluded that the substitution of a cysteine residue in the D variant for an arginine residue in the F1 variant is one of the significant differences between the two isoforms. (Presence of an arginine residue rather than the other possibility of a lysine residue would have to be confirmed by finding the peptide HQTVEQNTDGR and determining the terminal residue). Replacement of a positively charged arginine by a neutral cysteine in the D variant could account for the difference in electrophoretic mobility of the two variants.

Tryptic digestion of the * variant produced a large peak 27 in the * HPLC profile (see Fig 3.18). Sequence determination of peptides in * peak 27 confirmed that peptide H_{543}-K_{559} with a cysteine residue at position 553 was present. This leads to the important conclusion that the * variant has almost certainly arisen from the D allele rather than the F1 allele.

After completion of experimental work for this thesis, results from investigations on single stranded polymorphisms in the cDNA (M. Carpenter, personal communication) revealed the presence of the C/R polymorphism at position 553 confirming the findings of the present study.

(2) The amino acid residue 418.

Another difference in the HPLC profiles of tryptic digests from D and F1 is the presence of peak 38 in D (Fig 3.16) which is missing in F1 (Fig 3.17). Peak D 38 contains as the major component the peptide C_{404}-K_{437}, i.e. CGLVPVLAENYETRPGSACVDTPEEGYHAV(AVVK).

Examination of the D/F1 cDNA sequence (Fig 4.2) reveals a probable explanation for the absence of peak 38 in the F1 HPLC profile. The cDNA sequence showed polymorphism of the nucleotides coding for the amino acid at position 418 such that it could be either a serine or proline residue (M. Carpenter, 1992). It was not clear, however, whether this

was attributable to a difference between D and F1 and if so which of the two polymorphisms was present in the D variant and which was in the F1 variant.

In the peptide C_{404}-K_{437} there is a proline residue at this position following an arginine residue. The presence of a proline residue on the C-terminal side of an arginine residue would protect this site from tryptic cleavage as an arginine-proline peptide bond is not normally cleaved by trypsin.

If the corresponding residue in the F1 sequence was serine, then tryptic cleavage would occur between R_{417} and S_{418} replacing the peptide C_{404}-K_{437} by two smaller peptides CGLVPVLAENYETR (C_{404}-R_{417}) and SGSACVDTPEEGYHAVAVVK (S_{418}-K_{437}). It should be possible to find these two new peptides in the F1 tryptic digest.

Peak 28a in the F1 HPLC profile (Fig 3.17) was seen to be absent from the D HPLC profile (Fig 3.16). Peptide sequencing of the F1 28a constituents revealed the peptide S_{418}-K_{437} (Table 3.2b) thus confirming the prediction that the F1 variant contained a serine residue at position 418.

These results lead to the conclusion that the cDNA sequence containing the code for a proline residue at position 418 has been derived from a D allele while the sequence with serine at this position has been derived from a F1 allele.

Peak 38 is also present in the * variant and contains the peptide C_{404}-K_{437} with proline at position 418 confirming that the * variant is derived from the D allele not the F1 allele.

4.4 The transferrin variant *

The comparison of the sequence of those tryptic peptides discussed in the previous section, in which specific sequence differences between the D and F1 variants have been identified, with those of the * variant clearly indicate that the * phenotype has arisen from the D allele and not the F1 allele. However, there must be structural difference(s) between the D and * variants since they migrate differently on PAGE. Although many of the tryptic peptides from the * variant transferrin have been sequenced, unequivocal identification of a sequence difference between the * and D variants has not been obtained. Some possibilities are discussed below.

(i) Peak 5a

Comparison of the HPLC profiles of the tryptic digests of the D (Fig 3.16) and * (Fig 3.18) variants reveals the presence of a distinctive peak 5a in the * variant which is barely detectable in the D variant. Sequencing revealed that peak 5a of the * digest contained the tripeptide VPK. Inspection of the amino acid sequence predicted from the D/F1 cDNA sequence reveals that this tripeptide is present in the amino acid positions 336-338, since this tripeptide is contained within the sequence DIRPEVPKDE (see Fig 4.2), its presence in a tryptic digest of the * transferrin requires explanation. An explanation could be one of the following.

(a) There was an atypical tryptic cleavage between the glutamate and valine residues.

(b) There was a sequencing error in the cDNA sequence so that the residue preceding the valine was actually an arginine or lysine residue not a glutamate residue. A C/T difference in the first position of the triplet codon would result in a lysine/glutamate substitution. This is more probable than an arginine/glutamate substitution which would require two nucleotide changes in the triplet codon.

(c) The cDNA sequence was correct for the D variant so that no tryptic cleavage occurred but in the * variant a lysine residue replaced the glutamate residue allowing tryptic cleavage at the K_{335}-V_{336} bond.

The third possibility would not only provide a sequence difference between * and D but would also impart a charge difference and account for the difference in mobilities on PAGE.

If the third possibility was correct then the tryptic peptide EDIRPK (resulting from tryptic cleavage within the sequence (...LR/EDIRPK/VPK... - see Fig 4.2) would be present in the HPLC profile. The MacProMass programme HPLC index can be used to predict the possible elution position of such a hypothetical peptide on the reverse phase HPLC elution system used for the tryptic peptides. On the basis of this prediction the mass values of peptides in the region from peak 9 to peak 30 were determined by mass spectrometry to see if a peptide with a mass value corresponding to that of EDIRPK (or R) could be found. No peak with an appropriate mass was identified. Subsequently N-terminal sequencing of peptides in peak 24 of the * variant revealed the presence of the tryptic peptide EDIRPEVP(K). This establishes that the sequence predicted from the cDNA is correct and that a glutamate residue is present in position 335 in the * variant. Therefore the presence of VPK in peak 5a of the * profile must have resulted from an atypical tryptic cleavage. There is no obvious explanation for its presence in the * variant HPLC profile but not in the D HPLC profile.

(ii) Peak 11

Another peak which showed a significant difference in size between the D and * HPLC profiles was peak 11 which was much larger in the transferrin D HPLC profile than in the * profile. The major peptide found in peak D 11 was the sequence XGSHCPDK in which the first residue was not detected by the peptide sequencer. Inspection of the amino acid sequence predicted from the D/F1 cDNA indicates that the first residue in this peptide should be an asparagine. The sequence NGS is a recognition sequence for glycosylation and this peptide is located in the C-terminal lobe of the horse transferrin sequence where both N-glycosylation sites are known to occur in human serotransferrin (Baker et al., 1987). A glycosylated asparagine residue at position 619 would explain why the sequencer failed to detect these residues. The important conclusion can therefore be drawn that the sequence NGS probably does signal for the attachment of a glycan moiety in horse transferrin and that one of the carbohydrate groups of horse transferrin is most probably located on the asparagine residue at position 619.

A small quantity of the peptide XGSHCPDK was detected in peak 11 of the * transferrin HPLC profile and a peptide with the same sequence was also found in peak 21 of the * variant. If indeed the N-terminal asparagine residue is a site of N-glycosylation in horse transferrin then an explanation for the appearance of the peptide N_{619}-K_{626} in both peaks 11 and 21 of the HPLC profile of the * variant could be that the two peptides contain

different glycan groups. This is possibly the case since any one of four known glycan groups could be attached to this one peptide (see sections 1.6 and 4.1). Therefore, an explanation could be put forward for the appearance of a larger peak 11 in the HPLC profile of the D variant than in the HPLC profiles of the * and F1 variants, i.e. the carbohydrate moiety that causes elution of the peptide N_{619}-K_{626} in peak 11 is present, attached to this peptide, in greater amounts in variant D than in variants * or F1. The occurrence of this specific peptide and carbohydrate group in larger amounts in the D transferrin variant than the * and F1 transferrin variants may be,

(a) the result of chance, i.e. upon obtaining blood from different sources (or another lot of blood from the same source) and isolating out the transferrin variants, peak 11 would appear in the HPLC profiles with a random variation in size,

or (b) inherent to the specific variant, i.e. the location of specific glycans on specific peptides in a certain ratio to the other glycopeptides is somehow determined by the structure of the individual variant.

These possibilities could be tested as follows.

(1) The transferrin variants D and * could be treated with deglycosylation enzymes to see if this removed the differences in mobility.

(2) The mass spectra for peaks 11 and 21 could be examined for molecular ion species with mass values greater than that for the deglycosylated peptide. Unfortunately mass spectra were only recorded up to a certain mass limit based on the likely maximum for deglycosylated peptides so glycosylated peptides would not have been evident in the mass spectra obtained in this study.

(3) HPLC profiles of tryptic digests of transferrin could be obtained from blood from different sources with the same transferrin genotype. Comparison of these profiles may show variation or no variation in size of the putative glycopeptide containing peaks.

(4) Glycopeptides might be isolated from the tryptic digests, perhaps by using a glycopeptide-binding adsorbent column e.g. Concanavalin A. Analysis of the isolated peptides by HPLC may reveal the relative amounts of glycopeptides present.

(iii) Peak 9

Peak 9 is present in the HPLC profile of the * variant (see Fig 3.18) but is barely detectable in the HPLC profile of the D variant (see Fig 3.16). Peptide sequencing of peak 9 in the * variant revealed that this peak contains a small amount of the sequence DLK. This same peptide was present in peak 10 (which is same size in D and *) and is probably found in peak 9 due to the collection of part of the leading edge of peak 10. A very large amount of tyrosine was recorded in the first sequencing cycle but no corresponding sequence could be found in the following cycle profiles. Tyrosine may therefore be present as a single amino acid residue, due to an atypical tryptic cleavage, and the large amount of tyrosine present may account for the appearance of peak 11 in the * profile. However, the mass spectrometry profile obtained for peak 9 does not show the presence of a molecular ion with the mass of tyrosine to confirm this result.

(iv) Peak 51

Peak 51 is very large in the HPLC profile of the D variant but relatively small in the * HPLC profile. The D 51 peak contains only one peptide, i.e. SAGWNIPIGLLYWQLPEPR (see Table 3.2a), whereas peak 51 of the * variant contains two peptides, S_{122}-R_{140} and A_{48}-K_{85} (see Table 3.2c). The A_{48}-K_{85} peptide in peak 51 is probably due to incomplete resolution from peptides present in later peaks since it is also found in peak * 54 (see Table 3.2c). An explanation cannot be provided for the large difference in size of peak 51 between the two variants. Examination of the region in which peak 51 elutes shows that the proportion of all peptides present is relatively higher in the D HPLC profile than in the * HPLC profile.

(v) Peak 28

Initial peptide sequencing showed that the peptide XSTLCNLCIGSASGPG(R) was present in peak 28 of the D HPLC profile. Examination of the D/F1 amino acid sequence revealed that the N-terminal amino acid may be an asparagine residue. If indeed the N-terminal was an asparagine residue then the first three amino acids (NST) could be a signal sequence for the attachment of a carbohydrate groups to the asparagine. If this residue was glycosylated then it would not be recorded in peptide sequencing. Thus, it is

probable that this residue is glycosylated. From these results it can be tentatively concluded that,

(1) the second site for N-glycosylation in horse serotransferrin is at the asparagine residue N_{496},

and (2) both sites of N-glycosylation in horse serotransferrin are found in the latter half of the C-terminal domain.

4.5 Agreement between amino acid sequence information determined by N-terminal sequencing and that predicted from cDNA sequence.

Apart from the polymorphisms that occur at amino acid positions 418 and 553 (as discussed in section 4.2) all sequencing data obtained in this study agrees with the amino acid sequence predicted from the cDNA sequence (M. Carpenter, thesis - 1992). Therefore, the N-terminal peptide sequencing performed in this study provides the necessary conformation that much of the predicted amino acid sequence is correct.

4.6 Further characterization of the * variant.

While this present study has indicated that the * variant has arisen from the D allele the nature of the difference has not been defined. This difference may be characterized by:

(a) obtaining further N-terminal peptide sequencing information, especially for those HPLC peaks that were found to contain molecular ions that did not correspond to any of the predicted tryptic peptides of the * variant.

(b) obtaining further mass spectra for the * variant, especially for those HPLC peaks for which there is, as yet, no N-terminal sequence information.

(c) digesting the protein with another enzyme or chemical that will produce large peptides and provide an HPLC profile with only a small number of peaks, thus probably providing a better peak separation, e.g. cyanogen bromide.

(d) examining the digest HPLC profiles at other wavelengths, e.g. at 280nm only those peptides containing tryptophan will be picked up spectrophotometrically.

(e) labelling specific amino acids so that only peptides containing labelled amino acids will show up at the specific wavelength at which the label absorbs.

(f) sequencing the * allele by cDNA techniques as this would allow a sequence to be predicted for the complete protein and it would also provide confirmation of the * variant peptide sequences derived by N-terminal sequencing.

References

Elder, J. H. & Alexander, S. (1982) *endo-B-N-* Acetyiglucosaminidase F: Endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. *Proc. Nat. Acad. Sci., U.S.A.* **79**, 4540-4544

Andersson, L., Arnason, Th., Sandberg, K. (1987) Biochemical polymorphism in relation to performance in horses. *Theor. Appl. Genet.***73**, 419-427.

Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W., Baker, E. N. (1989) Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 Å resolution. J. Mol. Biol. 209: 711-734.

Baer, A., Schwendimann, B. (1972) Biochemical and immunological comparison of phenotypes of horse serum transferrins. *Experientia* **28**, 16-17.

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., Rumball, S. V., & Anderson, B. F. (1987) *Trends Biochem. Sci.*, **12**, 350-353.

Bell, K., Pollitt, C. C., & Patterson, S. D. (1988) Subdivision of equine Tf into H_1 and H_2. *Animal Genetics*, **19**, 177-183.

Bezkorovainy, A. (1980) Biochemistry of non-heme iron. pp129-130.

Bezkorovainy, A. (1980a) Biochemistry of non-heme iron. pp159-160.

Bezkorovainy A. Iron Proteins; from Iron and Infection (1987) Bullen J.J. & Grffiths E.

Bezkorovainy, A., & Zschocke, R. H. (1974) Structure and function of transferrins. I. Physical, chemical and iron-binding properties. *Arzneim. Forsch.* **24**, 476-485.

Bowling, A. T. (1991) An unusual transferrin variant in a Thouroughbred stallion. *Animal Genetics* **22**, (Suppl. 1): 18.

Bowling, A. T., & Clark, R. S. (1985) Blood group and protein polymorphism gene frequencies for seven breeds of horses in the United States. *Anim. Blood Groups Biochem. Genet.* **16**, 93-108.

Bowman, B.H., Yang, F., Adrian, G. S. (1988) Transferrin: Evolution and genetic regulation of expression. *Advances in Genetics* **25**, 1-38.

Braend, M. (1964) Serum types of Norwegian horses. *Nordisk Veterinaermedicin* **16**: 363-373.

Brock, J. H., & Mainou-Fowler (1983) Immunology Today **4**, 347-351.

Carpenter, M. (1992) The cDNA sequence and polymorphism of horse transferrin. Thesis, Ph. D. Massey University.

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

Chung M. C.-M., & Mckenzie H. A. (1985) Studies on equine transferrin - 1. The isolation and partial characterization of the D and R variants. *Comp. Biochem. Physiol.* Vol. **80B**, No. 2, 287-297.

Coddeville, B., Stratil, A., Wieruszeski, J. M., Strecker, G., Montreuil, J., & Spik, G. (1989) *Eur. J. Biochem.* **186**, 583-590.

Cothran, E. G., Henney, P. J., & King, J. A. (1991) Inheritance of the equine Tf F3 allele. *Animal Genetics*, **22**, 187-190.

Crichton, R. R. (1991) Inorganic Biochemistry of iron metabolism, p 120.

de Jong, G., & van Eijk, H. G. (1989) Functional properties of the carbohydrate moiety of human transferrin. *Int. J. Biochem.* **21**, 253-263.

Dubrovskaya, R. M., & Starodumov, I. M. (1989) Reproductive ability of mares with different levels of heterozygosity for alleles at the transferrin, albumin, esterase and D blood group loci. *Sovrem. sostoyanie i perspektivy razvitiya nauch. po konevod. Tez dolk. vses. nauch. soveshch.* **1989**, 77-78.

Evans, R. W., & Williams J. (1980) The electrophoresis of transferrins in urea-polyacrylamide gels. *Biochem. J.* **189**, 541-546.

Evans, R. W., Meilak, A., Aitken, A., Patel, K. J., Wong, C., Garratt, R. C., & Chitnavis, B. (1988) Characterization of the amino acid change in a transferrin variant. *Biochemical Society Transactions* **16**, 834-835.

Farndale, B. M., Anderson, I. L., Moore, C. H. Identification of an apparent transferrin mutation in a commercial Thouroughbred Stallion. *Animal Genetics* **22**, (Suppl. 1): 19.

Hayashi, I. & Sato, G. H. (1976) Replacement of serum by hormones permits growth of cells in a defined medium. *Nature* **259**, 132-134.

Hesselholt, M. (1966) Studies on blood and serum types of the Icelandic horses. *Acta Vet. Scand.* **7**, 206-225.

Holmberg, C.G., & Laurell, C.-B. (1947). Investigation of serum copper. I. Nature of serum copper and its relation to the iron-binding protein of human serum, *Acta Chem. Scand.* **1**, 944-950.

Hudson, B. G., Ohno, M., Brockway, W. J., & Castellino, C. J. (1973) Chemical and physical properties of serum transferrins from several species. *Biochemistry. N.Y.* **12**, 1047-1053.

Huebers, H. A., & Finch, C. A. (1987) Physiological Reviews **67**, 520-582.

Hutchings, S. E. & Sato, G. H. (1978) growth and maintenance of Hela cells in seum-free medium supplemented with hormones.*Proc. Natl. Acad. Sci. USA* **75**, 901-905.

Janiszewska, Jolanta, & Tomaszewska-Guszkiewicz, K. (1990) Polymorphism of blood proteins in Ardenne horses reared in Poland: I. Bielin Horse Stud. *Genet. Pol.* **31**, No. 2, 137-144.

Juneja, R.K., Sandberg, K., Kuryl, J., & Gahne, B. (1989) Genetic polymorphism of horse serum protein 3 (SP3). *Animal Genetics*, **20**, 43-49.

Kaminiski M., Didkowski S. & Skiotis M. (1981) Polymorphism of transferrin locus in horses: immunological evedence of two structurally different subgroups of the allelic proteins. *Acta Biochemica Physiologica* **68B**, 505-507.

Kornfeld, S. (1968) The effects of structural modifications on the biologic activity of human transferrin. *Biochemistry* **7**, 945-954.

Kornfeld, S. (1969) The effect of metal attachment to human apotransferrin on its binding to reticulocytes. *Biochim. Biophys. Acta* **194**, 25-33.

Kuehnl, P., & Spielmann, W. (1978) Transferrin: evidence for two common subtypes of the Tf^C allele. *Human Genetics* **43**: 91-95.

Lane, R. S. (1971) DEAE-cellulose chromatography of human transferrin. The effect of increasing iron saturation and copper (II) binding. *Biochim. Biophys. Acta.* **243**, 193-202.

Maeda, K., Mckenzie, H. A., & Shaw, D. C. (1980) Nature of the heterogeneity within genetic variants of bovine serum transferrin. *Anim. Blood Grps biochem. Genet.* **11**, 63-75.

Makey D. G. and Seal U.S. (1976) The detection of four molecular iron binding forms of human transferrin during the iron binding process. *Biochim. biophys. Acta* **453**, 250-256.

Malagolini, N., Dall'Olio, F., Serafini-Cessi, F., & Cessi, C. (1989) Effect of acute and chronic ethanol administration on rat liver α2,6-sialyltransferase activity responsible for sialylation of serum transferrin. *Alcohol Clin. Exp. Res.* **13**, No. 5, 649-653.

Morgan, E. H. (1981) Transferrin, biochemistry, physiology and clinical significance. *Mol. Aspects Med.* 4: 1-123

Murate, Takashi, Kaneda, T., Hotta, T., & Goto, S. (1988) Differentiation of human erythroleukemia K562 cells without chemical inducer and the role of iron-saturated transferrin. *Acta. Haematol. Jpn.* **51**, No. 6, 1006-1015.

Norris, G. E., Baker, H. M., Baker, E. N. (1989) Preliminary Crystallographic Studies on Human Apolactoferrin in its Native and Deglycosylated Forms. *J. Mol. Biol.* (1989) **209**, 329-331.

Nozawa, K., Shotake, T., & Ohkura, Y. (1976) Blood protein variations within and between the east Asian and European horse populations. *Z. Tierzüchtg Züchtgsbiol.* **93**, 60-74.

Osterhoff, D. R., Azzie, M. A. J., Op't Hof, J. (1974) Biochemical genetics and performance ability in horses. *J. S. Afr. Vet. Assoc.* **45**, 311-316.

Penhallow, R. C., Mason, A. B., & Woodworth, R.C. (1991) Electrophoretic characterization of human, equine and bovine transferrins. *Comparative Biochemistry.* **98**, No. 1, 41-45.

Pepper, D. S. (1968) The sialic acids of horse serum with special reference to their virus inhibitory properties. *Biochim. Biophys. Acta* **156**, 317-326.

Putt, W., & Whitehouse, D.B. (1982). Genetics of four plasma protein loci in *Equus przewalskii* : new alleles at the prealbumin, postalbumin and transferrin loci. *Animal Blood Groups Biochemical Genetics* **14**, 7-16.

Regoeczi, E., Wong, K. -L., Ali, M., Hatton, M. W. C. (1977) The molecular components of human transferrin type C. *Int. J. Pept. Protein Res.* **10**, 17-26.

Reisinger, P. W. M., & Soyka, M. (1990) The diagnosis of alchoholism on the basis of detection of a transferrin variant by polyacrylamide gel electrophoresis and immunoblotting. *Blutalkohol.* **27**, No. 6, 427-433.

Rodriguez-Gallardo, P. P., & de Andres-Cara, D. F., (1990) The TfJ allele in the transferrin genetic system in Spanish Thoroughbred horses. *Archivos-de-Zootecnia.* **39**, 145: 233-238.

Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H., Lindley, P. (1990) High-resolution X-ray studies on rabbit serum transferrin: preliminary structure analysis of the N-terminal half-molecule at a 2.3 Å resolution. Acta Crystallogr. B46: 763-771.

Schellenberg, F., Bernard, J. Y., Le Goff, A. M., Bourdin, C., & Weill, J. (1989) Evaluation of carbohydrate-deficient transferrin compared with Tf index and other markers of alcohol abuse. *Alcohol Clin. Exp. Res.* **13**, No. 5, 605-610.

Schimd D.O., Ek, N.,\ & Braend M. (1990) Further evidence for a silent allele in the transferrin locus of the horse. *Animal Genetics* **21**, 423-426.

Schreiber, G., Dryburgh, H., Millership, A., Matsuda, Y., Inglis, A., Philips, J., Edwards, K., Maggs, J. (1979) The synthesis and secretion of rat transferrin. *J. Biol. Chem.* **254**, 12013-12019.

Scott, A. M. (1980) A furthur allele in the transferrin system of the horse. *Anim. Blood Groups Biochem. Genet.* **11** (Suppl. 1): 67.

Seligman, P. A. (1983) Prog. Hematol. **13**, 131-147.

Singh, A., & Choudhary, R. P. (1989) Serum transferrin polymorphism and its association with growth and reproductive traits in sahiwal and crossbreds. *Indian vet. J.* **66**, 933-937.

Spik, G., Streker, G., Fournet, B., Bouquelet, S., Montreuil, J. (1975) *FEBS Letts.*, **50**, 296-299.

Spooner, R. L., Oliver, R. A., Richardson. N., Buttress, N., Feinstein, A., Maddy, A. H., & Stratil, A. (1975) Isolation and partial characterization of sheep transferrin. *Comp. Biochem. Physiol.* **52B** 515-522.

Starodumov, I. M. (1989) Reproductive ability of mares in relation to alleles at the transferrin locus. *Sovrem. sostoyanie i perspektivy razvitiya nauch. po konevod. Tez dolk. vses. nauch. soveshch.* **1989**, 69-71.

Stibler et al (1986) *Alcohol Clin Exp Res* **10**:534-544.

Stibler H., Borg S. (1986) Carbohydrat composition of serum transferrin in alcoholic patients. *Alcohol Clin Exp Res* **10**, 534-544.

Stratil, A., & Glasnak, V. (1981) Partial characterization of horse transferrin heterogeneity with respect to the atypical type, Tf C. *Animal Blood Groups and Biochemical Genetics* **12**, 113-122.

Stratil, A., & Kúbek, A. (1974) Heterogeneity of pig transferrin. *Int. J. Biochem.* **5**, 895-899.

Stratil, A., & Spooner, R. L., (1971) Isolation and properties of individual components of cattle transferrin: the role of sialic acid. *Biochem. Genet.* **5**, 347-365.

Stratil, A., Tomasek, V., Bobák, P., & Glasnak, V. (1984) Heterogeneity of horse transferrin: the role of carbohydrate moiety. *Anim. Blood Groups Biochem.Genet.***15**, 89-101.

Thorstensen, K., Romslo, I. (1990) The role of transferrin in the mechanism of cellular iron uptake. *Biochem. J.* **271**, 1-10.

Tsuji, S., Kato, H., Natsuoka, Y., & Fukushima T. (1984) Molecular Weight Heterogeneity of Bovine Serum Transferrin. *Biochemical Genetics* Vol. **22**, Nos. 11/12 1145-1159.

Valenta, M., Stratil, A., Slechtová, V., Kálal, L., Slechta, V. (some letters need symbols) (1976) Polymorphism of transferrin in carp (*Cyprinus carpio* L.): Genetic determination, isolation and partial characterization. *Biochemical Genetics* **14**: 27-45. **(35)**

Van Niekerk, F. E., Van Niekerk, C. H., Reid, G., & Coetzee, J. (1990) The bent-leg syndrome in sheep II. Association with genetic markers. *J. S. Afr. Vet. Assoc.* **61**, No. 4, 151-154.

Watanabe, K., Sohara, T., Takeda, M., Ueno, K., Suzumura, N., Rokurouda, Y., Rokurouda, I., & Yamamoto, S. (1989) Difference in content ratio of components among horse serum transferrin variants. *Japanese Journal of Zootechnical Science.* **51**, No. 5, 969-974.

Weitkamp,L. R., Kenney, R. M., Bailey, E., MacCluer, J. W., Brown, J. S., Blanchard, T. L., Sertich, P. L., Love, C. C., Hunt, P. R. (1991) Pathological changes of the mare endometrium and genotypes for transferrin and ELA. *J. Reprod. Fertil. Suppl.* 44, 275-282.

Welch, & Langmead, S & L. (1990) A comparison of the structure and properties of normal human transferrin and a genetic variant of human transferrin. *Int. J. Biochem.* **22**, No. 3, 275-282.

Williams, J. (1962) A comparison of conalbumin and transferrin in the domestic fowl. *Biochemical Journal* **83**, 355-364.

Wong, K. -L., & Regoeczi, E. (1977) Some observations on the carbohydrate composition of purified transferrin. *Int. J. Pept. Protein Res.* **9**, 241-248.

Wong, K. -L., Debanne, M. T., Hatton M. W. C., & Regoeczi, E. (1978) Human transferrin, asialotransferrin and the intermediate forms. *Int. J. Pept. Protein Res.* **12**, 27-37.

Yamashita, K., Koide, N., Endo, T., Iwaki, Y., & Kobata, A. (1989) Altered glycosylation of serum transferrin of patients with hepatocellular carcinoma. *J.Biol.Chem* **264**, No. 5, 2415-2423.

Yokohama M., Mogi K. & Hosoda T. (1982) Relationship between serum iron and transferrin concentrations in light horse breeds. *Japanese Journal of Zootechnical Science* **53**, 766-773.

Yokohama, M., Watanabe, Y., Kobayashi, E., Snotake, T., Nozawa, K., & Mogi, K. (1989) Classification for transferrin and esterase types in horses.*Japanese Journal of Zootechnical Science*. **60**, No. 2, 115-120.

Yoshinari, Kaoru, Yuasa, K., Iga, F., & Mimura, A. (1989) A growth-promoting factor for human myeloid leukemia cells from horse serum identified as horse serum transferrin. *Biochem. Biopys. Acta.* **1010**, No. 1, 28-34.