The cDNA Sequence and Polymorphism of

Horse Transferrin

A thesis presented in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Veterinary Clinical Science
at Massey University, Palmerston North,
New Zealand.

Margaret Ann Carpenter

1992

ABSTRACT

Transferrin, the serum protein responsible for iron transport, is highly polymorphic in many species, including the horse. In this study, the cDNA sequence of horse transferrin was determined and used to identify sequence polymorphisms which distinguish some of the many variants of horse transferrin.

A horse liver cDNA library was constructed, and was screened using the human transferrin cDNA as a probe. The clones isolated gave 1800 bp of sequence. The remainder of the cDNA was obtained using PCR. The 2305 bp horse transferrin cDNA sequence included part of the 5' untranslated region and extended to the poly(A) tail. It had 80% sequence identity with the human transferrin cDNA, and encoded a protein of 706 residues, including a signal sequence of 19 amino acids. The amino acid sequence was compared to those of related proteins, i.e. the serum transferrins and lactoferrins of several species and human melanotransferrin. The horse transferrin sequence had the duplicated structure and 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% amino acid sequence identity to human transferrin, 62% identity with human lactoferrin, and 53% identity with chicken transferrin.

Sequence polymorphisms distinguishing the variants of horse transferrin which are found in thoroughbreds (D, F_1, F_2, H_2, O, R and *) were identified. First, comparison of the horse transferrin cDNA and protein sequences with two reported amino acid substitutions distinguishing the D and R variants indicated that exons 12 and 15 were likely to be polymorphic. Therefore, these two regions were analysed by Southern blotting and by sequencing PCR products. The D and R variants differed by 10 nucleotide substitutions which encoded 6 amino acid substitutions. The F_1, F_2, H_2 and * variants were identical to D, and the O variant was very similar to R, in the regions studied. The data indicated that the horse transferrin variants comprise two distinct groups. Secondly, the positions of differences between the D and F_1 alleles were established by producing single stranded conformation polymorphisms. Sequencing then revealed 3 nucleotide substitutions that encode 2 amino acid substitutions. All 8 of the amino acid substitutions which were observed occurred at positions which are variable between members of the transferrin family. Location of the polymorphic residues on the 3-dimensional structure of human lactoferrin revealed that all were clustered at one end of the C-lobe.

ACKNOWLEDGEMENTS

I wish to thank Dr. Tom Broad (DSIR) for making this project possible and for providing unfailing support throughout, and also Dr. Eric Terzaghi (Department of Microbiology and Genetics, Massey University) and Dr. Brian Goulden (Department of Veterinary Clinical Sciences, Massey University) for their advice and encouragement. I gratefully acknowledge the Biotechnology Division of DSIR, DSIR Grasslands and the New Zealand Pastoral Agriculture Research Institute, Ltd. for providing facilities and funding.

Thanks to Dr. Ross MacGillivray, University of British Columbia, for providing the human transferrin cDNA; Dr. Ian Anderson, Bruce Farndale and other staff of the Equine Blood Typing and Research Centre, Massey University, for blood typing; Dr. John Tweedie, Department of Chemistry and Biochemistry, Massey University, for guidance with RNA preparation and providing pGEM plasmids; and Dr. Ted Baker, Department of Chemistry and Biochemistry, Massey University, for providing a photograph of human lactoferrin.

I am grateful to members of the Gene Mapping Unit of the N.Z. Pastoral Agriculture Research Institute (formerly of DSIR Grasslands and Biotechnology Division, DSIR), particularly Pauline Lewis, Sin Hen Phua and Andrea Ede for their advice and company in the lab. Thanks also to other staff of DSIR, such as John Forrest and Bruce Davidson for providing horse genomic DNA; Nick Ellison for providing the Clustal V package, λgt10 and *E. coli* C600 and C600 *hfl*A; Peter Spring for photographic services; and Paul Ealing for synthesis of PCR primers.

Thanks to friends, flatmates and family who listened, sympathised, supported and tried to understand, especially Tim and Grant.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ATP	adenosine triphosphate
bp	base pairs
BRL	Bethesda Research Laboratories
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
cpm	counts per minute
cps	counts per second
dCTP	deoxycytosine triphosphate
DTT	dithiothreitol
EBTRC	Equine Blood Typing and Research Centre, Massey University
kb	kilobases
LF	lactoferrin
mRNA	messenger RNA
MTF	melanotransferrin
NTP	nucleotide triphosphates
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
SSCP	single stranded conformation polymorphism
TEMED	tetramethylethylenediamine
TF	transferrin
U	units

Abbreviations for other reagents are defined in the Reagents section (Section 2.6)

ABBREVIATIONS

Amino Acids

A	alanine
C	cysteine
D	aspartate
E	glutamate
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophan
Y	tyrosine

CHAPTER 1: INTRODUCTION

"The transferrin gene is of biological interest not only because of transferrin's importance in iron binding and cell growth, but also because of its interesting evolutionary history." Yang *et al.* (1985)

"The transferrins of all animal species so far examined show a high degree of genetic polymorphism, yet disorders of transferrin function which are dependent on gene mutations are extremely rare." Morgan (1981)

"One would like to know, for example, what is the precise nature of the structural differences between the variant forms of a given enzyme, and whether these are reflected....in differences in functional activity." Harris (1966)

Transferrin is a serum protein involved in iron metabolism. It is essential for the survival of the animal through its role in iron transport, and has important roles in cell proliferation and resistance to infection. It is a member of a family of structurally similar proteins, including lactoferrin, which occurs in milk, and ovotransferrin, a protein in avian egg white. Serum transferrin occurs in all vertebrate classes, and related proteins have been detected in invertebrates. Transferrin is a highly polymorphic protein; multiple variants have been described for many species, but little is known of the molecular basis of this variation.

1.1 FUNCTIONS OF TRANSFERRIN

1.1.1 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 Fe^{3+} oxidation state. The solubility of these ferric ions in such fluids is very low. Free iron can be toxic through its activity in redox reactions, which involve production of radicals that can damage cells (Aisen and Listowsky, 1980). Consequently, iron *in vivo* must be bound by proteins which can stabilize it in a soluble form, making it available for biological processes. Ferritin, an iron storage protein, stabilizes

aggregates of iron ions by surrounding them with a protein shell. Transferrin binds iron deep within clefts in the protein, allowing it to be transported in a soluble form.

Iron is a necessary component of several proteins. The major requirement for iron in vertebrates is for the production of haemoglobin, the oxygen-carrying protein of the blood. Consequently, the haematopoietic tissues which produce erythrocytes must be efficiently supplied with iron. The liver is the major source of iron, being a site of both erythrocyte catabolism, and of iron storage. The intestinal mucosa is another important source, supplying iron derived from the diet. Erythrocyte catabolism occurs in the Kupffer cells of the liver and macrophages of the spleen. Transferrin is required to transport iron from these sites of storage, absorption, and catabolism to the haematopoietic system, and to proliferative cells such as lymphocytes. During pregnancy it also facilitates delivery of iron to the placenta for the foetus (Brock, 1985).

The accepted model for the delivery of transferrin-bound iron to cells is receptor mediated endocytosis. This involves the binding of transferrin, with its attached iron, to a specific receptor on the cell surface. The iron-transferrin-receptor complex is internalized as part of an endocytotic vesicle. The receptor returns to the cell surface, and the transferrin is released back into the bloodstream. Enns *et al.* (1983) demonstrated the co-internalization of transferrin and its receptor in an erythroid cell line. They also observed that binding of transferrin to the receptor promotes capping, a process known to lead to internalization.

Transferrin binds iron with a high affinity, so a specific mechanism is required for the release of iron. Removal of iron may occur within endocytotic vesicles, where the pH could be lowered to reduce the transferrin-iron affinity. At physiological pH, binding to the receptor increases the affinity of transferrin for ferric ions, so that iron is not lost at the cell surface. At lower pH (pH 5.6), that of endocytotic vesicles, iron dissociates more readily from receptor-bound transferrin than from free transferrin. Therefore, once inside the cell, the receptor facilitates iron release (Bali *et al.*, 1991). This effect is mainly due to changes in the C-lobe, the N-lobe being more consistent (Bali and Aisen, 1991).

The uptake of transferrin-bound iron may also occur by other mechanisms. Zaman *et al.* (1980) showed that iron uptake by rat reticulocytes exceeded transferrin uptake, and concluded that two different processes must occur. It has been suggested that reductive iron release may be a more important process than receptor mediated endocytosis (Thorstenson and Romslo, 1990).

In the reductive iron release model, diferric transferrin is again bound to the transferrin receptor. Ferric ions are reduced to ferrous ions by NADH:ferricyanide oxidoreductase located in the cell membrane close to the receptor, which weakens the transferrin-iron bond. The ferrous ion is released from transferrin and transferred to a membrane protein which is specific for Fe^{2+}. The ferrous ion passes through the membrane and is released to a cytosolic protein. Iron uptake may occur by both processes, but their relative importance may differ between cell types or circumstances.

Apotransferrin has a lower affinity for the receptor than iron-saturated transferrin (Kornfeld, 1969), so it is readily released from the receptor after iron removal. Some transferrin is degraded within the cell, but most is released into the bloodstream for reuse (Brock,1985).

The transferrin receptor is a glycoprotein of molecular weight 170,000 - 200,000, made up of two identical subunits (Fernandez-Pol and Klos, 1980 (rat); Trowbridge and Omary, 1981 (human); Goding and Harris, 1981 (mouse)). It is covalently bound to palmitic acid, which is thought to anchor the receptor in the cell membrane (Omary and Trowbridge, 1981). Transferrin receptors are found at highest density on erythrocyte precursor cells, but also occur on mammary gland cells, fibroblasts, lymphocytes, hepatocytes, spleen cells, thymus cells (Brock, 1985) and on many tumour and leukemic cell lines (Sutherland *et al.*, 1981). The number of receptors on erythroid precursors decreases as the cell matures, indicating a decreasing requirement for iron. The receptor is specific for transferrin and will not bind the related molecules lactoferrin and ovotransferrin. Optimal binding of transferrin to its receptor occurs at pH 7-8 (Brock, 1985).

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 and Connor, 1990).

1.1.2 Growth factor

Transferrin appears to be essential for cell proliferation. Most serum-free cell culture media require transferrin as an essential component for cell transformation and proliferation. Transferrin stimulates proliferation and differentiation of muscle cells (Beach *et al.*, 1983), kidney cells (Ekblom *et al.* (1983), lymphocytes (Tormey and Mueller, 1972) and hepatoma

cells (Shapiro and Wagner, 1989). Transferrin also enhances the mitogenic effects of thyroid stimulating hormone, forskolin and dibutyryl cAMP on rat thyroid follicular cells, but has little effect on its own (Lombardi *et al.*, 1989). Transferrin is required for the growth and differentiation of rat embryos in culture, and prevents anaemia in the embryos (Gulamhusein *et al.*, 1990).

The growth promoting properties of transferrin are thought to be a consequence of its ability to supply iron. Laskey *et al.* (1988) have produced evidence that iron is essential for DNA synthesis, and that transferrin functions as a growth factor solely by supplying iron. It has been suggested that transferrin has a role in cell proliferation other than that of supplying iron, but this has not yet been demonstrated.

1.1.3 Bacteriostasis

All microorganisms, with the single possible exception of the lactobacilli (Archibald, 1983), require iron for growth. Therefore transferrin, with its high affinity for iron, can retard microbial growth by making iron relatively unavailable. This effect is reversed if iron is in excess. A wide range of organisms show increased virulence in an iron-loaded host (Weinberg, 1978). Many microorganisms produce iron chelators capable of removing iron from transferrin in order to overcome its antimicrobial effect (Neilands, 1981), whereas others have cell surface receptors for transferrin (Schryvers and Gonzalez, 1990).

Transferrin also performs a more active antimicrobial role. It releases lipopolysaccharide from the outer membrane of Gram-negative bacteria which damages the membrane and alters its permeability (Ellison *et al.*, 1988).

1.2 IRON BINDING

Transferrin binds two ferric ions per molecule, at specific sites. It can also bind ferrous ions which are rapidly converted to Fe^{3+} in the presence of oxygen (Kojima and Bates, 1981). Transferrin can bind other di-, tri- and tetravalent metal ions, but few have been shown to be bound *in vivo*. Al^{3+} is bound (Roskams and Connor, 1990), and possibly Zn^{2+} (Harris, 1983). Apotransferrin is colourless, whereas transferrin with iron bound is red-brown.

Iron binding requires synergistic binding of an anion. Gelb and Harris (1980) produced evidence that the anion bound *in vivo* is bicarbonate, although carbonate has also been suggested. Iron binding induces a conformational change which makes the transferrin protein more compact. Kourilsky and Burtin (1968) found immunochemical differences between iron saturated transferrin and apotransferrin which suggested that iron binding causes conformational changes which hide certain antigenic sites.

Binding does not appear to be co-operative between the two sites, as apo-, mono- and diferric transferrin all occur (Wenn and Williams, 1968). However the two sites are not identical chemically, nor in their iron affinity. Under physiological conditions the affinity constants of the two sites are in the range $1\text{-}6 \times 10^{22}/M$ (Aisen and Listowsky, 1980). Harris (1977) observed differences in the metal binding properties of the two sites, and found that at physiological pH, iron is preferentially bound to the A (C-lobe) site. Zak and Aisen (1986) analysed the relative amounts of iron bound at the N- and C-lobes in monoferric transferrin and found results varied considerably between samples, but there was a weak relationship with iron saturation. The distribution of iron between the two was non-random. The relative iron-donating abilities of the two sites have been the subject of conflicting reports. This may be because the ability of each site to donate iron is influenced by environmental variables such as atmosphere and medium (Zak and Aisen, 1990). The N-lobe releases iron less easily than the C-lobe at pH 7.4, but at pH 5.6 the situation is reversed. Despite numerous studies, no clear evidence has been produced for functional differences between the two sites (reviewed by Brock, 1985).

1.3 OCCURRENCE, SYNTHESIS AND DEGRADATION

Transferrin has been found in representatives of all vertebrate classes and has also been detected in some invertebrates (Palmour and Sutton, 1971; Welch, 1990). Transferrin is primarily a protein of the blood, but it is also present in other tissues, particularly the testes and brain. Aldred *et al.* (1987) studied the levels of transferrin synthesis in different tissues of the rat by detecting levels of mRNA using Northern hybridization. The highest rates of synthesis were found in the liver and parts of the brain. Synthesis was also detected in the testes, placenta, stomach, spleen, kidney, muscle and heart. Yolk sac, small intestine and adrenal glands showed no detectable transferrin synthesis. Transferrin synthesis has also been demonstrated in bovine liver and testes (Gilmont *et al.*, 1990) and human lymphocytes (Soltys and Brody, 1970). Transferrin has been detected in the milk of some mammals, such as the

rabbit, rat (Grigor *et al.*, 1988), possum (Grigor *et al.*, 1991), echidna and platypus (Teahan and McKenzie, 1990). Mouse milk contains both transferrin and lactoferrin (Leclercq *et al.*, 1987). Transferrin, referred to as ovotransferrin, is also synthesized in the chicken oviduct, where its glycosylation is different from the serum transferrin (Williams *et al.*, 1962).

The liver is the major site of transferrin catabolism. The removal of transferrin from circulation occurs when the carbohydrate chains attached to the transferrin molecule have lost sialic acid residues, enabling transferrin to be bound by the hepatic asialoglycoprotein receptor (Young *et al.*, 1983).

1.4 STRUCTURE OF TRANSFERRIN

The molecular weights of the serum transferrins lie in the range 72,000-83,000 (Welch, 1990). The molecular weight of human transferrin has been calculated from amino acid sequence data (MacGillivray *et al.*, 1983) to be 79,570.

MacGillivray *et al.* (1983) determined the amino acid sequence of human transferrin. It consists of 679 amino acids. The sequences of the two lobes, residues 1-336 and 337-679, can be aligned by introducing gaps, revealing an internal homology of 42%. Human transferrin has 8 disulphide bridges in the N-lobe and 11 in the C-lobe. There are two N-linked glycan chains attached to residues in the C-lobe of human transferrin.

The primary structure of rabbit transferrin has been partially determined, and has 78% identity to human transferrin (Junanker *et al.*, 1990). N-terminal amino acid sequences of the serum transferrins of a few other species have been determined, including that of the pig (Baldwin *et al.*, 1990), cattle (Maeda *et al.*, 1980), horse (Chung and McKenzie, 1985), possum (Grigor *et al.*, 1991), echidna and platypus (Teahan and McKenzie, 1990).

The cDNA sequences of the serum transferrins of several species have been determined. Yang *et al.* (1984) published a human transferrin cDNA sequence of 2.3 kb, including 57 bp coding for a leader sequence, and 2037 bp coding for the mature protein. This predicts a 679 amino acid sequence which is identical to the protein-derived sequence reported by MacGillivray *et al.* (1983), with the exception of nine residues. Porcine transferrin cDNA was sequenced by Baldwin and Weinstock (1988) and found to encode a 696 residue protein with 70% identity to

human transferrin. This is not a complete sequence; it lacks the signal peptide. Banfield *et al.* (1991) reported the cDNA sequence of rabbit transferrin. Its derived protein sequence consists of 694 amino acids with 79% and 69% identity to human and pig transferrins respectively. Moskaitis *et al.* (1990) determined the transferrin cDNA sequence for the frog *Xenopus laevis*, using liver tissue. The 2279 bp cDNA encodes a protein of 717 residues, with 46% and 47% identity to human and chicken transferrins respectively.

Partial cDNA sequences of the serum transferrins of rat (Aldred *et al.*, 1984, Huggenvik *et al.*, 1987), mouse (Chen and Bissell, 1987), hamster (Duguid *et al.*, 1989), cattle (Gilmont *et al.*, 1990) and sheep (Tu *et al.*, 1991) have been determined. In addition, Idzerda *et al.* (1989) isolated clones comprising the entire mouse transferrin gene, and published sequence data for each end of the gene.

The 3-dimensional structure of rabbit transferrin has been studied by Bailey *et al.* (1988). 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. The two lobes can be superimposed by a rotation of 167° and a translation. Each lobe is made up of two domains, and the iron binding site is located within the interdomain cleft (Figure 1). The iron is bound by 2 tyrosine residues (numbers 95 and 188 in the N-lobe), 1 histidine (249) and 1 aspartic acid (63) (see Figure 2). The molecule is stabilized by 19 disulphide bridges. Greater detail of the iron binding site, particularly the residues binding the anion, were reported by Sarra *et al.* (1990) after studying the N-terminal half-molecule. The anion is bound by four residues (threonine 120, arginine 124, alanine 126, and glycine 127) which lie at one end of an α-helix.

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, remotely situated from the iron binding site and does not include iron-binding ligands. The equivalent region in human transferrin is encoded by a single exon, exon 2 (Schaeffer *et al.*, 1987), which suggests that it may have been incorporated into the transferrin gene by exon shuffling. The amino acid sequence of this region is homologous to that of the B lym transforming proteins (Section 1.7.5), 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).

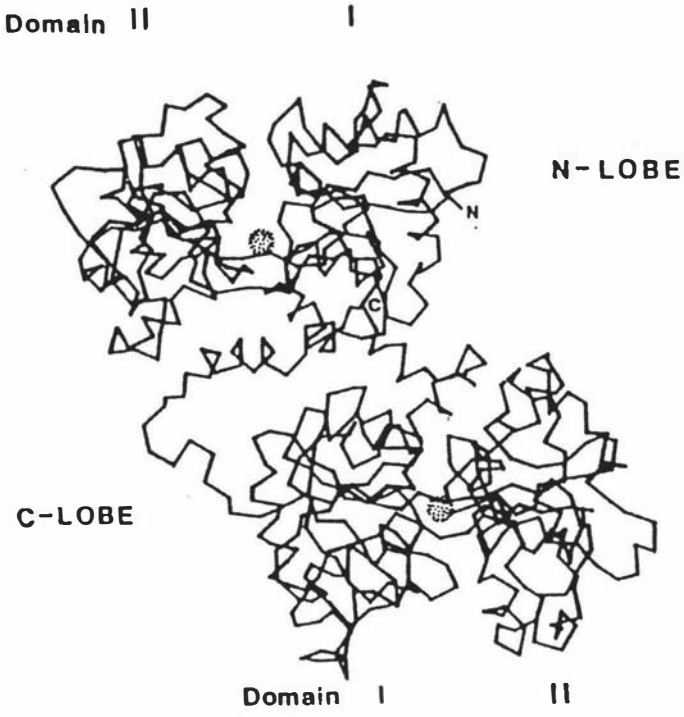


Figure 1. The structure of rabbit transferrin, represented by its α-carbon chain. The molecule consists of two structurally homologous lobes, each made up of two dissimilar domains. A cleft between the domains contains the iron-binding site. Iron atoms are represented by spheres. The N- and C-termini of the molecule are labelled. From Sarra *et al.* (1990).

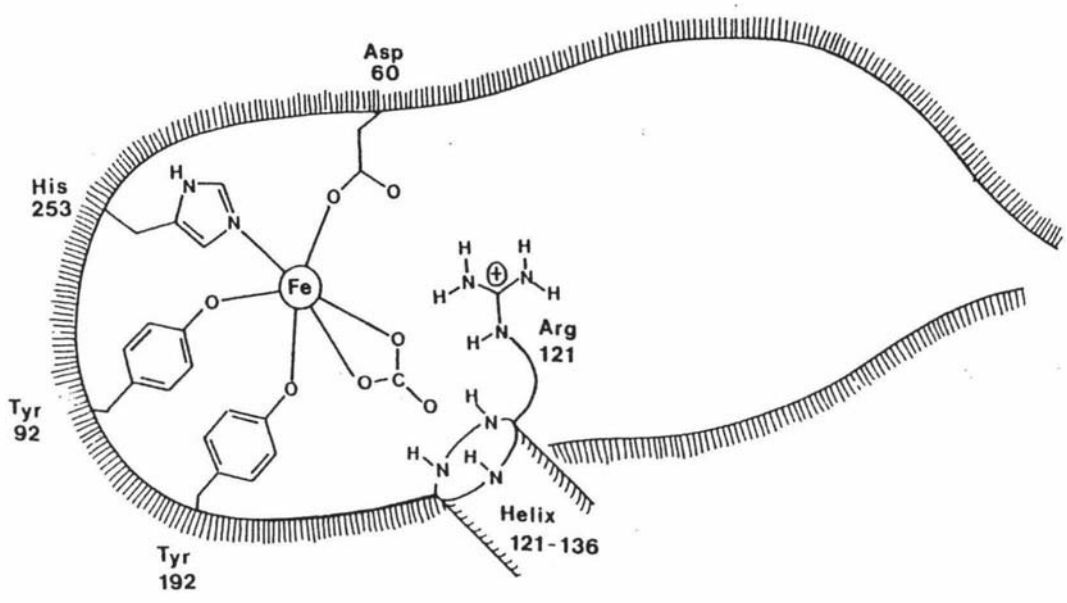


Figure 2. Schematic representation of the iron binding site in the N-lobe of human lactoferrin, from Anderson *et al.* (1989). The same arrangement is found in the C-lobe, and also in rabbit transferrin (Sarra *et al.*, 1990). The ferric ion (labelled Fe) and carbonate anion are shown.

1.5 HETEROGENEITY OF TRANSFERRIN

Electrophoresis of transferrin reveals multiple bands, indicating that transferrin is not a homogeneous substance. The structural differences causing this heterogeneity are of three main types. These are variations within the polypeptide chain, differences in glycosylation, and degrees of iron saturation.

1.5.1 Variation in the polypeptide chain

Heterogeneity occurs through genetically determined amino acid substitutions or deletions in the polypeptide chain. Polymorphism of transferrin has been detected in most species studied (Morgan, 1981). Human transferrin occurs in at least 22 allelic forms. Most of these are uncommon, as more than 90% of the population have type C transferrin (Kuhnl and Spielmann,1978). Type C is actually made up of three subtypes C_1, C_2 and C_3 (Kuhnl and Speilmann, 1979). D_{Chi} is the only other common variant. It occurs mainly in Asian and American Indian populations.

The variants of transferrin appear to differ by very minor changes to the amino acid sequence. Single amino acid substitutions have been detected among human transferrin variants. An aspartate in type C was replaced by a glycine in D_1 (Wang and Sutton, 1965) while a glutamic acid residue in B_2 was replaced by a glycine in C (Wang *et al.*, 1966). The C and D_{Chi} variants differ by the substitution of a histidine for an arginine (Wang *et al.*, 1967) at position 300 (Yang *et al.*, 1984). No other differences were detected between the pairs of variants, but the possibility of other changes being present was not eliminated. A similar study of bovine transferrin by Maeda *et al.* (1984) found three amino acid substitutions. These were glutamic acid, lysine and aspartic acid in type A substituted for aspartic acid, arginine and glycine in D_2 respectively. These differences all occurred in the C-terminal half of the molecule.

A functionally abnormal variant of human transferrin has been identified (Evans *et al.*, 1982, 1988). The C-terminal lobe of this variant cannot bind iron as securely as common variants do. This is due to a substitution of arginine for the glycine at position 394 in type C. This is close to the aspartate (position 392) which is an iron ligand. A second abnormal variant, with properties much like the first, was described by Welch and Langmead (1990). It has an isoleucine substituted for an asparagine, the position of which is uncertain.

Attempts have been made to identify functional differences among the variants of human transferrin. Wong and Saha (1986) could find no significant differences in total iron binding capacity in relation to the transferrin variants, but did find differences among the sub-types of transferrin C. The total iron binding capacity of C_1 homozygotes was significantly greater than that of C_1C_2 heterozygotes and C_2 homozygotes. Auconi *et al.* (1982) found that the C_2 sub-type was more common among infants born prematurely than in the control sample. In contrast, Wong and Saha (1991) reported that maternal and foetal transferrin phenotypes had no effect on foetal growth.

Functional studies have also been performed in other species. Kanda *et al.* (1983) found that rat transferrin A was a better donator of ferric ions to reticulocytes than type B. Several groups have found an association between transferrin phenotypes and reproductive performance. Rasmusen and Tucker (1973) found that in sheep the transferrin types of both parents and offspring influenced survival of the offspring. In cattle, transferrin heterozygosity appeared to be advantageous (Ashton, 1965). In laboratory populations of mice, transferrin B appeared to be superior to the A variant, as determined by numbers of progeny.

Another type of heterogeneity has been observed in bovine transferrin. Two electrophoretic bands were observed in a homozygous sample after removal of sialic acid. This was found to be due to a scission in the polypeptide chain of the faster-moving band. Reduction of the protein to remove disulphide bonds allowed the break to be detected (Maeda *et al.*, 1980).

1.5.2 Glycosylation

Transferrin is a glycoprotein. The carbohydrate content varies between species, and tissues within a species. Spik *et al.* (1988) studied glycosylation in the serum transferrins of a number of mammals and birds. Carbohydrate content ranged from 2.2 to 5.8%. This was made up of one or two glycan chains, with a biantennary (two-branched), triantennary or tetraantennary structure (Irie *et al.*, 1988). The glycan chains are attached by N-glycosidic linkages to specific residues in the amino acid sequence.

Human serum transferrin has two chains located in the C-terminal half of the molecule. These may be both biantennary (82%), both triantennary (1%), or one of each (17%) (Spik *et al.*, 1988). The quantities of the various types of glycan chains on transferrin found in

cerebrospinal, amniotic and synovial fluids differ from those on transferrin in the serum (van Eijk *et al.*, 1983). Increased proportions of triantennary glycans occur during pregnancy (Leger *et al.*, 1989). The functional significance of these variations is not clear.

Differences also occur in the sialic acid content of the glycan chains. The three bands of human transferrin type C are reduced to a single band by treatment with neuraminidase, an enzyme which removes sialic acid residues. This demonstrates that the bands differ in sialic acid content (Regoeczi *et al.*, 1977). Likewise, bovine transferrin has between 0 and 5 sialic acid residues per molecule (Stratil and Spooner, 1971), and rat transferrin has 2 or 3 (Schreiber *et al.*, 1979).

Glycosylation does not appear to be necessary for efficient iron transport by transferrin. Kornfeld (1968) found that removal of most of the carbohydrate from human transferrin caused no significant alteration in iron binding capacity, receptor interaction, or the ability to transfer iron to reticulocytes.

Glycosylation is important in determining the rate of elimination of transferrin from circulation. Transferrin is removed by interaction of the carbohydrate moiety with the hepatic asialoglycoprotein receptor. Removal is more rapid when the sialic acid content is low (Wong and Regoeczi, 1977).

1.5.3 Iron saturation

In physiological conditions, transferrin is 30% saturated with iron. There are four types of transferrin present: apotransferrin, monoferric transferrins with iron bound in either the N- or C-lobe, and diferric transferrin (de Jong and van Eijk, 1989).

1.6 DISORDERS OF TRANSFERRIN FUNCTION

There have been a few cases reported of humans with exceedingly low levels of transferrin in the blood. A total absence of transferrin does not occur; presumably this situation would be fatal. The transferrin deficiency cases reported by Heilmeyer (1966) and Goya *et al.* (1972) concerned children with low levels of iron in the serum, only traces of transferrin present and a

deficiency of haemoglobin. The organs, particularly the liver, contained excessive amounts of iron, except the bone marrow and spleen which had low levels. Treatment with iron had no effect, while blood transfusions were helpful, as was the administration of transferrin. Iron absorption from the gut was elevated, as is usual during anaemia. In both cases, the parents' transferrin levels were approximately half normal values but the parents were not anaemic. The disorder was inherited as an autosomal recessive. The problems of acute transferrin deficiency illustrated by these cases reveal the importance of transferrin in iron distribution and erythrocyte production. They also show that transferrin is not vital for iron absorption.

A similar disorder occurs in the Hp strain of mice. Homozygotes have less than 1% of normal transferrin levels in the serum, and die soon after birth with severe anaemia, unless treated with mouse serum or transferrin. Treated mice accumulate iron in the liver and pancreas, are anaemic, and have low fertility. The mode of inheritance is autosomal recessive (Bernstein, 1987). The disorder is due to incomplete processing of the transferrin mRNA transcript. Most of the transferrin mRNA present in brain and liver is 5 kb instead of 2.5 kb, and contains the last two introns. The gene has no gross abnormality, as determined by Southern blotting, so the disorder is probably caused by a small deletion or point mutation which disrupts splicing (Huggenvik *et al.*, 1989).

A few other abnormal transferrin variants have been observed, but as these occurred in heterozygotes the effects were minimal. A single functional transferrin allele produces a serum transferrin concentration which is half the normal value. This is sufficient for a normal phenotype. Silent transferrin alleles have been observed in horses. The carriers appear to be homozygous for their other allele when analysed by electrophoresis of blood proteins, but the offspring phenotypes fail to show the parental transferrin type (Schmid *et al.*, 1990; Bowling *et al.*, 1991). This situation may parallel the human and mouse examples above. Two abnormal human transferrin variants, which bind iron defectively in the C-lobe, are caused by single amino acid substitutions (Evans *et al.*, 1982, 1988; Welch and Langmead, 1990), (Section 1.5.1).

Several inherited anaemias in mice and humans could be due to disorders of transferrin or its receptor, but this has not been demonstrated (Morgan, 1981). Reduction in levels of transferrin and/or altered glycosylation of transferrin, are associated with several conditions, such as infection, malignancy, rheumatoid arthritis, liver cirrhosis, haemochromatosis and a paediatric neurological syndrome (Morgan, 1981; de Jong and van Eijk, 1989; Debruyne *et al.*, 1984).

1.7 RELATED PROTEINS

1.7.1 Lactoferrin

Numerous biological fluids of mammals, including milk, semen, tears, saliva, bile, nasal and genital secretions, contain the protein lactoferrin, also known as lactotransferrin (Masson *et al.*, 1966). Lactoferrin is also present in neutrophilic leukocytes (Bullen and Armstrong, 1979). The lactoferrins of milk and leukocytes differ in their glycosylation (Derisbourg *et al.*, 1990). The protein occurs largely devoid of iron although its affinity for iron is greater than that of serum transferrin (Aisen and Leibman, 1972). The molecular weights of lactoferrins range from 78,000-82,000 (Shimizaki *et al.*, 1991; Metz-Boutigue *et al.*, 1984).

The amino acid sequence of human lactoferrin was determined by Metz-Boutigue *et al.* (1984). They established a 703 amino acid sequence, with a two-fold internal homology. The sequence was 59% identical to human transferrin and 49% identical to chicken ovotransferrin. The positions of disulphide bridges were predicted, and found to be homologous with those in serum transferrin except that three less were present. One glycosylation site was present in each of the two lobes, in homologous positions.

A cDNA encoding 40% of human lactoferrin was isolated from a neutrophil library and sequenced by Rado *et al.* (1987). Its derived protein sequence was very similar to the C-terminal of the human lactoferrin amino acid sequence determined by Metz-Boutigue *et al.* (1984); only 3 residues differed. The human mammary gland lactoferrin cDNA sequence was reported by Powell and Ogden (1990), who obtained sequence encoding the complete mature protein, and 17 residues of the signal peptide. Rey *et al.* (1990) also used mammary gland tissue to produce a human lactoferrin cDNA which was almost full length. The remaining 5' sequence was obtained from a genomic DNA clone, resulting in a complete sequence of 711 amino acids including a 19 residue signal peptide.

Lactoferrins have also been isolated from the milk of the cow (Wang *et al.*, 1984), horse (Jolles *et al.*, 1984), monkey (Davidson and Lonnerdal, 1986), pig (Hutchens *et al.*, 1989) and sheep (Buchta, 1991). N-terminal amino acid sequences were determined for these lactoferrins, and comparison revealed considerable homology (Hutchens *et al.*, 1989).

Pentecost and Teng (1987) isolated and sequenced a 2224 bp cDNA for mouse lactoferrin. It encoded a protein of 688 amino acids plus 17 residues of the signal sequence. The mouse lactoferrin sequence was 70%, 56% and 49% identical to human lactoferrin, human transferrin and chicken ovotransferrin, respectively. The remaining 2 residues of the mouse lactoferrin signal sequence have been determined by Liu and Teng (1991), who isolated and sequenced the promotor region and first exon of the mouse lactoferrin gene.

Several authors have reported sequences for bovine lactoferrin. Mead and Tweedie (1990) obtained a cDNA sequence encoding most of the mature protein, and completed the protein sequence by sequencing the N-terminal of the protein. A submaxilliary gland library yielded a cDNA encoding the mature bovine lactoferrin and 16 residues of signal peptide (Pierce *et al.*, 1991). The complete bovine lactoferrin sequence was obtained by Goodman and Schanbacher (1991) from a 2351 bp cDNA from a mammary gland library, combined with direct mRNA sequencing. The translation product was 708 residues, including a 19 amino acid signal peptide. Comparison of the bovine lactoferrin sequence with other lactoferrins and transferrins revealed the following identities: human lactoferrin (68%), mouse lactoferrin (64%), pig transferrin (61%), human transferrin (60%), chicken transferrin (53%), human melanotransferrin (41%).

A cDNA sequence for pig lactoferrin has been reported (Alexander *et al.*, 1992). The 2259 bp sequence encodes a protein of 703 amino acids including a 19 residue signal peptide.

The carbohydrate chains of lactoferrins make up 6-11% of the proteins mass, a higher proportion than in serum or ovotransferrins (Spik *et al.*, 1988). The chains are biantennary, and 1, 2 or 4 may be present, depending on the species.

The three-dimensional structure of human milk lactoferrin was studied by Anderson *et al.* (1989) using X-ray crystallography. They observed two lobes, made up of the N- and C-terminals of the protein respectively, much like serum transferrin. The two lobes were each made up of two domains, with the iron binding site in the cleft between the two domains. The two lobes were connected by a short α-helix and showed similar folding patterns, with differences occurring in surface loops of the protein. Each lobe contained one iron binding site in which the iron atoms were co-ordinated by two tyrosine residues, a histidine and an aspartic acid (Figure 2). The carbonate anion was co-ordinated with the ferric ion in a bidentate mode, and was also bound by an arginine side chain and a threonine, and alanine and glycine residues

in the N-terminal of an α-helix. Part of the N1-domain of lactoferrin, comprising residues 1-47 in human lactoferrin, is responsible for its bactericidal properties (see below). The equivalent regions of serum transferrin and ovotransferrin do not appear to share this property (Bellamy *et al.*, 1992) but may be associated with the role of transferrin as a growth factor (see Section 1.4).

The major function of lactoferrin is to prevent bacterial infection. Lactoferrin protects the mammary glands from infection, particularly when not lactating, and gives breast-fed infants resistance to gastrointestinal infections (Reiter, 1978). This bacteriostasis is thought to occur through iron depletion, caused by binding of iron by lactoferrin. In piglets, *Escherichia coli* infection in the blood caused a rise in plasma lactoferrin levels and a drop in total serum iron (Gutteberg *et al.*, 1989). Lactoferrin was thought to be released from granulocytes in order to remove iron from the blood. Molloy and Winterbourn (1990) showed that when neutrophils and *E. coli* were mixed, many of the bacteria were phagocytosed, and the iron which was released as a result was bound by lactoferrin released from neutrophils. Thus iron would be removed from sites of inflammation.

Lactoferrin was shown to be bactericidal against *Streptococcus mutans* and *Vibrio cholerae* (Arnold *et al.*, 1977). This effect was not due to withholding of iron, as iron in the media was in excess of the iron-binding capacity of the lactoferrin present. The bactericidal effect of lactoferrin is enhanced by the presence of lysozyme. Suzuki *et al.* (1989) found that although apolactoferrin decreased the rate of bacterial growth, the presence of iron reduced this effect. However if lysozyme was present as well as lactoferrin, bacterial growth was strongly inhibited even with iron present. Lysozyme on its own did not produce this effect. Bacteria treated with lactoferrin and lysozyme were shown to agglutinate.

Lactoferrin may have a role in the regulation of development of cells of the immune system. Iron-saturated lactoferrin inhibits the production of granulocytes and macrophages (Broxmeyer *et al.*, 1980). The ability to synthsize lactoferrin is associated with normal granulocyte maturation. Therefore, lactoferrin may have a role in the development of mature granulocytes (Rado *et al.*, 1987).

1.7.2 Ovotransferrin

Avian egg-whites contain a protein known as ovotransferrin or conalbumin. The ovotransferrin of the chicken shares the same polypeptide chain as chicken serum transferrin, but the two proteins differ in the attached carbohydrate chains (Williams, 1962). Ovotransferrin synthesis occurs in the oviduct, (Palmiter, 1972). Chicken ovotransferrin has the bilobal structure characteristic of the transferrin family (Abola *et al.*, 1982).

A partial amino acid sequence for chicken ovotransferrin was reported by Williams *et al.* (1982a), who calculated a 37% homology between the two lobes, and revealed the positions of 15 disulphide bonds. Of the nine disulphide bonds in the C-terminal half, six are in positions homologous to the six in the N-terminal half. Jeltsch and Chambon (1982) derived the amino acid sequence of chicken ovotransferrin from the cDNA sequence, revealing a putative protein of molecular weight 77,700, made up of 705 residues including a leader sequence of 19 residues.

Chicken ovotransferrin has a single glycan chain attached to the C-terminal portion of the molecule (Jeltsch and Chambon, 1982). There is evidence for heterogeneity in either the structure of this chain or in its site of attachment to the polypeptide chain, or perhaps both (Williams, 1968; Iwase and Hotta, 1977; Dorland *et al.*, 1979). Chicken ovotransferrin is unusual in that its glycan chain does not contain sialic acid residues, so the heterogeneity in sialic acid content common in other transferrins does not occur (Dorland *et al.*, 1979).

Ovotransferrin is thought to have an antimicrobial role in protecting the developing embryo, but may also function as a nutritional store in the egg. Schade and Caroline (1944) found that chicken egg-white was bacteriostatic against a number of microorganisms, an effect which was reversed by excess iron.

1.7.3 Melanotransferrin

Melanotransferrin, or melanoma antigen p97, is a cell surface glycoprotein found in elevated numbers on most human melanomas, but is also present in trace amounts on normal adult cells. Its molecular weight is 97,000. Brown *et al.* (1982) sequenced the N-terminal and found it to be homologous with the N-terminal sequence of human serum transferrin. He also established

a functional similarity by demonstrating that melanotransferrin binds iron. The complete amino acid sequence was derived by Rose *et al.* (1986) from the 3900 bp cDNA sequence. The protein has 719 amino acids, structured in two lobes like those of the other transferrins, with an extra 25 amino acids at the C-terminal; these predominantly hydrophobic residues are thought to form a membrane anchor. Homology between human serum transferrin and the equivalent regions of melanotransferrin is 39%. The positions of the disulphide bridges are also conserved, suggesting that the overall structure is similar. The function of melanotransferrin is not known. It binds iron, but does not appear to facilitate its entry into the cell (Richardson and Baker, 1990, 1991).

1.7.4 Rat hemiferrin

Stallard *et al.* (1991) detected a 0.9 kb mRNA in the germ cells of rat testis, by hybridization to human transferrin cDNA. A corresponding cDNA was isolated and sequenced. The 216 amino acid sequence it encoded had 64% identity to the C-terminal half of human transferrin, and 42% identity to the N-terminal half.

1.7.5 B lym transforming gene proteins

Two proteins which are homologous to serum transferrin have been predicted from gene sequences. Transforming genes were isolated from the tissue of chicken B cell lymphomas (Goubin *et al.*, 1983) and human Burkitt's lymphoma (Diamond *et al.*, 1984). The chicken and human B lym transforming genes encoded proteins of 65 and 58 residues, respectively, and the two protein sequences were 33% identical. Both genes consisted of two exons. The amino acid sequence derived from the second exon of the chicken gene had significant homology to the sequence derived from the second exon of the human transferrin. Sequences homologous to the B lym gene occur in normal human and chicken DNA. The structural relationship of the B lym proteins with the transferrins may be paralleled by a functional relationship with respect to the role of serum transferrin in cell proliferation.

1.7.6 Invertebrate transferrins

An iron-binding glycoprotein of molecular weight 80,000 occurs in the haemolymph of the tobacco hornworm (insect), *Manduca sexta.* It binds a single iron ion per molecule (Huebers *et al.*, 1988). A cDNA isolated from a library prepared from the larval fat body was sequenced (Bartfeld and Law, 1990). The 2183 bp sequence encoded a mature protein of 633 residues and a signal peptide of 18. It has the duplicated structure characteristic of the transferrins, and is 28% identical to human transferrin and lactoferrin, and 26% identical to chicken transferrin and human melanotransferrin. The level of similarity to human transferrin was shown to be significant.

A transferrin-like protein is present in the ascidian *Pyura stolonifera.* It has a molecular weight of 41,000 and binds one ferric ion (Martin *et al.*, 1984). Hybridization studies indicated that the *Pyura* transferrin DNA sequence was homologous to the human transferrin gene (Yang *et al.*, 1985). Iron-binding proteins have also been found in a crab (Heubers *et al.*, 1982) and a spider (Lee *et al.*, 1978b) but their relationships to the transferrins are uncertain.

1.8 TRANSFERRIN GENES

1.8.1 Human transferrin

The human transferrin gene, not including flanking regions, has a total size of 33.5 kb, and is made up of 17 exons separated by 16 introns (Schaeffer *et al.*, 1987). The internal homology between the two domains seen in the protein structure is reflected in the intron/exon pattern of the gene. The exon sizes in the 5' half of the gene correspond to those in the 3' half. The intron sizes are more variable and do not show the internal homology.

Sequence data for the human transferrin gene has been reported by several authors. Park *et al.* (1985) and Schaeffer *et al.* (1987) between them reported the sequences of all the exons, including the intron/exon boundary sequences.

Schaeffer *et al.* (1987) also isolated and sequenced part of a human transferrin pseudogene. The non-processed pseudogene showed 72% nucleotide identity with the true transferrin gene, but had extra stop codons present.

The sequence of the 5' region of the human transferrin gene has been examined for possible regulatory sequences. Adrian *et al.* (1986) sequenced 3.6 kb of the 5' flanking region, including exons 1 and 2. They located possible regulatory elements homologous to the metallothionein gene heavy metal receptor element, the glucocorticoid receptor element and the putative acute phase reaction signal. They also found 14 bp sequences like those occurring 5' to the interleukin 2 and gamma interferon genes. All three of these genes are expressed in T-lymphocytes. The sequence of the 5' untranslated region of the human transferrin mRNA revealed a probable stem-loop structure and a binding site for the 18S rRNA. These features have also been observed in the sequence of chicken transferrin (Cochet *et al.*, 1979a).

In a similar study, Lucero *et al.* (1986) sequenced 620 bp in the human transferrin 5' region, which included exon 1. Several putative regulatory elements were located: 5 metal binding regulatory elements, 2 Sp1 transcription factor binding sites, a progesterone receptor element, a glucocorticoid regulatory sequence, a cAMP regulatory domain and 4 *cis*-acting DNA enhancers.

Schaeffer *et al.* (1985) detected polymorphisms in the transferrin gene by Southern blot analysis. They found that *Eco*RI fragments of transferrin type D differed from those of types B, C_1, C_2 and C_3. Similarly, Yang at al (1985) identified polymorphic fragments using *Eco*RI and *Pvu*II digests of human genomic DNA.

1.8.2 Chicken transferrin/ovotransferrin

Chicken serum transferrin and ovotransferrin appear to be two different glycoproteins produced by a single gene. The two proteins differ only in their carbohydrate chains. Analysis of amino acid composition and peptide patterns showed that the polypeptide components were identical (Williams, 1962). Thibodeau *et al.* (1978) determined the N-terminal amino acid sequences of the two preproteins and found that ovotransferrin and serum transferrin have identical precursors. Finally, Lee *et al.* (1978a) demonstrated that transferrin mRNA's from liver and oviduct, when hybridized to ovotransferrin cDNA, produced identical hybridization and melting curves. They concluded that chicken serum transferrin and ovotransferrin are products of the same gene, under different regulation in the liver and oviduct respectively.

A cDNA clone for chicken ovotransferrin was isolated from a chicken oviduct library by Cochet *et al.* (1979b). The cDNA was 2350 bp in size, and represented an almost complete mRNA sequence. Its identity was confirmed by sequencing the 5' end. Perrin *et al.* (1979) used this clone to isolate parts of the chicken ovotransferrin gene from a genomic library. They then used electron microscopy of mRNA-DNA hybrids to establish intron/exon boundaries, and found that the first 940 bp of the mRNA exists in the gene as 7 exons. Cochet *et al.* (1979a) continued this approach and discovered that the complete gene contains 17 exons, of 60-200 bp each. The 5' region (-260 to +200) was sequenced and found to contain two regions complementary to the 18S rRNA sequence, a Hogness box and a region which would form a base-paired loop structure in the mRNA. These features may be involved in regulation of gene expression.

Jeltsch *et al.* (1987) published the sequence of the whole ovotransferrin gene. The 10.6 kb sequence includes 17 exons and 16 introns, a TATA box (31 bp upstream of the start of exon 1) and a polyadenylation signal (at position 10549-10555). Repetitive sequences were located in introns B and C, but not in the corresponding introns in the C-terminal half of the gene, indicating that they originated after the intragenic duplication which produced the transferrins.

The human and chicken transferrin genes were compared by Schaeffer *et al.* (1987). Both have 17 exons, the sizes of which are conserved, producing proteins of a similar size. The introns however differ widely in size between the two genes, which results in a considerable difference in the total gene size: 33.5 kb for human and 10.6 kb for chicken.

1.8.3 Mouse lactoferrin

Fragments of the mouse lactoferrin gene have been isolated and 11 kb of sequence determined (Shirsat *et al.*, 1992). This includes the 5' flanking region and the first 8 exons and introns, and additional fragments from the remainder of the gene. The sequence indicates an intron-exon pattern similar to that of human and chicken serum transferrins, and a total size between that of the chicken and human transferrin genes. The 5' flanking region contains several regulatory elements which are also present in the human transferrin gene, and also a regulatory element which occurs in the myeloperoxidase gene. Both the myeloperoxidase and lactoferrin genes are expressed in neutrophilic granulocytes.

1.8.4 Chromosomal location of transferrin family genes

The serum transferrin gene, and several related genes, are located on human chromosome 3. Le Beau *et al.* (1985, 1986) located the human serum transferrin gene at 3q21-23, the human melanotransferrin gene at 3q28-29, and the transferrin receptor gene at 3q26. The human lactoferrin gene at 3q21-23 is situated very close to the serum transferrin gene (McCombs *et al.*, 1988). A transferrin pseudogene isolated by Schaeffer *et al.* (1987) is also located on human chromosome 3. Another gene found on human chromosome 3 encodes ceruloplasmin (Bowman *et al.*, 1985), which is an oxidase converting ferrous to ferric ions, indicating a possible location-function relationship with the transferrin gene family.

The transferrin family genes in rodents occur in similar syntenic groups to the human genes. Baranov *et al.* (1987) used *in situ* hybridization to locate the serum transferrin and ceruloplasmin genes on rat and mouse chromosomes. The rat chromosome 7 has the serum transferrin gene at 7q31-34, and ceruloplasmin at 7q11-13. On mouse chromosome 9 the serum transferrin gene is at 9F1-3, and ceruloplasmin at 9D. Teng *et al.* (1987) found that the mouse lactoferrin gene is also on chromosome 9.

1.8.5 Transferrin gene regulation

Transferrin synthesis in the liver is increased by iron deficiency. Chickens (McKnight *et al.*, 1980) and rats (Idzerda *et al.*, 1986) fed iron deficient diets had increased levels of transferrin in the serum, and increased transferrin mRNA in the liver. Transferrin mRNA levels in the spleen and kidney were not altered by iron deficiency. Iron overload in cultured rat hepatocytes depresses transferrin synthesis (Lescoat *et al.*, 1989).

Transferrin gene expression is regulated by a variety of hormones, with the responses differing among tissues. The chicken transferrin gene is regulated by estrogen and progesterone. Estrogen increased transferrin and transferrin mRNA synthesis in the oviduct and liver, while progesterone only stimulated transferrin synthesis in the oviduct (Lee *et al.*, 1978a). Increased transferrin synthesis in the oviduct was due to both an increase in transferrin mRNA transcription, and an increase in the stability of the transcript (McKnight and Palmiter, 1979). The Sertoli cells of rat testis also responded to hormones (testosterone, follicle stimulating hormone, insulin and retinol) by increasing synthesis of transferrin mRNA (Huggenvik *et al.*,

1987). This response was specific to the transferrin gene. Lactoferrin gene expression is also regulated by hormones. Pentecost and Teng (1987) showed that the major estrogen-inducible protein in mouse uterus was lactoferrin.

Regulation of the human transferrin gene by iron and hormones is thought to be mediated via specific sequences found in the 5' flanking region of human transferrin (Lucero *et al.*, 1986; Adrian *et al.*, 1986). These include elements relating to regulation by progesterone, glucocorticoids and cAMP, as well as metal responsive elements similar to those found in the metallothionein gene. Similarly, the mouse transferrin gene contains a growth hormone responsive element (Idzerda *et al.*, 1989), and the mouse lactoferrin gene 5' flanking region contains an estrogen responsive element (Liu and Teng, 1991).

Transferrin gene expression is specific to certain tissues. The 5' flanking region of the chicken transferrin gene stimulated expression of a reporter gene in oviduct and hepatocyte cells, but not fibroblasts or kidney cells (Dierich *et al.*, 1987), and caused high liver specific expression in transgenic mice (McKnight *et al.*, 1983). The region of the human transferrin gene from -620 to -45 conferred gene expression in hepatic cell lines, but not non-hepatic cells (Brunel *et al.*, 1988). Further studies found that only the region -76 to -51 and the TATA box were required for liver specific expression (Mendelzon *et al.*, 1990). This region includes a 13 bp element similar to HP1, the hepatocyte-specific promoter element in the *Xenopus* albumin gene (Kugler *et al.*, 1988). In conflict with the reports mentioned above, Adrian *et al.* (1990) reported that the region -622 to -152 of the human transferrin gene gave liver-specific expression. The mouse transferrin gene has a liver-specific regulatory element in the region -139 to +50 (Idzerda *et al.*, 1989).

Transferrin gene expression is specific to particular developmental stages as well as to particular tissues. Levin *et al.* (1984) studied the expression of the transferrin gene in several tissues of rats of different ages, using Northern blots. Levels of transferrin mRNA in the liver were low in the foetus and stabilized at a high level after birth. Brain tissues contained little transferrin mRNA before birth, but levels increased as the animal aged. In muscle and other non-hepatic, non-nervous tissues, the mRNA levels were high in the foetus, dropping to very low levels after birth and in the adult. The high levels in foetal tissues generally may be related to the role of transferrin as a proliferative agent. A study of transferrin mRNA levels in mouse tissues produced similar results (Yang *et al.*, 1990). Transferrin mRNA was detected in liver, brain, macrophages, placenta, uterus and ovaries of adult mice. In full term foetuses, the lungs,

heart, stomach and kidneys contained transferrin mRNA. The levels of transferrin mRNA in placenta and foetal liver increased throughout pregnancy, and the level in the liver was high in young mice, lower in adulthood, then showed an increase during aging.

Transferrin mRNA levels in the rat mammary gland follow a biphasic pattern of expression, increasing until birth, then decreasing up to 10 days of lactation and rising again towards the end of lactation (Grigor *et al.*, 1990).

1.9 EVOLUTION OF THE TRANSFERRIN FAMILY

The amino acid sequences of human serum transferrin (MacGillivray *et al.*, 1982), chicken ovotransferrin (Williams *et al.*, 1982a) and human lactoferrin (Metz-Boutigue *et al.*, 1984) each revealed a two-fold internal homology. This led to the proposal that the transferrin family had originated through the duplication of a gene encoding a protein of approximately 340 amino acids, with a single iron binding site (MacGillivray *et al.*, 1982). MacGillivray *et al.* (1983) also detected a weaker four-fold homology in the human transferrin sequence. This was supported by Jeltsch and Chambon (1982) who demonstrated that the chicken ovotransferrin sequence could be arranged in four blocks with 44% homology. However Mazurier *et al.* (1983) suggested a six-fold homology based on a partial amino acid sequence for human lactoferrin.

A comparison of the known and predicted disulphide bonds of human transferrin, human lactoferrin and chicken ovotransferrin showed six disulphide bonds to be conserved in both domains of the three proteins (Metz-Boutigue *et al.*, 1984). This is further evidence for the duplication theory.

Park *et al.* (1985) proposed a model for the duplication of an ancestral gene, in which a 10-exon gene is duplicated by an unequal cross-over between the first and last introns, followed by the loss of exon 4 in the 5' half (Figure 3). This model is supported by Schaeffer *et al.* (1987) who compared the sizes of exons in the two halves of the human and chicken transferrin genes. The intron sizes do not show the internal homology in either human or chicken transferrins, indicating post-duplication modification.

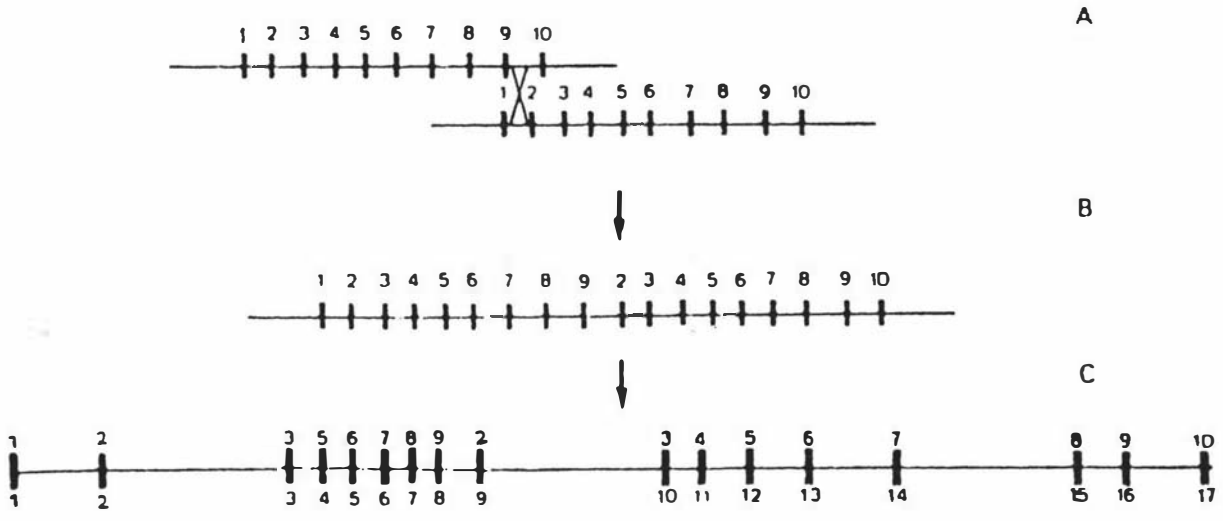


Figure 3. A scheme for the origin of transferrin genes, using the human transferrin gene as a model. The common ancestor of transferrin (with ten exons) was duplicated by an intragenic crossing-over (A), producing a duplicated ancestor which had lost one of each of the first and last exons (B). Subsequent loss of the 5' exon 4 and changes to the introns produced the modern 17-exon transferrin genes, including the human transferrin gene (C). Exons are represented by bars and are not to scale. Intron sizes are arbitrary in (A) and (B), but are to scale in (C). In (C), the upper numbers correspond to the numbers in (A) and (B) and the lower numbers correspond to the exons of the present-day human transferrin gene. From Bowman *et al.* (1988).

Attempts have been made to characterize the half-sized precursor of the transferrins. An iron binding protein of a suitable size (41,000 Da) has been detected in the ascidian *Pyura* (Martin *et al.*, 1984). It bound only one ferric ion per molecule, and was able to donate iron to rat reticulocytes. Yang *et al.* (1985) showed by Southern hybridization that *Pyura* DNA contained sequences homologous to human transferrin. This transferrin-like gene was expressed in the eggs and the digestive gland, which is analogous to the vertebrate liver. Another single iron-sited transferrin has been found in the germ cells of rat testis, referred to as hemiferrin (Stallard *et al.*, 1991).

Williams *et al.* (1982b) demonstrated that transferrin half-molecules were rapidly lost from the blood via the kidneys. They concluded that the ancestral transferrin was either not a serum protein, or was not as small as a single domain of transferrin. A third possibility is that the ancestral transferrin did not occur in an organism with kidneys.

The discovery and sequencing of a transferrin in an insect, the tobacco hornworm *Manduca sexta* is an interesting addition to this story. The *M. sexta* transferrin is internally duplicated, but only binds one ferric ion, probably in the N-terminal domain (Bartfeld and Law, 1990, Huebers *et al.*, 1988). Unless the *M. sexta* transferrin originated through an independent intragenic duplication, the origin of the two-sited transferrins occurred much earlier than previously thought, and the *Pyura* transferrin would be a derived rather than ancestral type.

A hypothetical scheme for the evolution of the genes encoding the transferrin family was proposed by Yang *et al.* (1985) and elaborated by Bowman *et al.*, (1987). According to this proposal, the transferrins originated via an intragenic duplication. Subsequent gene duplication events gave rise to melanotransferrin and lactoferrin. Figure 4 shows a modified representation of this model. The model illustrated by Bowman *et al.* (1987) indicated a gene duplication at the divergence of birds and mammals instead of at the divergence of melanotransferrin from serum transferrins and lactoferrins.

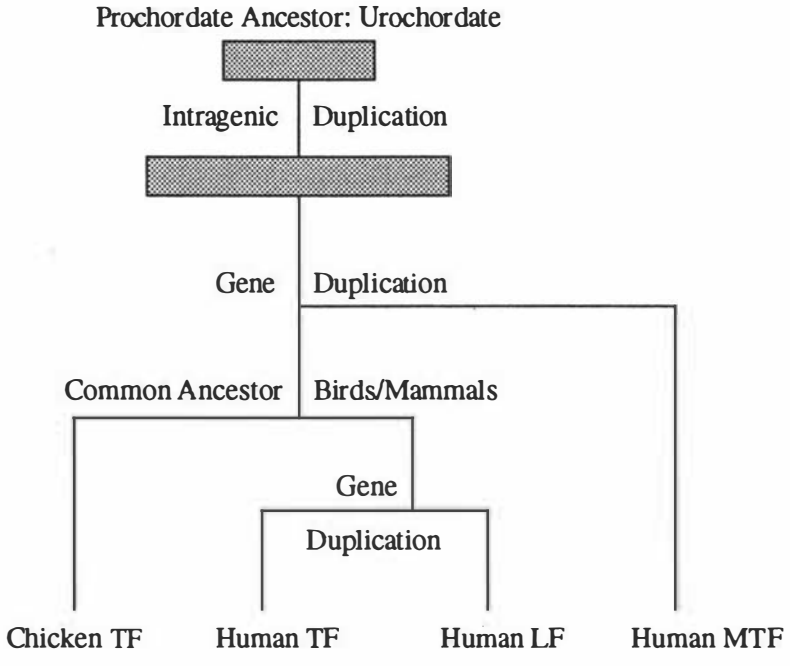


Figure 4. Evolutionary scheme for genes encoding the transferrin family. The transferrins are thought to have arisen via an intragenic duplication. Subsequent gene duplications resulted in serum transferrin, lactoferrin and melanotransferrin. TF=transferrin. LF=lactoferrin. MTF=melanotransferrin. (Modified from Bowman *et al.*, 1987).

1.10 HORSE TRANSFERRIN

The molecular weight of horse transferrin has been determined by several authors. Hudson *et al.* (1973) reported a molecular weight of 79,100, determined by sedimentation equilibrium centrifugation. They also reported a similar value for the reduced protein, indicating a single polypeptide chain. Using the same method, Stratil *et al.* (1984) identified two components of horse transferrin, with molecular weights of 75,200 and 80,500. These were thought to differ in the number of attached glycan chains. Penhallow *et al.* (1991) also detected two components of horse transferrin, with molecular weights of 78,000 and 80,000, established by SDS-polyacrylamide gel electrophoresis.

In horses, as in other mammals, transferrin occurs in a range of variant forms (Figure 5). Fourteen variants of horse transferrin, designated D, D_2, D_3, F_1, F_2, F_3, G, H_1, H_2, J, M, O, R, X, are recognized, according to Bell *et al.*(1988)and Cothran *et al.*, (1991). The D, F and H subtypes can be difficult to distinguish, and some authors treat each group as a single type. In addition, silent alleles have been reported (Schmid *et al.*, 1990; Bowling *et al.*, 1991). The variants show autosomal co-dominant inheritance.

Several groups have studied the gene frequencies of the variants in different breeds of horses (Braend, 1964; Gahne, 1966; Kaminski and Urbanska-Nicolas, 1979; Stratil and Glasnak, 1981; Bowling and Clark, 1985; Bell *et al.*, 1988). Not all the variants occur in all breeds of horses. For example, in thoroughbreds, only six variants normally occur: D, F_1, F_2, H_2, O and R (Bell *et al.*, 1988) (Figure 5). The transferrin system is useful for parentage testing, in conjunction with other blood protein systems (Gahne, 1966, Bowling and Clark, 1985).

A new variant of horse transferrin (designated *) which is likely to become widespread in the thoroughbred population, was recently reported (Farndale *et al.*, 1991). It was discovered in a commercially valuable stallion which is being used extensively for breeding. The * variant does not occur in either of the stallion's parents, and must have been derived from D or F_1 variants. It has been inherited by approximately half the stallion's offspring.

Each variant appears as a major and a minor band by starch or polyacrylamide gel electrophoresis, except the D_2 (or C) variant, which has two bands of equal intensity (Scott, 1980, Stratil and Glasnak, 1981). Several weaker bands can also be detected after electrophoresis of purified transferrin variants (Stratil and Glasnak, 1981).

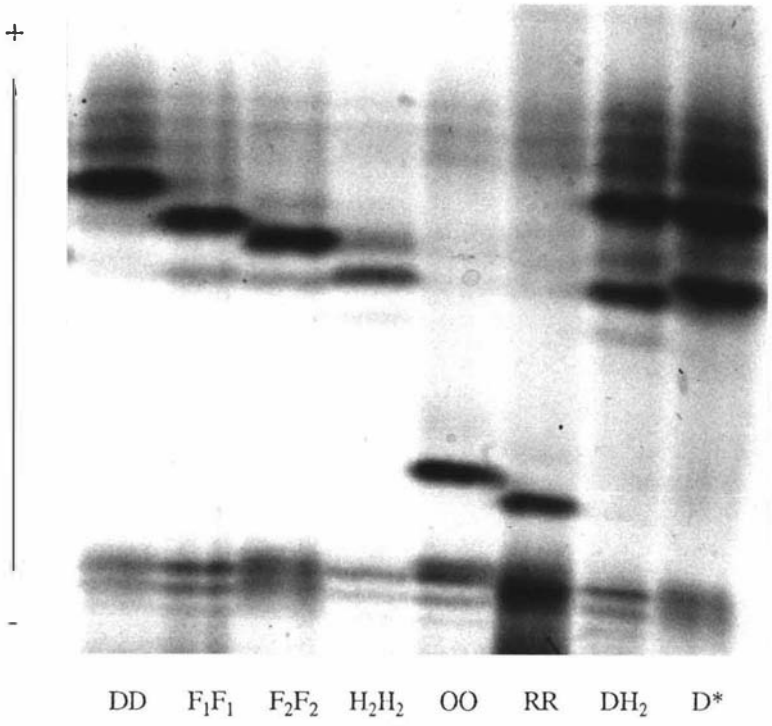


Figure 5. Variants of horse transferrin as revealed by horizontal polyacrylamide gel electrophoresis of plasma samples. Transferrin phenotypes are indicated below the photograph. The variants shown are those typically found in thoroughbreds and the new variant *. Each variant has a major and a minor band. The minor bands of the O and R variants are partially obscured by the haptoglobins. Note the substantial difference in electrophoretic mobility between the D, F_1, F_2, H_2 and * group of variants and the O and R group. Electrophoresis was carried out in a 10% polyacrylamide gel at pH 7.9, and the gel was stained with Coomassie Blue G-250. Photograph courtesy of Bruce Farndale, EBTRC.

The heterogeneity of the variants with respect to their carbohydrate chains was analysed by Stratil and Glasnak (1981). They isolated the variants C (D_2), D and O, and treated them with neuraminidase to remove sialic acid residues from the carbohydrate chains. In all three variants, either two or four sialic acid residues were present, resulting in the major and minor bands observed. Stratil *et al.* (1984) further studied type O, and determined that the major and minor bands have the same amino acid composition, but differ in their carbohydrate composition. They suggested that this difference was due to the presence of either one or two attached carbohydrate chains. This was verified by Coddeville *et al.* (1989), who also showed that the structure of the glycan chains varies, accounting for the additional weak bands.

The differences between the variants of horse transferrin have been investigated by several groups. Didkowski *et al.* (1984) established that the variants have a common molecular weight. The amino acid composition of variants analysed by Didkowski *et al.* (1984), Chung and McKenzie (1985) and Baer and Schwendimann (1972) were found to differ only slightly. Likewise peptide maps revealed a few differences (Chung and McKenzie, 1985; Baer and Schwendimann, 1972). Chung and McKenzie (1985) determined the amino acid sequences of the two peptide fragments which are variable between types D and R, and found that the two variants differ by two amino acid substitutions: aspartic and glutamic acids in type D were both replaced by glycine in type R. They also determined the N-terminal sequence (10 residues) which shows clear homology with human, bovine and rabbit transferrin sequences.

Genetic variants usually display similar antigenic properties. However immunological studies by Baer and Schwendimann (1972) revealed that types D and F are immunologically similar, while type R is considerably different. In a more extensive analysis, Kaminski *et al.* (1981) used immunological evidence to divide the horse transferrin variants into two subgroups: D, F, H, and J, the faster migrating variants, act as homologous antigens, and the slower migrating M, O and R are likewise immunologically similar to each other. However the two groups behave as heterologous antigens, as shown by spur formation during immunoelectrophoresis.

Didkowski *et al.* (1984) studied the level of transferrin in the blood in relation to the phenotype, using immunoelectrophoresis. The faster migrating types D, F and H were present in the serum at higher levels than the slower migrating types O and R. This supports the concept of two subgroups proposed by Kaminski *et al.* (1981).

Table 1

Features Distinguishing the Two Groups of

Horse Transferrin Variants

D group	R group	Feature	Reference
D, F	R	antigenic properties	Baer & Schwendimann (1972)
D, F, H, J	M, O, R	antigenic properties	Kaminski *et al.* (1981)
D, F, H, J	M, O, R	electrophoretic mobility	Kaminski *et al.* (1981)
D, F, H	O, R	serum levels	Didkowski *et al.* (1984)
D	R	peptide map & sequence	Chung & McKenzie (1985)
D, F_1, F_2, G, H_2	O, R	restriction fragment length polymorphism	Bailey *et al.* (1991)

Further evidence for two groups of horse transferrin was provided by DNA studies. Bailey *et al.* (1991) detected polymorphism in the horse transferrin gene by probing Southern blots of horse genomic DNA with a 700 bp fragment of the human transferrin cDNA. DNA cut with *Msp*I gave a 2.4 kb band when D, F_1, F_2, G or H_2 alleles were present, and a 0.5 kb band for O or R alleles. The features distinguishing the two groups of horse transferrin variants are summarized in Table 1.

Attempts have been made to correlate transferrin phenotypes with traits of horses. Osterhoff *et al.* (1974) compared transferrin types with racing performance in thoroughbreds, and found an excess of FF phenotypes among race winners. In a similar study of Swedish trotters, Andersson *et al.* (1987) did not find any relationship between transferrin types and racing performance. An association between transferrin heterozygosity and endometrial health, an important component of fertility in mares, has been reported by Weitkamp *et al.* (1991).

1.11 AIMS OF THIS STUDY

The horse has a serum transferrin which appears to be much like those of other vertebrates, but its primary structure has not been reported. Some work has been done on the differences between variants of horse transferrin but the molecular basis of these is not known. Therefore the first aim of this study is to determine the primary structure of horse transferrin. This is achieved by creating horse transferrin cDNA clones, sequencing them and deriving the amino acid sequence from the cDNA sequence. The primary structure is analysed by comparing it to other transferrins to identify differences and similarities which could be related to function. The evolution of the transferrins is also considered.

The second part of the study is the analysis of the natural variants of horse transferrin at a molecular level. Two differences between the D and R variants, identified by Chung and McKenzie (1985) by protein sequencing, are confirmed by DNA sequencing. This analysis is extended by sequencing the equivalent regions of five other alleles. Other differences between the alleles are located, and characterised by DNA sequencing. The horse transferrin intraspecies variation is compared to the variation between species, and between different members of the transferrin family. The natural variation of horse transferrin is discussed in terms of structure, function and evolution.

CHAPTER 2: METHODS AND MATERIALS

2.1 CONSTRUCTION OF A HORSE LIVER cDNA LIBRARY

2.1.1 Tissue collection and storage

The liver was removed from a freshly killed domestic horse, cut into cubes of approximately 1 cm^3, and immediately frozen in liquid nitrogen. The samples were transferred to plastic containers and stored at -80°C.

A 10 ml blood sample was collected before the horse was killed. The sample was analysed by starch gel electrophoresis at the Equine Blood Typing and Research Centre (EBTRC), Massey University, to determine which transferrin variants the animal expressed.

2.1.2 Isolation of RNA from horse liver

The methods for preparation and electrophoresis of RNA and isolation of poly(A)$^+$ RNA were obtained as a personal communication from Dr. John Tweedie, Massey University. For working with RNA, most solutions and plasticware were treated with 0.1% diethylpyrocarbonate (DEPC), and glassware was baked, to avoid contamination with RNases.

Samples of frozen horse liver (approximately 1 g) were transferred to liquid nitrogen, then, still frozen, reduced to a powder using a hammer and metal block, both of which had been chilled in liquid nitrogen. The powder was transferred into 10 ml of guanidine isothiocyanate/sarkosyl solution (see Section 2.6.1) in plastic centrifuge tubes, and homogenized for 30 s with an Ultraturrax homogenizer, on maximum speed. The homogenate was then centrifuged for 10 min at 12,000*g* and 10°C to pellet cellular debris.

The supernatant was layered on top of 4 ml 5.7 M CsCl/100 mM EDTA solution, in ultracentrifuge tubes which had been treated for 30 min in 0.1% DEPC. After centrifugation for 25 h at 33,000 rpm and 15°C in a Beckman ultracentrifuge (SW41 rotor), the solution was aspirated to within 1 cm from the bottom, and the remainder poured off. With the tube still

inverted, the bottom 1 cm of the tube, containing the RNA pellet, was cut off. The RNA pellet was resuspended in guanidine isothiocyanate buffer (400 µl) by pipetting, added to 4 ml guanidine isothiocyanate/sarkosyl, then centrifuged through 4 ml CsCl/EDTA as before. The pellet was again resuspended in guanidine isothiocyanate buffer, and precipitated by the addition of 0.1 volume of 2 M potassium acetate and 2.5 volumes of absolute ethanol, overnight at -20°C. The RNA was pelleted by centrifugation at 1700*g* and -10°C for 30 min, and the supernatant poured off. The pellet was washed twice with 70% ethanol/30 mM NaCl, centrifuged as above after washing, and the tube drained at 4°C for 30 min.

The pellet was resuspended in 1 ml 1% sarkosyl/20 mM EDTA pH 7.0, and the solution extracted with phenol/chloroform. The organic phase was back extracted with sarkosyl/EDTA and the pooled aqueous phases were extracted once with phenol and twice with chloroform. Each extraction involved mixing by inversion for 5 min and centrifugation at 200*g* and 5°C for 5 min. The final aqueous phase was precipitated in 0.2 M potassium acetate and 70% ethanol at -20°C overnight. Centrifugation at 2000*g* for 30 min produced a pellet which was washed with 70% ethanol/30 mM NaCl, centrifuged again at 2000*g* for 30 min, drained for 30 min at 4°C and resuspended in 200 µl 10 mM HEPES/1 mM EDTA pH 7.6.

Purity and concentration of the RNA solution were estimated by measuring the absorbance at 260 and 280 nm of a 1/200 dilution (Section 2.5.1). The RNA was stored at -80°C.

2.1.3 Preparation of poly(A)$^+$ RNA

Total RNA in HEPES/EDTA solution was precipitated in 0.2 M potassium acetate and 70% ethanol at -20°C, pelleted by centrifugation in a microcentrifuge and resuspended in 500 µl of sterile distilled water. The solution was heated to 65°C for 5 min to disrupt secondary structure, added to 500 µl 2x loading buffer (Section 2.6.2), cooled to room temperature and applied to an oligo(dT) cellulose column.

The column was prepared by suspending 0.2 g of oligo(dT) cellulose in 1x loading buffer. This suspension was pipetted into a sterile 1 ml syringe plugged with glass wool, and washed through with loading buffer giving a volume close to 1 ml. The column was washed with 3 ml sterile distilled water, 3 ml 0.1 M NaOH/5 mM EDTA, then 3 ml sterile distilled water and the pH of the eluate checked to see that it was less than 8.

After washing the column with 5 ml loading buffer, the RNA solution was applied. The eluate was collected, heated to 65°C for 5 min again, cooled and reapplied to the column. Fractions (1 ml) were collected from the column while 7 ml loading buffer, 4 ml washing buffer and 3 ml elution buffer (Section 2.6.2) were applied sequentially.

To determine which fractions contained RNA, 50 µl aliquots were taken from each fraction and added to 500 µl HEPES/EDTA. The absorbance of each diluted aliquot was measured at 260 nm.

Fractions containing RNA were precipitated in 0.2 M sodium acetate and 70% ethanol at -20°C, pelleted by centrifugation at 2000*g* for 30 min and resuspended in 50 µl HEPES/EDTA. The concentration was determined by measuring the absorbance at 260 nm.

2.1.4 Northern blotting

Agarose gel electrophoresis and Northern hybridization were used to check the quality of RNA preparations and to determine the size of horse liver mRNA that would hybridize to the human transferrin cDNA.

Total, poly(A)$^+$ and poly(A)$^-$ RNA were electrophoresed in a 1.2% agarose gel which contained 1x MOPS buffer (see Section 2.6.1) and 3% formaldehyde. The electrophoresis buffer was 1x MOPS. The RNA samples (3µg poly(A)$^+$, 10 µg total and poly(A)$^-$) were dissolved in a sample buffer containing 50% deionized formamide, 16% formaldehyde and 1x MOPS. The RNA in sample buffer was heated to 60°C for 10 min and cooled on ice. To it were added 1 µl ethidium bromide (10 mg/ml) and 1 µl RNA loading buffer (50% glycerol/0.4% bromophenol blue/1 mM EDTA). The sample was then pipetted into a well and electrophoresed at 80 V for 3-4 hours, until the dye front had migrated approximately 2/3 the length of the gel. The gel was then rinsed in deionized water and photographed in ultra-violet light.

The RNA was transferred to nitrocellulose membrane by capillary transfer as described for Southern blotting in Section 2.5.9, except for the following differences. The gel received no pretreatment and the nitrocellulose was prewetted in deionized water. The transfer solution was 20x SSC (Section 2.6.4) and the transfer was run overnight. After disassembling the transfer

system, the nitrocellulose was allowed to dry in air for 1 h and then baked at 80°C under vacuum for 1 h (Sambrook *et al.*, 1989).

A strip carrying one lane of total RNA was cut from the nitrocellulose membrane and stained with methylene blue to show the rRNA bands for use as size markers. The staining involved soaking the membrane in 5% acetic acid for 15 min, 0.5 M sodium acetate/0.04% methylene blue for 10 min and washing with running water for 10 min (Sambrook *et al.*, 1989).

The remainder of the nitrocellulose membrane was prehybridized in 50% formamide, 5x Denhardt's solution, 0.5% sodium dodecyl sulphate (SDS), 5x SSPE and 100 µg/ml sheared denatured herring sperm DNA (Section 2.6.4), using 10 ml of solution /100 cm^2 of membrane. Prehybridization was at 42°C for 4 h, after which the probe was added. This was human transferrin cDNA labelled with [α-^{32}P]dCTP (Section 2.5.11), which was boiled for 3 min, cooled in ice and added to the hybridization solution to a concentration of 2 ng/ml. Hybridization proceeded overnight at 42°C. The nitrocellulose was then removed from the hybridization solution and washed twice in 2x SSC/0.1% SDS at room temperature for 10 min, twice in 0.1x SSC/0.1% SDS at room temperature for 10 min and finally once in 0.1x SSC/0.1% SDS at 42°C for 1 h. Hybridization of the probe to RNA was detected by autoradiography.

2.1.5 cDNA synthesis

cDNA was synthesized using an Amersham cDNA synthesis system kit and following the instructions supplied with the kit. Initially a pilot experiment was carried out using [α-^{32}P]dCTP labelling to determine the efficiency of synthesis. The first strand synthesis was analysed using a reaction containing 1 µg of mRNA and 10 µCi of [α-^{32}P]dCTP. For the second strand synthesis, a similar first strand reaction was done, omitting the labelled dCTP, and 10 µCi of [α-^{32}P]dCTP was instead included in the second strand synthesis reaction. The incorporation of labelled nucleotide was determined by spotting 1 µl of a 1/20 dilution of the reaction products on to duplicate pieces of Whatman DE81 paper, one of which was washed to remove unincorporated nucleotides, as follows: six 5 min washes in 0.5 M Na_2HPO_4, two 1 min washes in deionized water and two 1 min washes in absolute ethanol, after which the paper was left to dry. Cerenkov counting (i.e. without scintillation fluid) of these samples enabled

calculation of the percentage incorporation of [α-^{32}P]dCTP, from which the yield of cDNA produced was calculated.

Synthesis of cDNA on a larger scale followed, using 5 μg mRNA in the reactions. To check the reaction efficiency, an aliquot was removed from the second strand synthesis reaction and added to 1 μl of [α-^{32}P]dCTP (10 μCi). This sample was used to calculate the quantity of cDNA synthesized as in the pilot reaction.

The size range of synthesis products was determined by electrophoresis of the labelled cDNA. Second strand synthesis products totalling 20,000 cpm were electrophoresed together with ^{32}P labelled, *Hin*dIII digested lambda DNA, in a 1% agarose gel in TBE buffer (Section 2.5.7), for 3.8 h at 100 V. The gel was dried between sheets of Whatman 3MM paper and paper towels for 2 h, then sealed into a plastic bag and autoradiographed.

2.1.6 Methylation of cDNA

Before methylating the cDNA, a trial methylation of bacteriophage λ DNA was performed to estimate the quantity of *Eco*RI methylase enzyme required. The reaction conditions were as specified in the Bethesda Research Laboratories catalogue and reference guide (1988). λ DNA (1 μg) was incubated with 80 μM S-adenosyl methionine and *Eco*RI methylase ranging from 0-8 U, in buffer (0.1 M Tris/ 0.1 M NaCl/ 1mM EDTA), for 1 h at 37°C, followed by 10 min at 65°C. The results were checked by digestion with *Eco*RI. This involved the addition of 5 U *Eco*RI and $MgCl_2$ to 5 mM, adjusting the solution by the addition of restriction enzyme buffer, then incubation at 37°C for 90 min. Agarose gel electrophoresis was used to determine the extent of methylation of each sample.

The cDNA produced by a synthesis reaction using 5 μg mRNA was purified by extracting twice with phenol/chloroform and twice with diethyl ether and by precipitating twice with ammonium acetate and ethanol to remove unincorporated nucleotides, as described in the kit instructions. The cDNA was then methylated using the same reaction conditions as in the λ DNA trial, with 20 U of *Eco*RI methylase. A control reaction using 1 μg of λ DNA was run in parallel. The methylated DNA solutions were then extracted with phenol/chloroform and the organic phase back extracted with TE. The pooled aqueous phases were extracted with diethyl

ether and precipitated with ammonium acetate and ethanol. The pellet was resuspended, then precipitated and resuspended again.

The methylated λ DNA was digested with *Eco*RI along with unmethylated λ DNA as a control, and the digests electrophoresed in an agarose minigel to check that the methylation had worked.

2.1.7 Ligation of *Eco*RI linkers to cDNA

*Eco*RI linkers with the sequence GGAATTCC (New England Biolabs) were phosphorylated before being ligated to the cDNA. The reaction used 5 µg of linkers, 15 U of T4 polynucleotide kinase, 10 µCi [γ-^{32}P]ATP in buffer (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 0.6 mM ATP) and was incubated at 37°C for 1 h.

The methylated cDNA pellet was resuspended in 10 µl of ligation buffer (Section 2.6) with 500 ng phosphorylated *Eco*RI linkers and 2 U T4 DNA ligase. After reacting overnight at room temperature, the reaction mixture was heated to 70°C for 10 min to inactivate the ligase and then cooled on ice. The linkers were cleaved by digestion with 25 U *Eco*RI for 3 h at 37°C to produce *Eco*RI ends. The cDNA was extracted with phenol/chloroform, the organic phase back extracted twice with TE, and the pooled aqueous phases extracted twice with ether. After ethanol precipitation, the pellet was resuspended in 20 µl TE with 5 µl DNA loading dye (Section 2.5.7).

2.1.8 Removal of linker fragments from cDNA

The cDNA was passed through a Sepharose column to remove unligated linkers and cleaved fragments of linkers which would interfere with subsequent ligation of the cDNA into a vector (Sambrook *et al.*, 1989). Sepharose CL-4B was equilibrated by washing repeatedly with TE-NaCl (10 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl). The column was formed in a siliconized 1 ml pipette plugged with glass wool, and washed with several ml of TE-NaCl. The cDNA solution was applied to the column and washed through with TE-NaCl until all the radioactivity had been eluted, as determined by monitor readings. Fractions (2 drops) were collected in microcentrifuge tubes, placed in vials, and counted in a scintillation counter. The

smaller peak of radioactivity eluted first was assumed to be cDNA, so the corresponding fractions were pooled and precipitated with sodium acetate and ethanol, ready for ligation into λgt10.

2.1.9 Preparation of λgt10 vector

For preparation of vector DNA, a stock of bacteriophage λgt10 containing a minimum of clear-plaque-forming phages was required, in order to reduce the number of false recombinants in the library (Huynh *et al.*, 1985). To obtain such a stock, λgt10 was plated on a lawn of *Escherichia coli* C600 (see Section 2.7) at low density. Plating cells were prepared by growing *E. coli* C600 overnight at 37°C in 10 ml LB medium containing 0.2% maltose. The cells were pelleted by centrifugation at 1000*g* for 3 min, resuspended in 5 ml 10 mM $MgSO_4$ and stored at 4°C for up to two days before use. Plating cells (200 µl) were incubated with λgt10 at 37°C for 30 min, added to 3 ml LB top agarose and plated on LB agar plates. After incubation at 37°C overnight, five turbid plaques were picked by removing a disc of top agarose, which was placed into 1 ml SM buffer (Section 2.6.3) with 40 µl chloroform, vortexed and stored at 4°C. A 5 µl aliquot of each of the 5 stocks was plated on an *E. coli* C600 lawn as before, incubated at 37°C for 7 h and the plaques counted to determine the titre of the stocks, and to check that no clear plaques were present.

As all the plaques from each of the five stocks were turbid, all five stocks were used for λgt10 DNA preparation, following the method of Sambrook *et al.* (1989). Large (15 cm) plates were prepared by mixing 400 µl of the λgt10 stocks with 250 µl *E. coli* C600 plating cells and 10 ml LB top agarose, then pouring the mixture on to LB agar containing 0.2% glucose for a high phage titre. The plates were incubated at 37°C for 7 h, cooled at 4°C for 15 min and overlaid with 10-15 ml cold (4°C) SM buffer per plate. After standing at 4°C overnight to allow phages to diffuse into the buffer, the buffer was pipetted off and pooled in a centrifuge bottle, to which 1% chloroform was added. After thorough mixing, the solution was centrifuged at 4000*g* for 10 min to remove bacterial debris, and the supernatant poured into a new bottle. The phage were precipitated by the addition of 1/3 volume of ice cold 5 M NaCl and 50% polyethylene glycol (PEG) 6000, giving final concentrations of 1 M and 10% respectively. After standing on ice for 4 h, the phage were pelleted at 12,000*g* and 4°C for 10 min, the pellet drained thoroughly and resuspended in 15 ml SM. This solution was centrifuged at 2000*g* for 10 min to remove remaining insoluble particles. The supernatant was extracted with an equal volume of

chloroform to remove PEG. The phage solutions were further purified by ultracentrifugation at 25,000 rpm and 4°C for 2 h in a SW28 rotor of a Beckman ultracentrifuge to pellet the phages. The supernatant was poured off and the pellet resuspended in 1.4 ml SM buffer.

DNA was extracted from the phage particles by adding EDTA to 20 mM, proteinase K to 50 µg/ml and SDS to 0.5%. After incubation at 56°C for 3 h, the solution was cooled to room temperature, extracted sequentially with phenol, phenol/chloroform and chloroform, and precipitated by the addition of 2.5 volumes of ethanol. After precipitating at -20°C for 1 h, the DNA was pelleted in a microcentrifuge for 10 min, washed with 70% ethanol, dried and resuspended in 400 µl TE. The concentration of the λgt10 DNA was estimated by electrophoresis in an agarose minigel, using λ DNA for quantity standards.

To prepare the vector for ligation, 24 µg of λgt10 DNA was digested with 50 U of *Eco*RI. The vector was extracted with phenol/chloroform and chloroform, precipitated with sodium acetate and ethanol, and resuspended in 10 µl TE.

2.1.10 Ligation of cDNA into λgt10

The precipitated cDNA was resuspended in 10 µl 1x ligation buffer with 5 µg of λgt10 and 1 U T4 DNA ligase. A control ligation was also done, using 500 ng of λgt10, 15 ng of pUC 18 cut with *Eco*RI and 1 U of ligase in 5 µl of 1x ligation buffer. Ligation proceeded overnight at room temperature. To check the ligations an aliquot of the products was electrophoresed in an agarose minigel, beside an unligated equivalent of the control ligation.

2.1.11 Packaging the cDNA library

One vial of Promega Packagene Lambda DNA Packaging System (50 µl) was thawed on ice, and used according to the manufacturers instructions. The total cDNA ligation (10 µl) was added, and mixed gently. After 2 h at 22°C, 440 µl of SM buffer was added together with 25 µl of chloroform. The library was stored at 4°C.

To determine the library titre and success of the packaging reaction, serial dilutions were plated out on a lawn of *E. coli* C600. Non-recombinant λgt10 was also plated as a control.

2.1.12 Amplification of the cDNA library

The cDNA library was amplified by plating 200 µl of library and 300 µl of *E. coli* C600 *hfl*A (Section 2.7) plating cells in 10 ml LB top agarose, on a 15 cm plate of LB agar. After incubation overnight at 37°C, the plate of almost confluent plaques was overlaid with 15 ml SM buffer and left for 2 h at room temperature with occasional shaking. The buffer was collected and centrifuged at 7000*g* and 4°C for 30 min to remove cellular debris. Dimethyl sulphoxide (7%) was added and the library stored in 1 ml aliquots at -80°C with 25 µl chloroform per aliquot (Sambrook *et al.*, 1989). The titre of the amplified library was determined by plating dilutions on a lawn of *E. coli* C600 *hfl*A.

2.1.13 Analysis of clones chosen at random from the library

In order to assess the success of the library construction, the insert sizes of randomly chosen clones were determined. Phage were plated at low density, and 10 individual plaques picked for analysis.

For DNA preparation, the 10 clones were plated at high density on 9 cm LB agar plates containing 0.2% glucose, to give confluent lysis of the lawn of *E. coli* C600 *hfl*A. After overnight incubation at 37°C, the plates were overlaid with SM buffer and left at room temperature for 2 h. The buffer was collected, 100 µl chloroform was added, and the solution was vortexed and centrifuged at 4000*g* for 10 min to remove bacterial debris. Phage were precipitated by the addition of NaCl to 1 M and PEG 6000 to 10%, followed by 2 h on ice. The phage were pelleted by centrifugation at 12000*g* and 4°C for 10 min, resuspended in 500 µl SM buffer and incubated with 10 µg RNase A and 2.5 µg DNase I at 37°C for 30 min. DNA was isolated from the phage particles as described in Section 2.1.9, and resuspended in 40 µl TE.

The insert sizes were determined as follows: 10 µl of phage DNA was digested with 20 U of *Eco*RI at 37°C for 3 h. Remaining RNA was removed by the addition of 10 µg RNase A and

incubation at 37°C for a further 3 h. The DNA was then extracted with phenol/chloroform, followed by chloroform, and precipitated with sodium acetate and ethanol. After centrifugation, the pelleted DNA was washed with 70% ethanol, dried and resuspended in 15 µl TE, to which 5 µl DNA loading dye (Section 2.5.7) was added. Samples were heated to 65°C for 5 min, chilled on ice for 5 min, and loaded onto a 1% agarose gel. After electrophoresis at 80 V for 5 h in TBE buffer, the gel was stained in 0.5 µg/ml ethidium bromide in TBE, and photographed in UV light.

2.2 ISOLATION AND ANALYSIS OF HORSE TRANSFERRIN cDNA

2.2.1 Probes used to screen the library

In order to isolate the horse transferrin cDNA, the library was screened using several different fragments of the human transferrin cDNA as probes. Initially the complete cDNA was used, i.e. the two *Pst*I fragments (Figure 6), and later the 1500 bp 5' fragment bordered by *Pst*I sites, and the 430 bp 3' fragment bordered by *Hin*dIII and *Pst*I sites. Probes were isolated from unwanted vector and cDNA by restriction endonuclease digestion and isolation of desired fragments from an agarose gel. The human transferrin cDNA was the same as that published by Yang *et al.*, (1984) (Dr. R.T.A. MacGillivray, personal communication) and was cloned into the *Pst*I site of pKT218 (Talmadge and Gilbert, 1980).

2.2.2 Screening the cDNA library

The cDNA library was screened according to the method of Kaiser and Murray (1985). Diluted aliquots of the library were mixed with 300 µl of *E. coli* C600 *hfl*A plating cells, incubated for 30 min at 37°C, added to 10 ml LB top agarose and plated on 15 cm LB agar plates, at a density of 10,000-50,000 pfu per plate. After incubation at 37°C overnight, the plates were cooled to 4°C for 1 h. Circles of nylon membrane (Gelman Biotrace RP) were cut slightly smaller than the plates, and placed on the plates for 1-2 min. For later identification and orientation, the membrane was labelled with pencil and pierced with an 18 gauge needle giving 3 asymmetric holes, which matched points marked on the underside of the plate. The membrane was carefully peeled off and placed DNA side upwards on Whatman 3MM paper saturated in 1.5 M NaCl/0.5 M NaOH, for 5 min. The membrane was then placed on 3MM

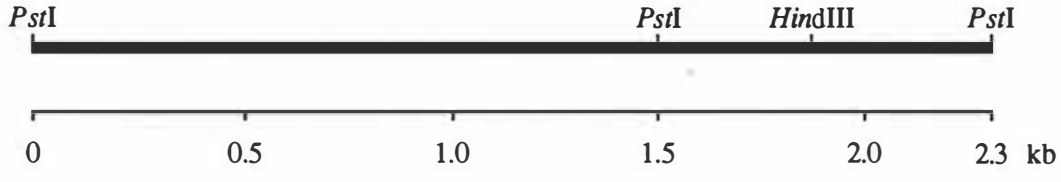


Figure 6. Restriction sites in the human transferrin cDNA which were used to create fragments for use as probes for screening the cDNA library. The probes used were: 1. The complete human transferrin cDNA i.e. the two *Pst*I fragments; 2. The 1500 bp *Pst*I fragment; 3. The 430 bp *Hin*dIII/*Pst*I fragment at the 3' end. The human transferrin cDNA was cloned into the *Pst*I site of pKT218.

paper saturated with 1.5 M NaCl/0.5 M Tris pH 7.2/1 mM EDTA for another 5 min, then washed in the same solution for a further 5 min. After two 5 min washes in 2x SSC the membrane was air dried.

The membranes were hybridized with a labelled probe to locate the positions of the desired clones. Up to four membranes were prehybridized in 25 ml of hybridization solution (containing milk powder, Section 2.5.10), sealed in a plastic bag, for 3-4 h at 65°C. Human transferrin cDNA labelled with ^{32}P was boiled for 3 min, cooled on ice and added to the hybridization solution. Hybridization occurred at 65°C overnight. The membranes were removed from the hybridization solution and washed in 3x SSC/0.5% SDS for 30 min at 65°C three times. Autoradiography was used to show the position of hybridizing plaques on the membrane, and thus the position of positive plaques on the plates. Positive plaques were picked using a pipette tip with the end cut off, placed into 500 µl SM and 30 µl chloroform, and stored at 4°C.

The plaques which gave the strongest signals were taken through a second round of screening. Aliquots from the first round stocks were plated on lawns of *E. coli* C600 *hfl*A on 9 cm plates, and the phage DNA transferred to Biotrace as before. The human transferrin probe was hybridized to the membrane using conditions identical to those used in the first round of screening. The second-round plates carried a low density of plaques so that single plaques could be picked. Aliquots from these stocks were plated and screened for a third time to check that the stocks were clonal.

2.2.3 Analysis of positive clones

DNA was prepared from the positive clones as described previously (Section 2.1.13). DNA was digested with *Eco*RI to remove the insert, or with *Bgl*II and *Hin*dIII, which cut out the insert in a 1.1 kb fragment of λgt10. The digests were analysed by agarose gel electrophoresis and by Southern blotting (Sections 2.5.9, 2.5.10) using parts of the human transferrin cDNA as probes.

2.2.4 Subcloning horse transferrin cDNA fragments

The inserts from positive λgt10 clones were subcloned into plasmids for sequencing. Inserts were removed from λgt10 DNA by digestion with *Eco*RI, and purified by phenol/chloroform extraction and ethanol precipitation. λgt10 arms were not removed from the insert preparation. The plasmids pGEM3Zf+, pGEM3Zf- and pUC18 were digested with *Eco*RI and dephosphorylated with calf intestinal phosphatase, and were purified by phenol/chloroform extraction and ethanol precipitation prior to ligation. Ligation reactions contained approximately 200 ng of plasmid DNA, 10-30 ng of insert DNA and 1 U T4 DNA ligase in 10 µl of 1x ligation buffer. Ligations were performed overnight at room temperature.

Aliquots from the ligation reactions were used to transform *E. coli* strain XL1. Plasmid DNA was prepared from transformants, and clones carrying the correct inserts were identified by electrophoresis of *Eco*RI digests. Large-scale preparations of plasmid DNA were done for those plasmids requiring further study.

2.2.5 Generation of subclones by exonuclease deletion

To sequence a large fragment of the horse transferrin cDNA, it was necessary to create a series of subclones by deleting sections of the insert from one end, thus bringing a range of different sections of the insert close to the sequencing primer binding site. This was done by cleaving the plasmid between the priming site and the insert, then digesting with exonuclease III. The plasmid had to be linearized using two different restriction enzymes, one producing an end susceptible to exonuclease III digestion adjacent to the insert, while the end towards the priming site had to be protected. Choosing restriction enzymes for linearization of the plasmid for deletion from one end of the insert was straightforward, but deletions from the other end required that the insert was in the reverse orientation within the cloning cassette of the plasmid to provide a suitable combination of restriction sites. To reverse the orientation, the insert was isolated from pGEM3Zf+ and ligated into pGEM3Zf+, pUC18 and pUC19. Clones were checked for the correct orientation by digestion with *Hin*cII which cut the plasmids once in the cloning cassette and once towards one end of the insert.

The exonuclease digestion was carried out on two plasmids each containing the insert in a different orientation with respect to the cloning cassette, so that both strands of the cDNA

could be sequenced. The method was obtained from the the Promega Protocols and Applications Guide. Plasmid DNA (3 µg) was digested with *Sma*I to produce an end susceptible to exonuclease digestion into the cDNA, and with *Sph*I to protect the vector sequences from digestion. Single digests with each enzyme were controls for both the digest and exonuclease reactions. The plasmid digests were purified by organic extractions and precipitation before being resuspended in 30 µl of exonuclease buffer (66 mM Tris pH 8, 0.66 mM $MgCl_2$) and preheated to 37°C. After the addition of 200 U of exonuclease III, the solution was incubated at 37°C for 5 min, with 3 µl aliquots removed at 30 s intervals. The aliquots were immediately added to 9 µl of a solution containing 40 mM potassium acetate pH 4.6, 340 mM NaCl, 1.4 mM $ZnSO_4$, 6.8% glycerol and 0.3 U/µl SI nuclease, to remove single stranded DNA. After standing at room temperature for 30 min the reaction was stopped by the addition of 1 µl of 0.3 M Tris pH 8, 0.05 M EDTA. At this stage aliquots were removed for electrophoresis in order to check the efficiency of the exonuclease reaction. The deleted fragments were blunt-ended by the addition of 1 µl of a solution containing 20 mM Tris pH 8, 100 mM $MgCl_2$ and 0.17 U/µl Klenow fragment (from *E. coli* DNA polymerase I). After incubation at 37°C for 3 min, 1 µl of dATP, dCTP, dGTP and dTTP each at 0.125 mM was added, and the reaction incubated a further 5 min at 37°C. The fragments were then ligated by the addition of 40 µl of 1x ligation buffer (Section 2.6) containing 5% PEG, 1 mM DTT and 1.5 U/ml T4 DNA ligase, followed by standing overnight at room temperature. The ligation products were used to transform *E. coli* XL1, and plasmid DNA was prepared for sequencing.

2.2.6 Amplification of the 3' end of horse transferrin cDNA

The polymerase chain reaction (PCR) was used to amplify the 3' end of the horse transferrin cDNA from a mixed cDNA population, as no clones containing this region were isolated from the library. This was done by a method based on that of Frohman *et al.* (1988). Oligonucleotide primers were synthesized with the following sequences:

1. 5' ACGGATCCTCTAGAGTCGACTTTTTTTTTTTTTTTTT 3'
2. 5' ACGGATCCTCTAGAGTCGAC 3'
3. 5' ATAGCGAGATCAAGCACTGT 3'

The first oligonucleotide consisted of three restriction enzyme sites (*Bam*HI, *Xba*I and *SalI*/*AccI*/*Hin*cII) which occur in the pGEM3Zf+ cloning cassette, and a run of 20 T's, and was used to prime cDNA synthesis. The second was simply the restriction site portion of the first.

The third primer was nucleotides 1499-1518 of the horse transferrin cDNA sequence, from the 3' end of the 1500 bp clone isolated from the library (Figure 22 and Section 3.2.2). The second and third primers were used for PCR.

Double stranded cDNA was synthesized using an Amersham cDNA synthesis kit, substituting primer 1 for the poly-T primer in the kit. The template for the cDNA synthesis was 1 µg of horse liver poly(A)$^+$ RNA. The products were diluted to 1 ml with TE buffer.

A Perkin Elmer Cetus GeneAmp kit provided the reaction buffer, deoxynucleotide triphosphates and *Taq* DNA polymerase for the PCR, together with template and primers for a positive control. The reaction used 1 µl of the diluted cDNA synthesized as described above, 25 pmol of each of primers 2 and 3 (a final concentration of 0.5 pmol/ml),dNTPs at 200 µM and 2 U of *Taq* polymerase in 50 µl of 1x buffer. This was overlaid with 50 µl of mineral oil to prevent evaporation. The PCR was run for 40 cycles as follows: denaturation at 94°C for 1 min, annealing at 58°C for 2 min, synthesis at 72°C for 2 min.

2.2.7 Cloning of PCR products

The PCR products were blunt ended by adding 1 U of Klenow fragment and incubating at 37°C for 30 min. The major product was isolated from a 1.2% agarose gel after electrophoresis at 100 V for 4 h, and digested with *Bam*HI. It was then ligated into pGEM3Zf+ cut with *Bam*HI and *Hin*cII (which produces a blunt end), using 100 ng of plasmid and 40 ng of insert in the reaction. Both insert and vector were purified by phenol/chloroform extraction and ethanol precipitation prior to ligation. The ligation reaction took place overnight at room temperature, in 10 µl of 1x ligation buffer with 1 U of T4 DNA ligase. The ligation products were used to transform *E. coli* XL1, and plasmid DNA was isolated.

As PCR can introduce errors in the DNA sequence (Higuchi, 1989), the section of the amplified product which had not previously been obtained in clones from the cDNA library was cloned from three independent PCR reactions. This was done by digesting the amplified material with *Bam*HI and *Eco*RI, and ligating into pGEM3Zf+ cut with the same enzymes. To sequence both strands of the resulting clones completely, subclones were produced using an *Xho*I site in the transferrin cDNA. Double digests of *Xho*I and enzymes cutting at either end of the cloning cassette (*Sal*I or *Bam*HI) were used to excise each half of the transferrin cDNA.

The fragments were blunt ended using the Klenow fragment of DNA polymerase I, religated and introduced into *E. coli* XL1. Plasmid DNA was isolated for sequencing.

2.2.8 Sequencing

The horse transferrin cDNA plasmid clones were sequenced by double stranded sequencing methods, using either the United States Biochemical Corporation Sequenase Version 2.0 or Amersham Multiwell microtitre plate DNA sequencing system. Both kits were used according to the manufacturers instructions, except that the primers used were 20 ng of M13 universal or reverse primers.

Sequencing gels of dimensions 400x300x0.4 mm were used, with sharkstooth combs with 5.7 mm spacing. Gels contained 6 % acrylamide (acrylamide:bisacrylamide 19:1, w/w) and 7 M urea in 1x TBE buffer. The gel solution was filtered through Whatman No.2 filter paper, degassed, and polymerized by the addition of ammonium persulphate to 0.1% and TEMED (tetramethylethylenediamine) to 0.05%. Gels were electrophoresed in 1x TBE buffer, at 1500 V, for 2-6 h, then transferred to Whatman 3MM paper, dried under vacuum at 80°C for 45 min, and autoradiographed.

2.2.9 Sequence analysis

Sequencing autoradiographs were read manually, and the sequence data assembled and manipulated using the Genetics Computing Group Sequence Analysis Software Package, Version 6.2 (GCG) (Devereux *et al.*, 1984). Transferrin, lactoferrin and melanotransferrin sequences for other species were obtained from Genbank. Multiple alignments and phylogenetic trees were done using the Clustal V package of Higgins *et al.* (1992). Pairwise sequence comparisons were made using the Gap program of the GCG package.

2.3 ANALYSIS OF KNOWN POLYMORPHIC REGIONS OF HORSE TRANSFERRIN cDNA

Comparison of the horse transferrin cDNA sequence and its derived protein sequence with the published data on differences between the D and R variants of horse transferrin (Chung and McKenzie, 1985) indicated two regions of the cDNA which were likely to be polymorphic (see Figure 24, Section 3.5). These two regions were then analysed in the genomic DNA of ten thoroughbred horses, carrying seven transferrin alleles (D, F_1, F_2, H_2, O, R, *), including a new mutation (Farndale *et al.*, 1991). The transferrin phenotypes of the horses were determined at the EBTRC, Massey University.

2.3.1 Exon 12: Southern blotting

The cDNA sequence and published data predicted a polymorphic *Sau*3AI site in the region corresponding to exon 12 in the human transferrin gene. Genomic DNA (5 µg), from ten horses of different transferrin phenotypes, was digested with *Sau*3AI. The digestion products were electrophoresed in a 1% agarose gel at 50 V for 12 h, and the DNA was transferred to Amersham Hybond N+ membrane by Southern blotting.

PCR was used to amplify 89 bp of exon 12 from horse genomic DNA, using primers based on the horse transferrin cDNA sequence, as follows:

12/1. 5' ATGCTGTGGCCGTGGTTAAG 3'

12/2. 5' TATAGAGCAGGCCCATGGGG 3'

The PCR reaction conditions were as described in Section 2.2.6 except that the template was 50 ng of horse genomic DNA. The reaction was incubated for 30 cycles of 1 min at 94°C and 1 min at 65°C. No synthesis step was included as the small product was synthesized during heating to the denaturing temperature. The major product was isolated from a 2% agarose gel, labelled with [α-^{32}P]dCTP, and hybridised to the *Sau*3AI digested genomic DNA.

2.3.2 Exon 12: cloning and sequencing

PCR was used to amplify exon 12 from four horses of transferrin phenotypes DD, DO, DR, OR, as described above. In some cases, products from a previous amplification were used as

template instead of genomic DNA. The products were blunt-ended, isolated from a 2% agarose gel and ligated into pGEM3Zf+ cut with *Hin*cII, as described in Section 2.2.7, except that a 10-fold molar excess of inserts over vector was used. After transformation and preparation of plasmid DNA, 1 or 2 clones from each horse were sequenced using an Amersham Multiwell sequencing kit.

2.3.3 Exon 15: amplification and sequencing

The region corresponding to exon 15 in the human transferrin gene was also predicted to have a polymorphic sequence. This region was studied by two slightly different methods of direct sequencing of PCR products, and by sequencing of cloned PCR products. For all three methods, PCR was used to amplify the exon 15 region from genomic DNA (50 ng), or from products of a previous amplification, using two of the following three primers:

15/1. 5' ACCCTGACGATTGGGCTAAG 3'
15/2. 5' TACAGCATGATTCGGGGCTC 3'
15/3. 5' CTGGTTGTGTAACTCCTGGC 3'

The first method of direct sequencing involved amplification using primers 1 and 2, and 30 cycles of 1 min at 94°C, 1 min at 60°C. After purification by agarose gel electrophoresis and spun column chromatography, the products were sequenced using one of the amplification primers. The sequencing reactions used an Amersham Multiwell sequencing kit, but the protocol was modified to reduce the reannealing of template DNA strands. The PCR products (100-200 ng) and 50 ng of primer, in a volume of 30 µl, were added to the annealing buffer from the kit. This solution was heated to 95°C for 5 min, then frozen in a dry ice/ethanol bath. Immediately after thawing and centrifugation, T7 DNA polymerase and labelling mix from the kit were added, and also 1 µl of [^{35}S]dATPαS. The solution was mixed by pipetting, then 6 µl added to each of the four termination solutions, which had been preheated to 50°C. The wells were sealed and the reactions incubated at 50°C for 4 min, after which 4 µl of stop dye were added. The reactions were heated to 95°C for 5 min, then 4 µl aliquots were loaded into a sequencing gel, and run at 1500 V for 1.5 h.

The second method for direct sequencing of PCR products differed from the first in that primers 1 and 3 were used for the amplification, while the sequencing reaction was primed by primer 2. Also the annealing temperature was 58°C.

The region of exon 15 lying between primers 1 and 2 was amplified from genomic DNA from horses with the phenotypes DO, DR, D*, F_1R, F_2H_2 and OR. The products were cloned and sequenced as described above for exon 12 (Section 2.3.2).

2.4 IDENTIFICATION AND ANALYSIS OF OTHER POLYMORPHIC REGIONS

Other polymorphic regions of the horse transferrin cDNA were identified by the single stranded conformation polymorphism (SSCP) method of Orita *et al.*(1989). This method involves amplifying fragments of DNA by PCR while incorporating a radioactive label. Fragments are denatured by heating in a formamide buffer and electrophoresed in a polyacrylamide gel under non-denaturing conditions. Base-pairing occurs within, rather than between, strands and the fragments form structures whose conformation, and therefore electrophoretic mobility, are dependent on sequence. Thus a single nucleotide substitution can be detected as a mobility shift.

2.4.1 Samples for SSCP analysis

Liver was obtained from two more horses, as described in Section 2.1.1, and blood samples were taken for blood typing. RNA was prepared by the method of Chomczynski and Sacchi (1987). Frozen horse liver (1 g) was homogenized in 10 ml guanidine isothiocyanate/sarkosyl solution as described previously (Section 2.1.2). To this, 1 ml of 2 M sodium acetate pH 4, 10 ml phenol and 2 ml chloroform/isoamyl alcohol were added, with mixing between each addition. The final suspension was shaken vigorously for 10 s and cooled on ice for 15 min. After centrifugation at 10,000*g* and 4°C for 20 min, the aqueous phase containing the RNA was transferred to a fresh tube. It was mixed with 10 ml isopropanol and placed at -20°C for 1 h to precipitate the RNA. After a further centrifugation at 10,000*g* and 4°C for 20 min, the resulting pellet was dissolved in 3 ml guanidine isothiocyanate/sarkosyl solution and precipitated again by the addition of 3 ml isopropanol and standing at -20°C for 1 h. The pellet produced by centrifugation at 10,000*g* and 4°C for 10 min was washed in 70% ethanol/0.3 mM NaCl, centrifuged at 10,000*g* for 5 min, dried under vacuum and resuspended in HEPES/EDTA buffer.

cDNA was synthesized as described in Section 2.2.6, except that total RNA was used instead of poly(A)+ RNA, and only the first strand was synthesized. These shortcuts were possible as the cDNA was to be used as a template for PCR instead of for library construction. These two cDNA samples, as well as the sample used previously for amplification of the 3' region of the horse transferrin cDNA, were used for the SSCP analysis.

PCR was used to amplify fragments of transferrin cDNA from the three cDNA samples, using some of the primers already described (Sections 2.2.6, 2.3.1, 2.3.3) and an extra one from the 5' end of the horse transferrin cDNA (bases 1-20, Figure 22), with the sequence: 5' GCGCTCCGCCGCAGACCAGG 3'. Three combinations of primers were used: the new 5' primer with the second exon 12 primer (12/2); the first exon 12 primer (12/1) with the third exon 15 primer (15/3); and the first exon 15 primer (15/1) with the 3' primer consisting of three restriction enzyme sites (Section 2.2.6, primer 2). The positions of these primer pairs in the horse transferrin cDNA sequence are shown in Figure 29 (Section 3.5.3). PCR was performed as described in Section 2.2.6 except that 1 µl of single or double stranded cDNA provided the template, 1 µl of [α-^{32}P]dCTP (10 µCi) was added in addition to 200 µmol dNTPs, and the reaction was run for 30 cycles of 94°C for 1 min, 58°C for 2 min, 72°C for 2 min. After products were isolated from a 2% agarose gel, they were digested with restriction enzymes (*Eco*RI, *Dra*I, *Fok*I or *Sau*96AI), to produce fragments of 100-550 bp for SSCP analysis (see Figure 29).

2.4.2 SSCP analysis

The fragments were electrophoresed in non-denaturing acrylamide gels containing 5% acrylamide (of which 1% was bisacrylamide), 1x TBE, 0.1% ammonium persulphate and 0.05% TEMED, using sequencing apparatus. Two sets of conditions were used: gels containing 5% glycerol were run at 30 W at room temperature; gels without glycerol were run at 50 W at 4°C.

Aliquots of the samples representing approximately 10 cps on a monitor were added to enough formamide dye solution (98% formamide, 10 mM EDTA, pH 8, 0.025% xylene cyanol, 0.025% bromophenol blue) to give a volume of 10 µl. Samples were heated to 95°C for 2 min, then 4 µl aliquots loaded on to the gel. Undenatured (i.e. unheated) samples were also loaded for comparison. After electrophoresis for 2-5 h, gels were vacuum dried at 80°C and autoradiographed.

Fragments found by SSCP analysis to be polymorphic were resynthesized using PCR, without the ^{32}P label, and digested with the appropriate restriction enzymes. If necessary, the fragments were blunt ended by the addition of 1 U Klenow fragment and deoxynucleotide triphosphates to a final concentration of 200 μM, and incubation at 37°C for 30 min. Fragments were then isolated from an agarose gel, and cloned into pGEM3Zf+ cut with *Hin*cII or *Eco*RI (as described in Section 2.2.7, but with a 10-fold molar excess of insert over vector) and sequenced.

2.5 GENERAL METHODS

2.5.1 Quantitation and purity of DNA and RNA

The concentrations of DNA and RNA solutions were determined by spectrophotometry or by comparing the intensity of fluorescence of ethidium bromide stained samples with standards (Sambrook *et al.*, 1989). The spectrophotometric method involved measuring the absorbance of a DNA or RNA solution at 260 nm, and calculating the concentration according to the following relationships. An absorbance reading of 1.0 is equivalent to 50 μg/ml of double stranded DNA, 40 μg/ml of RNA, or 20 μg/ml of single stranded oligonucleotides. The ratio absorbance at 260 nm:absorbance at 280 nm was used as an indication of purity. This ratio is 1.8 for a pure DNA solution and 2.0 for pure RNA.

To calculate DNA concentration by fluorescence intensity, the DNA sample was electrophoresed in an agarose minigel with quantity standards covering the range 10-80 ng. Ideally the molecular weight of the standards was similar to the sample. The concentration of the sample was estimated by comparing the intensity of its fluorescence to that of the standards.

2.5.2 Phenol/chloroform extractions

DNA and RNA samples were generally purified by extraction with phenol:chloroform:isoamyl alcohol, 25:24:1, v/v/v, followed by chloroform:isoamyl alcohol, 24:1, v/v (Sambrook *et al.*, 1989). If necessary, either or both extractions were repeated, and occasionally a phenol extraction was used first. Back-extraction of the organic phase with TE was done when it was

necessary to maximise the yield of DNA. Phenol was redistilled and 0.1% hydroxyquinolone was added. Phenol for DNA extractions was equilibrated in 0.1 M Tris pH 8, and for RNA extractions it was stored under 1% N-lauroylsarcosine/20 mM EDTA pH 5.

Extractions were performed by thoroughly mixing an equal volume of the organic solvent with the aqueous nucleic acid solution. The phases were separated by centrifugation, generally for 3 min in a microcentrifuge. The aqueous (upper) layer containing the nucleic acid was pipetted off, leaving the interface which contained protein.

2.5.3 Spun column chromatography

Spun column chromatography was used to purify DNA samples, particularly by removing unincorporated nucleotides, using a method based on that of Sambrook *et al.* (1989). Sephadex G-50 medium (Pharmacia) was prepared by suspending it in TE, autoclaving, pouring off excess TE and replacing it by an equal volume of fresh sterile TE. A column of Sephadex was formed in a 1 ml syringe plugged with glass wool. The column was equilibrated with deionized water before loading. The sample was applied and centrifuged at 1500*g* for 30 s. An aliquot of water was applied and similarly centrifuged, to elute the sample.

2.5.4 DNA precipitation

DNA was generally precipitated by the addition of 1/10 volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ethanol, with a 20 min period at -20°C. Pellets were obtained by centrifugation at 12,000*g* for 10 min. After removal of the supernatant, 70% ethanol was added, and the pellet recentrifuged for 5 min. The supernatant was pipetted off and the pellet dried.

2.5.5 Restriction enzyme digests

Restriction enzyme digests were generally performed at 37°C, using buffers supplied by the manufacturers, and using 2-10 U of enzyme per µg of DNA. The volume of enzyme added did not exceed 10% of the reaction volume.

2.5.6 Isolation of genomic DNA from blood

Two methods were used to isolate genomic DNA from horse blood. Most of the genomic DNA used in this work was isolated by Dr. John Forrest (DSIR) and Bruce Farndale (EBTRC), using a method based on that of Jeanpierre (1987).

Heparinized blood (10 ml) was added to 30 ml lysis solution (0.15 M NH_4Cl/10 mM $KHCO_3$/0.1 mM EDTA), mixed by inversion and left to stand until the red cells were lysed, as indicated by the solution turning dark red. After centrifugation at 1700*g* for 10 min, the pellet was washed two or three times in 10 mM EDTA/10 mM NaCl, by repeatedly resuspending and centrifuging it. The pellet was then resuspended in 7 ml of 6 M guanidine hydrochloride to which 500 µl of 7.5 M ammonium acetate, 500 µl 20% N-lauroylsarcosine and 100 µl proteinase K (10 mg/ml) were added. The solution was incubated at 60°C for 1-2 h, then the DNA was precipitated by the addition of 2 volumes of ethanol. It was spooled from the solution with a pasteur pipette and washed in 70% ethanol for 5 min. The excess ethanol was blotted off, and the DNA resuspended in TE.

Genomic DNA was also isolated from horse blood by Bruce Davidson (DSIR) using an Applied Biosystems 340A Nucleic Acids Extractor, following the manufacturers instructions.

2.5.7 Agarose gel electrophoresis

Agarose gels were used routinely to separate DNA fragments according to size. Agarose gels were prepared by dissolving the desired quantity of agarose in 1x TBE buffer (0.09 M Tris-borate, 0.002 M EDTA) by heating in a microwave oven (Sambrook *et al.*, 1989). The agarose concentration varied depending on the size of fragments to be separated, but was generally in the range 0.7-2%. For general analysis of DNA samples, a 11x14 cm minigel was used, while a larger gel (20x25 cm) was preferred for most Southern blotting, and isolation of DNA fragments. DNA samples were mixed with 1/4 volume of 5x DNA loading dye (50 mM Tris pH 8, 100 mM EDTA, 1% N-lauroylsarcosine, 15% Ficoll 400, 0.05% bromophenol blue, 0.05% xylene cyanol) prior to loading. Electrophoresis was conducted in 1x TBE buffer containing 0.25 µg/ml ethidium bromide, at 50-100V, until the bromophenol blue dye had moved approximately 3/4 the length of the gel. Gels were examined and photographed in ultraviolet light using Polaroid 667 film.

2.5.8 Isolation of DNA fragments from agarose gels

To isolate specific DNA fragments from mixtures produced by restriction digests or PCR, the fragments were separated by electrophoresis in a large (20x25 cm) agarose gel. Electrophoresis was at 100 V for 4-5 h. The desired band was isolated by electrophoresis on to DEAE-cellulose membrane (Schleicher and Schuell, NA45). The membrane was pretreated by washing in 10 mM EDTA for 5 min, 0.5 N NaOH for 5 min, then washing in 6 changes of sterile deionized water. It was inserted in a slit cut in the gel just below the band of interest. Electrophoresis was continued for 20 min at 100 V, or until the band had completely moved on to the membrane. The membrane was removed from the gel and placed in just enough high salt buffer (50 mM Tris pH 8, 1 M NaCl, 10 mM EDTA pH 8) to cover it. After incubation at 60°C for 30 min, the membrane was placed in fresh buffer, and incubated a further 15 min at 60 °C. The two aliquots of buffer were pooled and purified by either phenol/chloroform extraction or spun column chromatography, followed by ethanol precipitation and resuspension in TE (Sambrook *et al.*, 1989).

2.5.9 Southern transfer

DNA was transferred from agarose gels to nylon membranes (Gelman Biotrace RP or Amersham Hybond-N+) by capillary transfer. The transfer apparatus was similar to that described by Sambrook *et al.* (1989) but the denaturation and transfer solutions differed. Two glass plates were placed in a plastic tray to act as a support. Two layers of Whatman 3MM paper, saturated in transfer solution (0.02 M NaOH/1.5 M NaCl), were placed on top, extending beyond the glass support, to act as a wick. The tray was filled with transfer solution, to just below the support. The gel was pretreated by denaturing for 20 min in 0.4 M NaOH/1.5 M NaCl, rinsed with water and equilibrated for 20 min in transfer solution, then placed on the support. Plastic film was positioned around the edges of the gel, covering 5mm of the edge of the gel and reaching to the edge of the tray, to prevent transfer solution leaking around the edge of the gel. The nylon membrane was placed on next. Biotrace RP membrane was prewetted in hot deionized water (approximately 95°C), whereas Hybond-N+ membrane required no pretreatment. Two layers of Whatman 3MM paper, saturated in transfer solution were placed on the membrane, followed by a 5 cm pile of paper towels. As each layer was added, a glass rod was rolled over the surface to remove air bubbles. The transfer was run overnight, with paper towels and transfer solution replaced as required. After disassembling the apparatus, the

membrane was placed DNA side upwards on Whatman 3MM paper saturated in 0.4 M NaOH for 10 min to fix the DNA to the membrane (Broad *et al.*, 1988). The membrane was then rinsed briefly in 2x SSC, and soaked in fresh 2x SSC for 10 min.

2.5.10 Southern hybridization

Two methods were used for hybridizing probes to Southern blots. Membranes carrying cloned or PCR amplified DNA were prehybridized and hybridized in a solution of 6% polyethylene glycol 6000, 1.5x SSPE, 0.5% low fat milk powder and 1% SDS, after the method of Reed and Mann (1985). The prehybridization solution for membranes carrying genomic DNA digests was 6x SSPE, 5x Denhardt's solution, 0.5% SDS and 100 µg/ml denatured herring sperm DNA. The hybridization solution was the same as for prehybridization, except that the Denhardt's solution was omitted (Sambrook *et al.*, 1989).

Membranes were prehybridized in a plastic bag for 2-4 h at the temperature to be used for hybridization. The probe was boiled for 3 min, cooled on ice, and added to the hybridization solution at a concentration of 1-5 ng/ml. Hybridization was done overnight at 60-70°C, the temperature depending on the stringency required. Membranes were removed from the bag, and washed serially in 6x SSC, 3x SSC, 1x SSC each containing 0.1% SDS. Time, temperature, number, and final SSC concentration of the washes varied depending on the stringency required.

2.5.11 Labelling of DNA probes

Probes for Northern and Southern hybridizations were labelled with [α-^{32}P]dCTP using a BRL Random Primers DNA Labeling System, according to the manufacturers instructions. Probes for genomic blots were labelled using 5 µl of [α-^{32}P]dCTP (specific activity 3000 Ci/mmol, 10µCi/µl), and unincorporated nucleotides were removed by spun column chromatography. In other cases, 2 µl of [α-^{32}P]dCTP were used, and removal of the excess was not necessary.

The effectiveness of the labelling reaction was determined by thin layer chromatography of the products on PEI-cellulose (Schleicher and Schuell). Reaction products (1 µl) were spotted 1cm from the bottom of a strip of PEI cellulose, and the bottom of the strip placed in 0.75 M

KH_2PO_4 until the buffer had nearly reached the top. The strip was then autoradiographed revealing the proportion of [α-^{32}P]dCTP incorporated into DNA. For comparison, 1 µl of reaction components prior to the addition of enzyme was included.

2.5.12 End-labelling λ *Hin*dIII size standards

λ DNA (1 µg) was digested with *Hin*dIII. After digestion the following reagents were added: 1 µl of 10 mM dATP, dGTP and dTTP; 1 µl (10 µCi) [α-^{32}P]dCTP; and 2 U Klenow fragment. The solution (volume 40 µl) was incubated at 37°C for 30 min, and the products stored at -20°C. For use as a size standard, 1 µl was loaded per lane, or more depending on the age of the stock.

2.5.13 Removal of probe from membranes

Probes were removed from nylon membranes by a method based on that of Sambrook *et al.* (1989), allowing hybridization of a second probe to the membrane. The membrane was placed in a solution of 0.1x SSC and 0.1% SDS which had been heated to 95°C, and incubated in a 75°C shaking water bath for 2 h. The membrane was autoradiographed to check that all the probe had been removed.

2.5.14 Autoradiography

Fuji RX X-ray film was used for autoradiography. When ^{32}P was used, membranes/gels were wrapped in plastic prior to applying the film and an intensifying screen was used. For ^{35}S, the film was placed directly on to the dried gel. Film was exposed at -20°C for several hours or days. Film was developed for 3 min in Agfa G150 developer, rinsed for 1 min in 2% acetic acid, fixed for 2 min in Agfa G334 fixer, then washed thoroughly with water.

2.5.15 Transformation of *E. coli* XL1 with plasmid DNA

The methods for preparation of competent cells and transformation were based on those of Sambrook *et al.* (1989).

2.5.15.1 Preparation of competent cells

E. coli XL1 was cultured overnight at 37°C in 10 ml LB with 30 µg/ml ampicillin. Another 10 ml LB containing 30 µg/ml ampicillin was inoculated with 100 µl of the overnight culture, and grown at 37°C for 4 h with vigorous shaking. This gave an optical density at 600 nm close to the desired value of 0.5, indicating that the cells were growing exponentially. The culture was placed on ice for 30 min, then centrifuged at 1000*g* for 3 min to pellet the cells. After removal of the supernatant, the cells were resuspended gently in 10 ml of ice cold 100 mM $MgCl_2$, and centrifuged as before. The pellet was resuspended gently in 500 µl ice cold 50 mM $CaCl_2$, and stored on ice for 1 h before being transformed.

2.5.15.2 Transformation

Competent cells (100 µl) and plasmid DNA (20-50 ng) were mixed gently and stored on ice for 30 min. Cells were then heat-shocked by incubation at 42°C for 1 min, and placed back on ice for 2 min. LB medium (1 ml) was added, and the cells incubated at 37°C for 30-60 min. After the addition of 20 µl IPTG (25 mg/ml in water) and 20 µl X-gal (BCIG, 25 mg/ml in dimethyl formamide), the cells were spread on plates of LB agar containing 100 µg/ml ampicillin, and incubated overnight at 37°C.

2.5.16 Small scale preparation of plasmid DNA

Plasmid DNA was prepared by a small scale method based on that of Holmes and Quigley (1981). Cultures were grown overnight at 37°C in 10 ml LB medium containing ampicillin at 30 µg/ml, then centrifuged at 1000*g* for 7 min, and most of the supernatant poured off. Pellets were resuspended in the remaining medium, and transferred to 1.5 ml plastic microcentrifuge tubes. Bacteria were lysed by the addition of 700 µl of STET buffer (100 mM NaCl, 10 mM

Tris pH 8, 1 mM EDTA pH 8, 5% Triton-X 100) and 20 µl of 10 mg/ml lysozyme. After standing for 5 min, the samples were heated in boiling water for 2 min, and centrifuged for 10 min at 12,000*g*. The supernatant was removed to a fresh tube, to which 20 µl of RNase A (10 mg/ml, DNase-free) was added. After a 30 min incubation at 37°C, the solution was extracted once with phenol/chloroform and once with chloroform. DNA was precipitated by the addition of 700 µl of isopropanol, followed by 15 min at -20°C. The pellet produced by centrifugation at 12,000*g* for 15 min was dried, and resuspended in 100µl of 100 mM NaCl. It was then reprecipitated by adding 300 µl of ethanol and standing at -20°C for 20 min. The DNA was pelleted by centrifugation for 10 min, washed in 70% ethanol, vacuum dried and resuspended in 20 µl of TE buffer.

2.5.17 Large scale preparation of plasmid DNA

Methods for isolation and purification of plasmid DNA on a large scale were based on those of Sambrook *et al.* (1989).

A single bacterial colony was used to inoculate 10 ml LB containing 30 µg/ml ampicillin. After growing at 37°C for several hours with vigorous shaking, the whole culture was added to 500 ml LB containing 30 µg/ml ampicillin, and grown overnight at 37°C with vigorous shaking to aerate the medium. The culture was centrifuged at 8000*g* for 10 min to pellet the cells. Bacterial cells were resuspended in 15 ml of a solution containing 50 mM glucose, 25 mM Tris pH 8, 10 mM EDTA pH 8 and 5 mg/ml lysozyme, and left at room temperature for 10 min. An ice cold solution of 0.2 N NaOH and 1% SDS (30 ml) was added, and the solutions mixed by inversion, and chilled on ice for 10 min. After a further addition of 22.5 ml of ice cold 3M potassium acetate/2 M acetic acid, the solutions were mixed by swirling and again chilled on ice for 10 min. After centrifugation at 9000*g* for 20 min, the supernatant was poured into a new bottle, and the plasmid DNA precipitated by the addition of 0.6 volumes of isopropanol, and cooling at -20°C for 1 h. The plasmid DNA was pelleted by centrifugation at 1600*g* for 20 min, and the pellet washed twice with 40 ml of ice cold 70% ethanol. After vacuum drying, the plasmid DNA was resuspended in 5 ml TE buffer. Plasmid DNA was further purified by CsCl centrifugation or polyethylene glycol precipitation.

2.5.17.1 Cesium chloride centrifugation

The solution of plasmid DNA was centrifuged at 6000*g* for 10 min to remove insoluble debris. Ethidium bromide was added to a concentration of 630 µg/ml, and 1 g of CsCl was added per ml of solution. After gentle mixing to dissolve the salt, the solution was centrifuged at 12,000*g* for 15 min to remove insoluble matter. The density of the solution was measured by weighing a 100 µl aliquot, and adjusted to 1.55 g/ml by adding more CsCl or TE. The solution was transferred to a tube fitting a Beckman ultracentrifuge Ti65 vertical rotor. The tube was sealed and centrifuged at 40,000 rpm for 16 h. The large band of plasmid DNA in the centre of the tube was removed by inserting an 18 gauge needle, bevelled side upwards, at the bottom of the plasmid band. A 21 gauge needle was inserted in the top of the tube, and a syringe used to increase pressure in the tube. The plasmid DNA solution was collected as it escaped through the 18 gauge needle. Ethidium bromide was removed from the plasmid DNA solution by extracting 5 times with salt-saturated isopropanol. This was prepared by stirring isopropanol overnight with a solution of 5 M NaCl, 10 mM Tris pH 8, 1 mM EDTA, and leaving the phases to separate. To remove CsCl from the plasmid DNA solution, it was dialysed in 2 l of TES buffer (10 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl) for 4 h. The TES buffer was replaced and the 4 h dialysis repeated three more times. The DNA was then precipitated with sodium acetate and ethanol, pelleted, washed with 70% ethanol, vacuum dried and resuspended in TE.

2.5.17.2 Polyethylene glycol precipitation

An equal volume of ice cold 5 M LiCl was added to the solution of plasmid DNA in TE, in order to precipitate out the high molecular weight RNA. The solutions were mixed well and centrifuged at 12,000*g* for 10 min. The supernatant was added to an equal volume of isopropanol, mixed and cooled at -20°C for 10 min. After centrifugation at 12,000*g* for 10 min, the supernatant was removed and the pellet washed in 70% ethanol, vacuum dried and resuspended in TE. RNase A was added to 20 µg/ml, and the solution incubated at 37°C for 30 min. An equal volume of 1.6 M NaCl/13% PEG 6000 was added, and the solutions mixed and centrifuged at 12,000*g* for 5 min. The pellet was resuspended in TE, extracted sequentially with phenol, phenol/chloroform and chloroform, then precipitated with sodium acetate and ethanol, and resuspended in TE.

2.5.18 Storage of bacteria

Bacterial colonies were stored for up to one month on LB agar containing 100 µg/ml ampicillin or 15 µg/ml tetracycline as appropriate. For long term storage glycerol stocks were prepared (Sambrook *et al.*, 1989). Bacteria were cultured overnight in 10 ml LB medium with antibiotics as appropriate (30 µg/ml ampicillin or 15 µg/ml tetracycline). Aliquots of the cultures were stored at -20°C in 50% glycerol or at -80°C in 15% glycerol.

2.6 REAGENTS

TE was 10 mM Tris [Tris(hydroxymethyl)amino methane], 1 mM EDTA (ethylenediaminetetraacetic acid), pH 8.0. The pH of Tris solutions was adjusted with HCl.

Ligation buffer was 20 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol and 0.6 mM ATP.

TBE was 0.09 M Tris-borate, 0.002 M EDTA.

2.6.1 Reagents used for RNA preparation and electrophoresis

Guanidine isothiocyanate buffer was 4M guanidine isothiocyanate, 25 mM trisodium citrate, and 0.7% β-mercaptoethanol, pH 7.0. Guanidine isothiocyanate with sarkosyl was as above but with the addition of 5 mg/ml N-lauroylsarcosine.

HEPES/EDTA buffer was 10 mM N-2-hydroxyethylpiperazine-N'-2 ethanesulfonic acid and 1 mM EDTA, and was DEPC treated (i.e. 0.1% diethylpyrocarbonate was added, and the solution left overnight then autoclaved).

MOPS buffer was 0.1 M 3-(N-morpholino)propanesulphonic acid, 25 mM sodium acetate, 5 mM EDTA, pH 7.0.

2.6.2 Buffers used for the oligo(dT) cellulose column

Loading buffer was 0.5 M NaCl, 1 mM EDTA, 20 mM HEPES, pH 7.6. Washing buffer was 0.1 M NaCl, 1 mM EDTA, 20 mM HEPES, pH 7.6. Elution buffer was 1 mM EDTA, 10 mM HEPES, pH 7.5. All were treated with DEPC.

2.6.3 Solutions used with bacteria and phages

LB (Luria-Bertani medium) was 1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.5. LB agar was 15 g agar/l LB. LB top agarose was 7 g agarose/l LB.

IPTG (isopropylthio-β-D-galactoside) was 25 mg/ml in deionized water. X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) was 25 mg/ml in dimethylformamide.

SM buffer was 20 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$.

2.6.4 Solutions used for hybridizations

Denhardt's reagent (50x) was 1% Ficoll Type 400, 1% polyvinylpyrrolidine, 1% bovine serum albumin fraction V.

SSPE (20x) was 3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA, pH 7.4.

SSC (saline sodium citrate, 20x) was 3 M NaCl, 0.3 M sodium citrate, pH 7.0.

Herring sperm DNA (denatured, fragmented) was 10 mg/ml in water, sheared by repeatedly passing through an 18 gauge needle, boiled 10 min, and boiled again for 5 min just prior to use.

2.6.5 Oligonucleotide primers

Primers for cDNA synthesis and PCR were synthesized by Dr. Paul Ealing of DSIR Grasslands on an Applied Biosystems PCR-MATE 391 DNA synthesizer, or purchased from Oligos Inc., USA.

2.7 BACTERIAL STRAINS AND VECTORS

Escherichia coli strains C600, C600 *hfl*A, XL1 (XL1-Blue) and vectors λgt10, pUC18, pUC19 are described in detail by Sambrook *et al.* (1989). Details of pGEM3Zf+ and pGEM3Zf- are given in the Promega Biological Research Products Catalogue (1990/91). Relevant features of these strains and vectors are outlined below.

λgt10 is a bacteriophage λ derivative with a unique *Eco*RI site in the phage repressor gene (cI) which is used for cloning small fragments, up to approximately 6 kb. When propagated in *E. coli* C600, the parental λgt10, which is cI+, forms turbid plaques, while recombinants (cI-) form clear plaques. Recombinants are selected by growing them on *E. coli* C600 *hfl*A. When grown on an *hfl*A strain, the cI+ non-recombinants lysogenize at high frequency, whereas the cI- recombinants don't. Recombinants form large plaques which are ideal for screening by nucleic acid hybridization.

The plasmids pUC18, pUC19, pGEM3Zf+ and pGEM3Zf- all have multiple cloning sites in the *lacZ* (β-galactosidase) gene allowing blue/white selection of recombinants on X-gal. They also all have the β-lactamase gene conferring ampicillin resistance, and have binding sites for the pUC/M13 forward and reverse sequencing primers at each end of the multiple cloning site. The only difference between pUC18 and pUC19 is the orientation of the multiple cloning site with respect to the rest of the plasmid.

E. coli XL1 (or XL1-Blue) carries an F' plasmid with a gene for tetracycline resistance and one which allows blue/white selection on X-gal.

CHAPTER 3: RESULTS

3.1 CONSTRUCTION OF A HORSE LIVER cDNA LIBRARY

The construction of a horse liver cDNA library was a complex process involving several steps, from the collection of tissue to assessing the quality of the completed library. At almost every stage of this process the products were analysed to ensure that they were of a suitable quality, as detailed below. The only indication that the library might be flawed was observed when the inserts of several clones from the library were analysed.

3.1.1 Isolation of RNA

RNA was isolated from the liver of a horse with transferrin phenotype DF_1. The yield of RNA was approximately 0.7 mg/g of liver. A solution of the RNA had a ratio of absorbance at 260 nm to 280 nm of 1.97. This is close to the value of 2.0 which indicates a pure preparation of RNA. Electrophoresis of the total RNA in a formaldehyde-agarose gel revealed clear ribosomal RNA bands, indicating that the RNA was not extensively degraded (Figure 7A, lane 3).

When poly(A)$^+$ RNA was isolated from total RNA using an oligo(dT) cellulose column, the poly(A)$^-$ fraction was eluted first as a large peak. After the addition of low-ionic strength elution buffer, the smaller poly(A)$^+$ peak was recovered, as shown in the elution profile (Figure 8). The yield of poly(A)$^+$ RNA, calculated from its absorbance at 260 nm, was 18 µg, obtained from 2 mg of total RNA.

3.1.2 Northern blotting

A Northern blot carrying total RNA, poly(A)$^+$ RNA and poly(A)$^-$ RNA revealed hybridization to the human transferrin probe in both the total RNA and poly(A)$^+$ RNA lanes, but not in the poly(A)$^-$ lane. This is shown in Figure 7B. A single band was observed in the total RNA lane, while the more intense band in the poly(A)$^+$ lane was smeared, probably due to partial degradation of the transferrin mRNA.

Figure 7. Analysis of preparations of total, poly(A)$^+$ and poly (A)$^-$ RNA.

A. Electrophoresis of RNA isolated from horse liver.

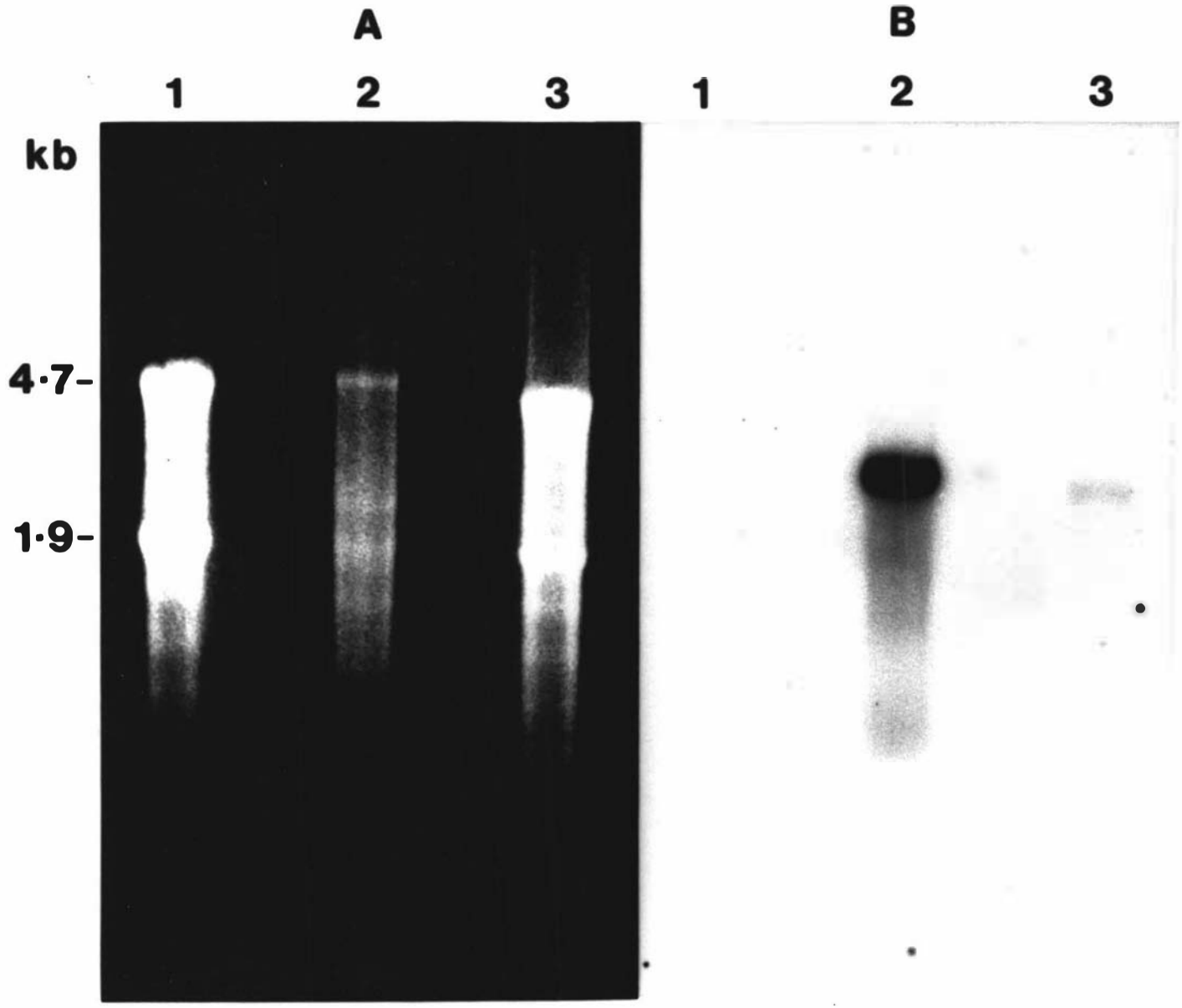
B. Northern blot of the gel in A, hybridized with the human transferrin cDNA.

Lane 1: Poly(A)$^-$ RNA, 10 µg

Lane 2: Poly(A)$^+$ RNA, 3 µg

Lane 3: Total RNA, 10 µg

Electrophoresis was in a 1.2% agarose gel, run at 80 V for 3.5 h. Hybridization was overnight at 42°C, and the final wash of the membrane was at 42°C in 0.1x SSC/0.1% SDS.


A
B
1
2
3
1
2
3
kb
4·7-
1·9-

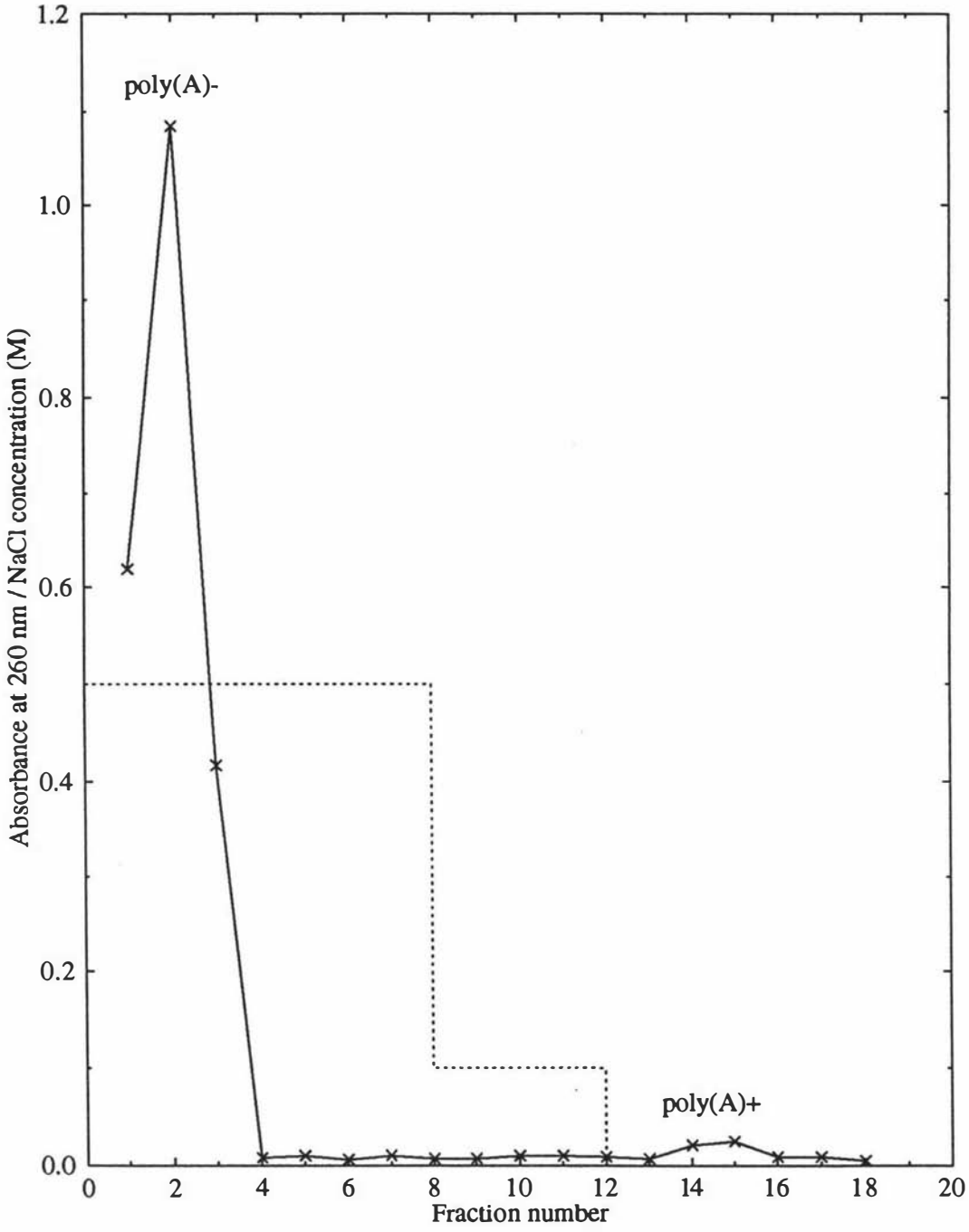


Figure 8. Elution of RNA from oligo(dT) cellulose column. Crosses represent absorbance readings of 1 ml fractions. Dotted line represents NaCl concentration of buffer. The large poly(A)- peak eluted first in loading buffer (0.5 M NaCl), whereas the small poly(A)+ peak was eluted after the addition of elution buffer containing no NaCl.

The size of the horse transferrin mRNA was estimated using the 18S and 28S rRNA bands as size markers, under the assumption that the sizes of horse rRNAs were approximately the same as rat liver rRNAs, that is, 1874 and 4718 bases respectively (Lewin,1990). The resulting size estimate for horse transferrin mRNA was approximately 2800 bases.

3.1.3 cDNA synthesis

In the pilot cDNA synthesis experiment, the proportion of [α-^{32}P]dCTP incorporated during second strand synthesis was 1.4%. Therefore 1.4% of the 40 nmol of nucleotide triphosphates in the reaction should have been incorporated, that is 0.56 nmol. As nucleotide triphosphates have a molecular weight of approximately 350, 0.56 nmol of nucleotides incorporated represents 196 ng of second strand cDNA. With the inclusion of the first strand, this becomes 392 ng of cDNA synthesized. As 1 µg of RNA template was used, the theoretical yield is 2 µg of cDNA. The actual yield was 20% of this, a yield within the range expected for the kit used.

In the large scale reaction, 4.1% of the 100 nmol (or 4.1 nmol) of nucleotide triphosphates were incorporated into second strand cDNA. This represents 1435 ng of second strand cDNA, or a yield of 2870 ng of cDNA. This is 29% of the theoretical yield of 10 µg from 5 µg of template mRNA.

Electrophoresis of the cDNA synthesis reaction showed the presence of products over a wide size range, most abundant between 1 and 3 kb, as shown in Figure 9.

3.1.4 Methylation of cDNA

In the trial methylation of λ DNA, designed to estimate the activity of the *Eco*RI methylase, 6 U of methylase was required to completely methylate 1 µg of λ DNA, as shown in Figure 10A. It was assumed that the conditions appropriate for methylation of lambda DNA would also be suitable for methylation of the cDNA. The λ DNA methylated in parallel with the cDNA (Section 2.1.6) remained intact when digested with *Eco*RI, whereas unmethylated λ DNA was cut to completion (Figure 10B). This indicated that the cDNA sample should be adequately methylated.

Figure 9. Products of the cDNA synthesis reaction.

Lane 1: λ DNA digested with *Hin*dIII and end-labelled with ^{32}P, for a size standard.

Lane 2: 2 µl aliquot from the cDNA synthesis reaction.

Electrophoresis was in a 1% agarose gel, run at 100 V for 4 h, and autoradiographic exposure was for 16 h.

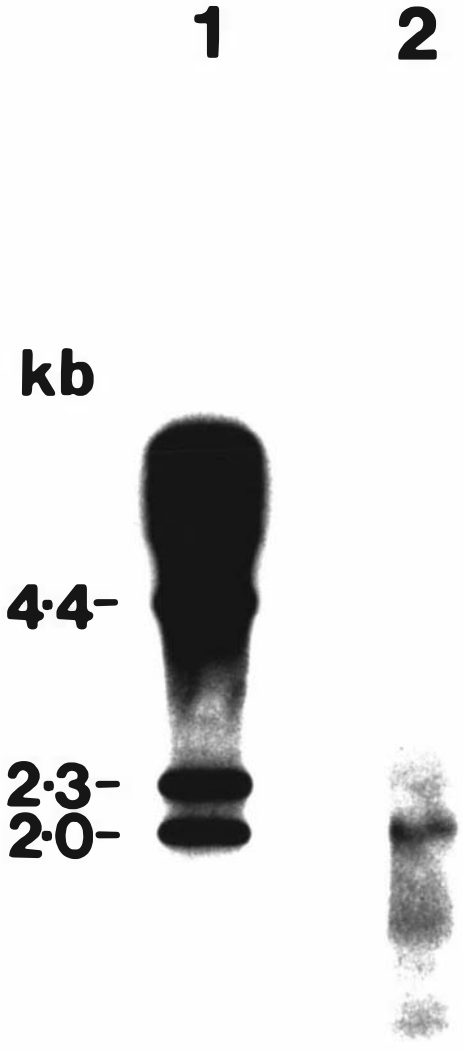

1
2
kb
4·4–
2·3–
2·0–

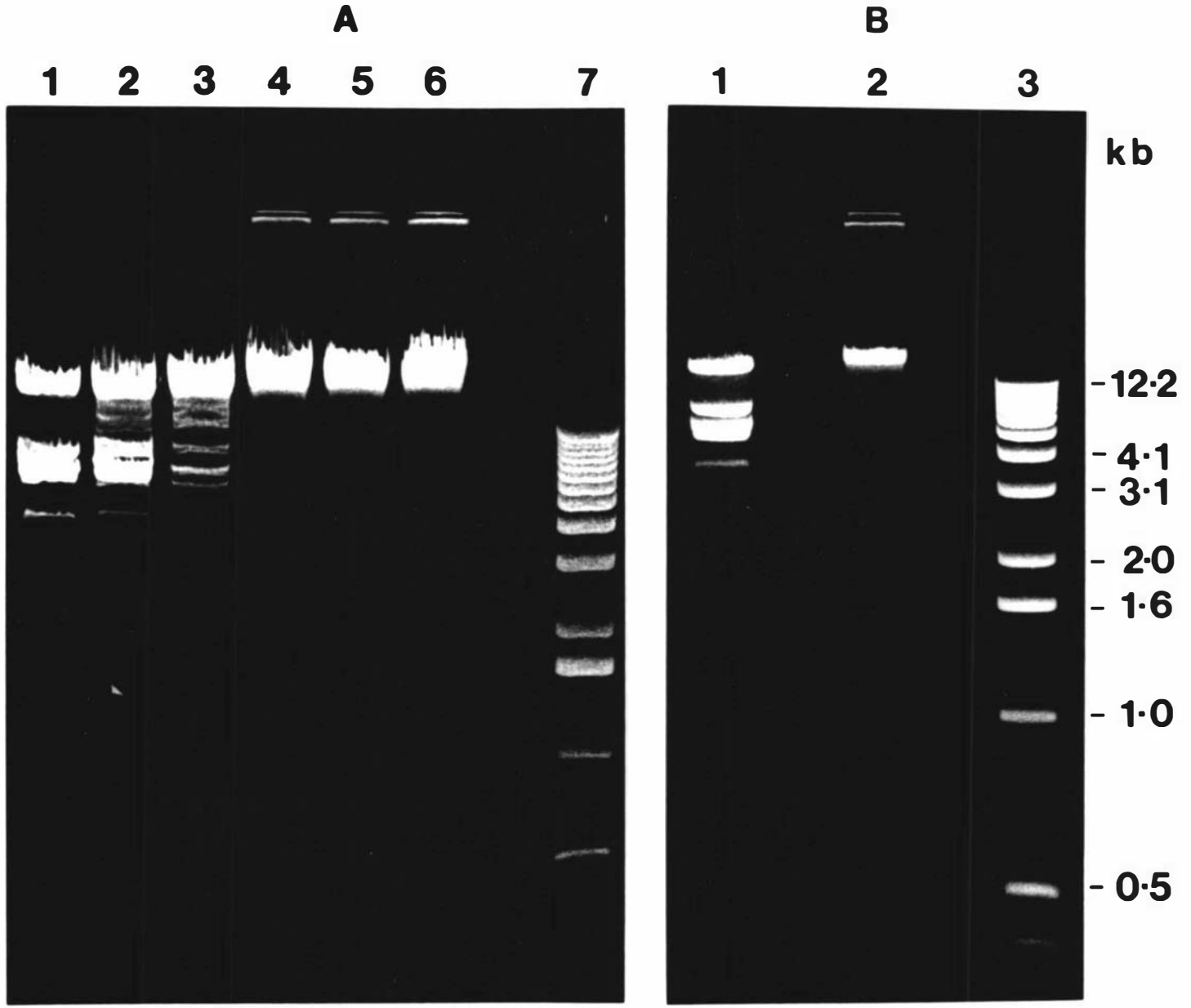
Figure 10. Trial methylation of λ DNA.

A. 1 μg of λ DNA methylated with a range of quantities of *Eco*RI methylase, then digested with *Eco*RI.

Lane 1:	0 U methylase
Lane 2:	1 U methylase
Lane 3:	2 U methylase
Lane 4:	4 U methylase
Lane 5:	6 U methylase
Lane 6:	8 U methylase
Lane 7:	BRL (Bethesda Research Laboratories) 1 kb ladder (size standard)

B. λ DNA methylated in parallel with cDNA as a control.

Lane 1:	1 μg unmethylated λ DNA cut with 5 U *Eco*RI
Lane 2:	1 μg λ DNA methylated with 20 U methylase, in parallel with cDNA sample, then digested with 5 U *Eco*RI.
Lane 3:	BRL 1 kb ladder


A
1
2
3
4
5
6
7
B
1
2
3
kb
12·2
4·1
3·1
2·0
1·6
1·0
0·5

3.1.5 Removal of linker fragments from cDNA

The elution profile of the Sepharose CL-4B column used to remove linker fragments from the cDNA sample following cleavage with *Eco*RI, is shown in Figure 11. The first, smaller peak corresponds to cDNA, and the larger peak to the linker fragments. The radioactivity in the fractions is due to the ^{32}P label on the linkers. The size of the cDNA peak relative to the linker peak reflects the small proportion of the linkers which are ligated to cDNA. As the linkers are in excess in the reaction, the cDNA peak is expected to be smaller than the linker peak.

3.1.6 Ligation of cDNA into λgt10

*Eco*RI digestion of λgt10 DNA produced two fragments corresponding to the expected sizes of 33 and 11 kb.

Electrophoresis of cDNA-λgt10 ligation reactions, together with controls (Section 2.1.10), gave evidence of a successful ligation (Figure 12). The unligated λgt10-pUC18 reaction (lane 2) had 3 fragments, the two larger ones being λgt10 arms. The smaller pUC18 fragment is not visible in Figure 12. The ligated equivalent (lane 3) has been reduced to a single high molecular weight fragment. Likewise, the ligated λgt10-cDNA appears as a single band (lane 4), indicating that ligation has occurred. The λgt10-cDNA ligation product (lane 4) appears to be smaller than the larger λgt10 arm in lane 2. The reason for this is not known. The important feature to note is that only a single band is present in lane 4, revealing that the λgt10 arms have ligated.

3.1.7 Titre of cDNA library

The packaged cDNA library, obtained from 5 µg of vector, had a titre of 3.7×10^6 plaque forming units (pfu). Of these, 5.3% formed the clear plaques indicative of recombinants. Therefore, the library was estimated to contain 2×10^5 recombinant clones. After amplification of the library on *Escherichia coli* C600 *hfl*A, the titre of recombinants was 5×10^6 pfu/µl.

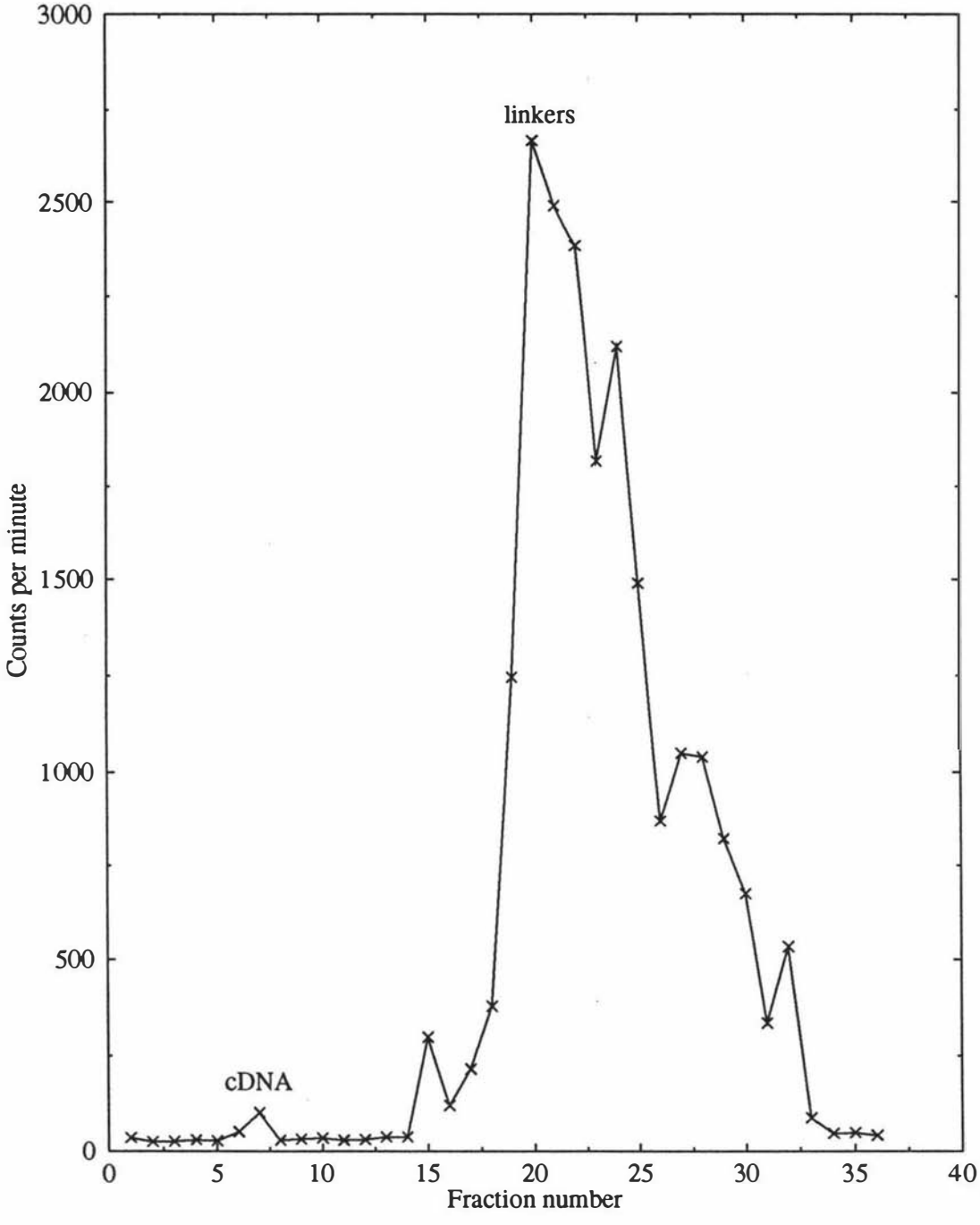
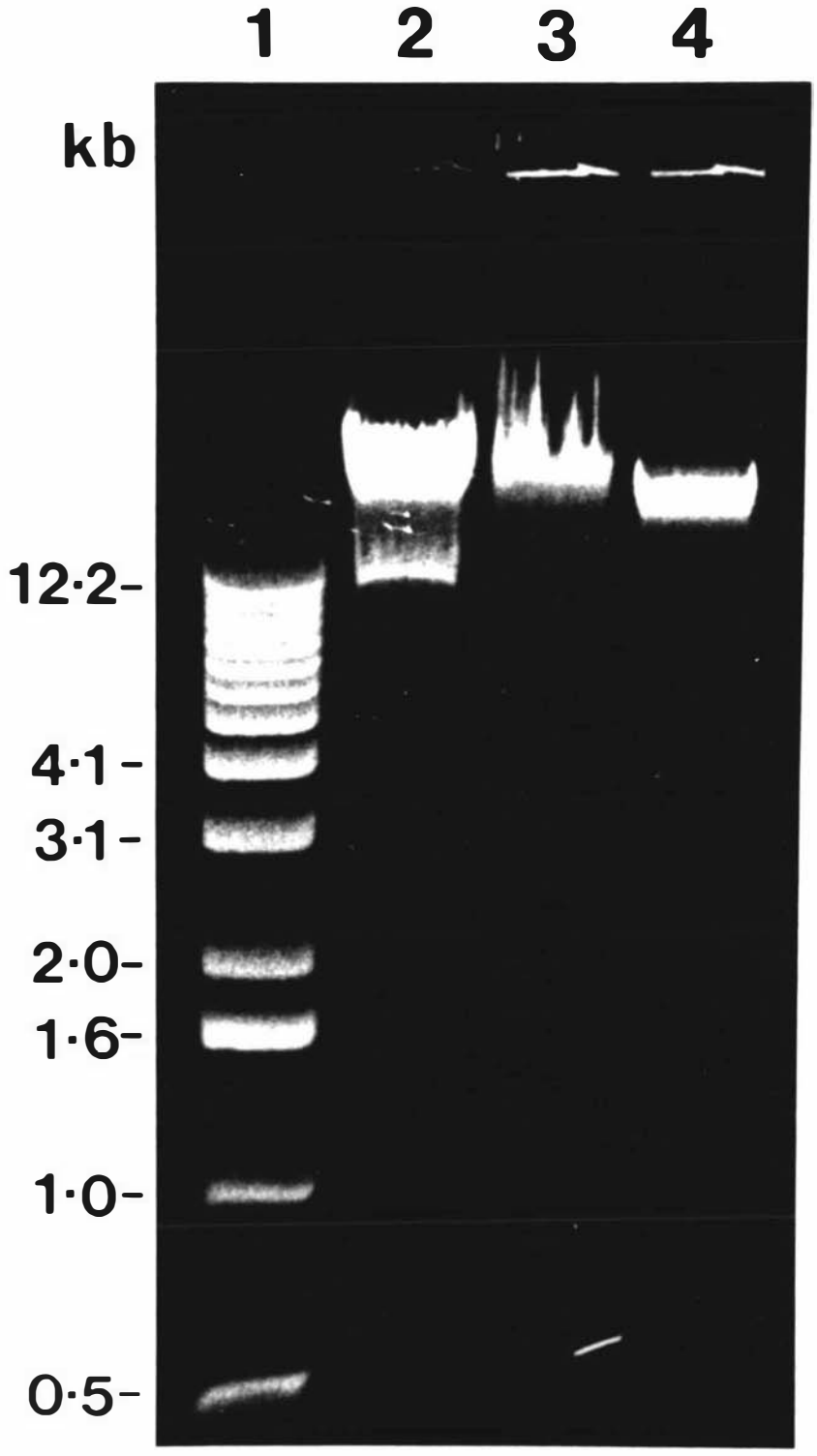


Figure 11. Elution of cDNA and linker fragments from Sepharose CL-4B column. The cDNA was eluted first followed by the linker fragments. The radioactivity in the cDNA was from the linkers ligated to it. The linker peak is much larger than the cDNA peak due to the large excess of linkers in the ligation reaction.

Figure 12. Ligation of cDNA to λgt10 arms.

Lane 1:	BRL 1kb ladder (size standard)
Lane 2:	Unligated λgt10 arms and pUC18 cut with *Eco*RI (control)
Lane 3:	Ligation of above (control)
Lane 4:	cDNA ligated to λgt10 arms

In both lanes 3 and 4 a single high molecular weight band is present indicating that ligation has occurred.


1
2
3
4
kb
12·2-
4·1-
3·1-
2·0-
1·6-
1·0-
0·5-

3.1.8 Assessment of library quality

Of ten clones chosen at random from the library, all but one contained inserts visible after agarose gel electrophoresis, as shown in Figure 13. The insert sizes were in the range 0.2 - 3.2 kb. This range includes inserts large enough to encode the complete horse transferrin cDNA. However, only single fragments were obtained for each clone, indicating an absence of internal *Eco*RI sites. As *Eco*RI would be expected to cut DNA once in every 4096 bp on average, a few of the inserts were expected to have internal *Eco*RI sites. This suggests that the methylation step may not have worked efficiently.

3.2 HORSE TRANSFERRIN cDNA-λGT10 CLONES

When the horse liver cDNA library was screened by hybridization with all or parts of the human transferrin cDNA, a number of positive clones were isolated. Closer inspection revealed that all the positive clones contained one of two contiguous fragments of horse transferrin cDNA, comprising 1800 bp of cDNA.

3.2.1 300 bp clones

Screening of 200,000 λgt10 clones using the complete human transferrin cDNA as a probe produced ten strong positive signals. When DNA isolated from these clones was cut with *Bgl*II and *Hin*dIII, a 1400 bp fragment was produced in place of the 1100 bp fragment which contains the cloning site in λgt10. This showed that there was a 300 bp insert in all ten clones. Digests of six of these clones are shown in Figure 14A. A Southern blot of these digests, hybridized with the human transferrin cDNA, produced a positive signal on all ten 1400 bp bands. Figure 14B shows the hybridization signal from six of the clones, together with the λgt10 negative control (lane 4). The insert from one of the ten clones (Figure 14, lane 5) was isolated and used to reprobe the Southern blot of the ten λgt10 digests. All ten gave signals of comparable intensity, confirming that all ten clones have similar inserts (see Figure 14C). The probe also hybridized to high molecular weight fragments. These are likely to be the products of incomplete digestion. Two of the 300 bp inserts (Figure 14, lanes 5 and 7) were subcloned and sequenced. The sequence had 80% identity to bases 1511-1804 of human transferrin cDNA.

Figure 13. *Eco*RI digests of ten clones chosen at random from the cDNA library, showing insert sizes.

Lanes 1 & 12: BRL 1 kb ladder (size standard)

Lanes 2 - 11: Clones chosen randomly from cDNA library

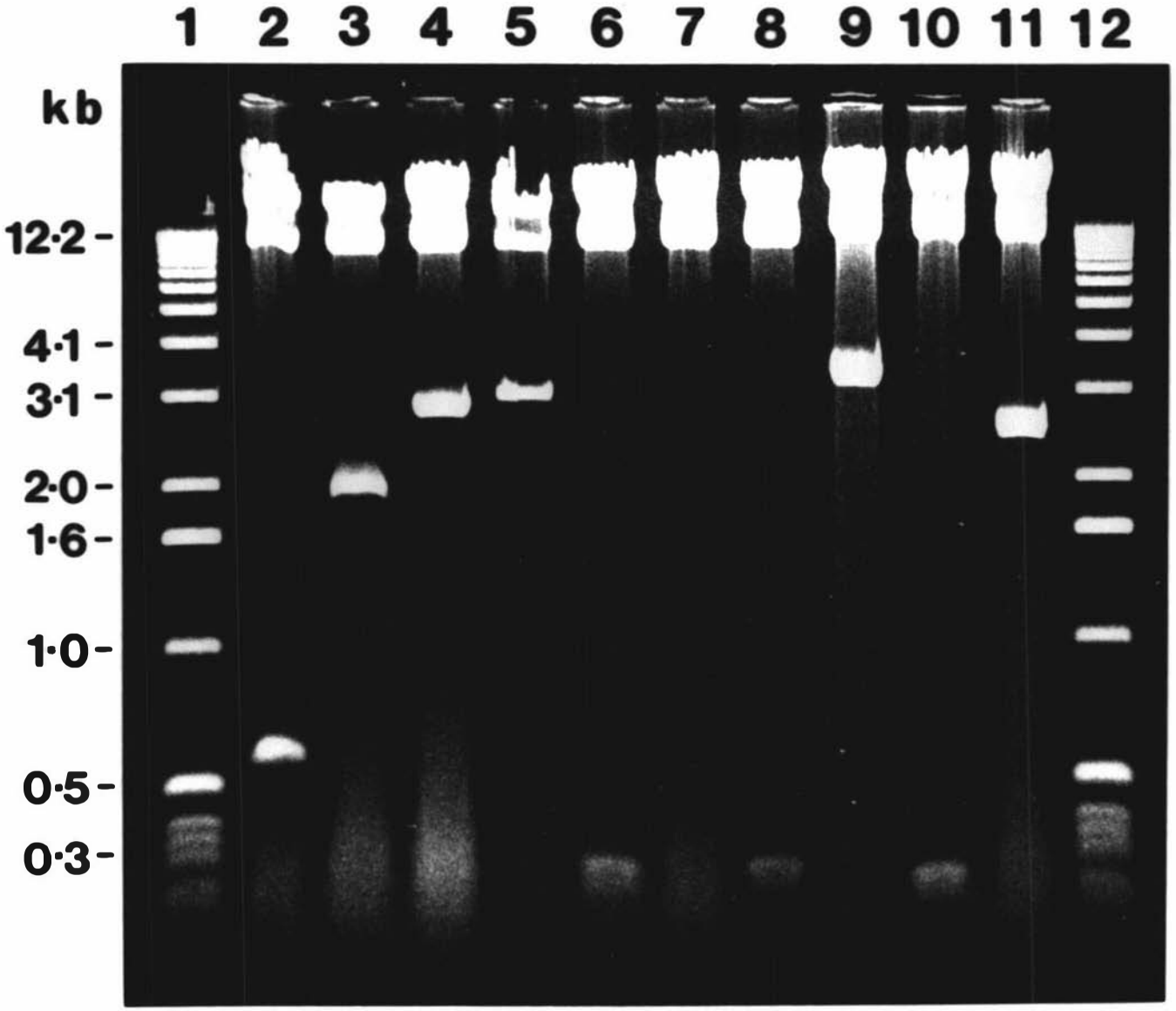

1 2 3 4 5 6 7 8 9 10 11 12
kb
12·2
4·1
3·1
2·0
1·6
1·0
0·5
0·3

Figure 14. λgt10-cDNA clones isolated from the cDNA library using the whole human transferrin cDNA as a probe.

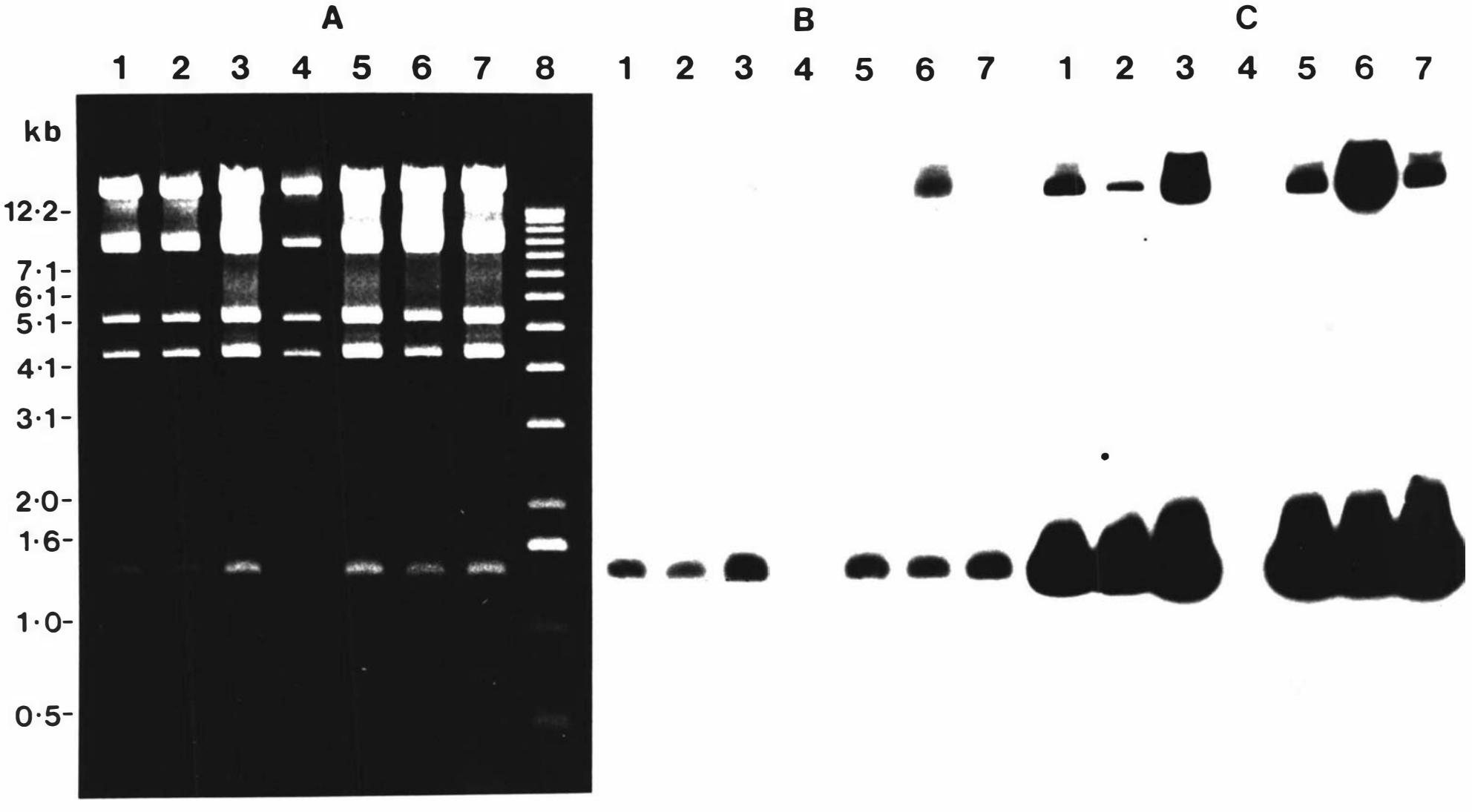
A.*Bgl*II/*Hin*dIII digests of λgt10-cDNA clones

B. Southern blot of A, hybridized with the human transferrin cDNA probe

C. Southern blot of A, hybridized with 300 bp insert

Lanes 1 - 3:	λgt10-cDNA clones
Lane 4:	λgt10
Lanes 5 - 7:	λgt10-cDNA clones
Lane 8:	BRL 1 kb ladder

Electrophoresis was in a 0.7% agarose gel, run at 30 V for 15 h. Hybridizations were at 65°C and final washes at 70°C in 1x SSC/0.1% SDS.


A
B
C
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7
1 2 3 4 5 6 7
kb
12·2-
7·1-
6·1-
5·1-
4·1-
3·1-
2·0-
1·6-
1·0-
0·5-

3.2.2 1500 bp clones

When the library was screened further using the 5' 1500 bp of the human transferrin cDNA as a probe, six positive clones were obtained, two of which gave a particularly strong signal. *Bgl*II/*Hin*dIII digests of the DNA from these clones revealed 2600 bp fragments in place of the 1100 bp λgt10 fragment in the two strong positives, indicating 1500 bp inserts (Figure 15A, lanes 2 & 8). Three of the weaker positives contained 300 bp inserts (Figure 15A, lanes 3, 4 & 7). Hybridization of the 5' 1500 bp of the human transferrin cDNA to a Southern blot of these digests gave strong signals for the 1500 bp inserts, and very weak signals for the three 300 bp inserts (Figure 15B). No signal was observed for the λgt10 control (lane 5) or the clone in lane 6.

The 300 bp inserts were shown by hybridization to be similar to the first ten 300 bp inserts (Figure 15C). This was unexpected as the 1500 bp human transferrin cDNA probe was constructed with the intention of excluding sequence equivalent to the 300 bp fragment of horse transferrin cDNA, in order to avoid isolation of more clones containing the 300 bp fragment. The isolation of these clones probably resulted from contamination of the probe preparation with other fragments of the human transferrin cDNA. This also explains the weak hybridization of the fragments containing the 300 bp insert in Figure 15B. The 1500 bp inserts did not hybridize to the 300 bp probe (Figure 15C, lanes 2 & 8) indicating that there is no overlap between these two fragments of horse transferrin cDNA.

After subcloning into pGEM3Zf+, the 1500 bp inserts were isolated and digested with a range of restriction enzymes to determine if there were major differences between the two clones. No differences were detected, as shown in Figure 16.

3.2.3 Subclones for sequencing the 1500 bp clone

For complete sequencing of both strands of the 1500 bp insert, it was necessary to create a range of subclones by deleting sections from either end of the insert. As explained in Section 2.2.5, the orientation of the insert within the cloning cassette had to be reversed to produce the second set of deletions. The orientation of the insert within the cloning cassette was determined by digestion with *Hin*cII. There is a *Hin*cII site approximately 600 bp from one end of the 1500 bp insert. Another *Hin*cII site is present in the cloning cassette and therefore lies at one end of

Figure 15. λgt10-cDNA clones isolated from the cDNA library using the 5' 1500 bp of the human transferrin cDNA as a probe.

A. *Bgl*II/*Hin*dIII digests of λgt10-cDNA clones.
B. Southern blot of A, hybridized with the 5' 1500 bp of the human transferrin cDNA.
C. Southern blot of A, hybridized with the 300 bp probe.

Lane 1:	BRL 1 kb ladder
Lanes 2 - 4:	λgt10-cDNA clones
Lane 5:	λgt10
Lanes 6 - 8:	λgt10-cDNA clones

Electrophoresis was in a 0.8% agarose gel, run at 30 V for 15 h. Hybridizations were at 70°C and final washes at 70°C in 1x SSC/0.1% SDS.

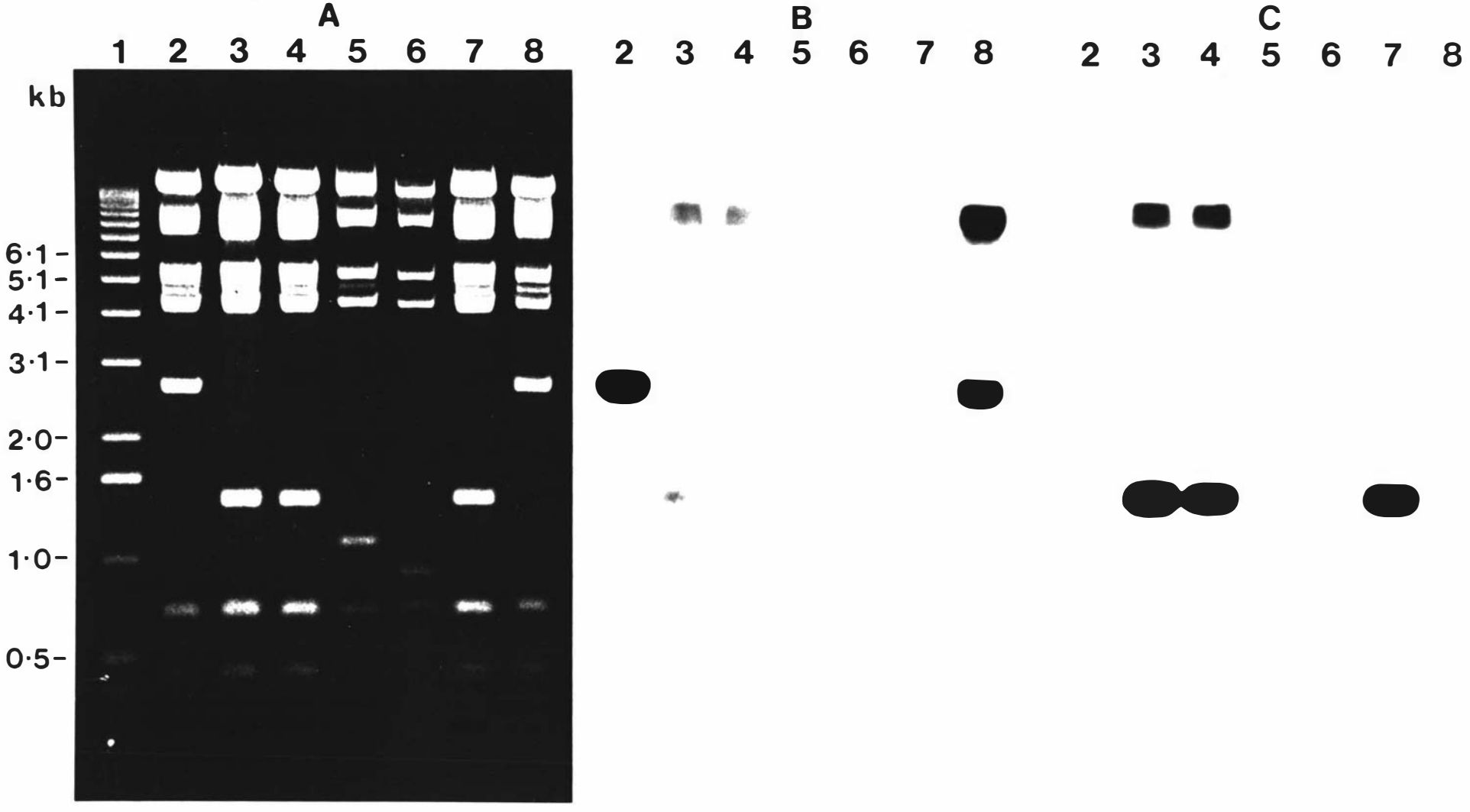
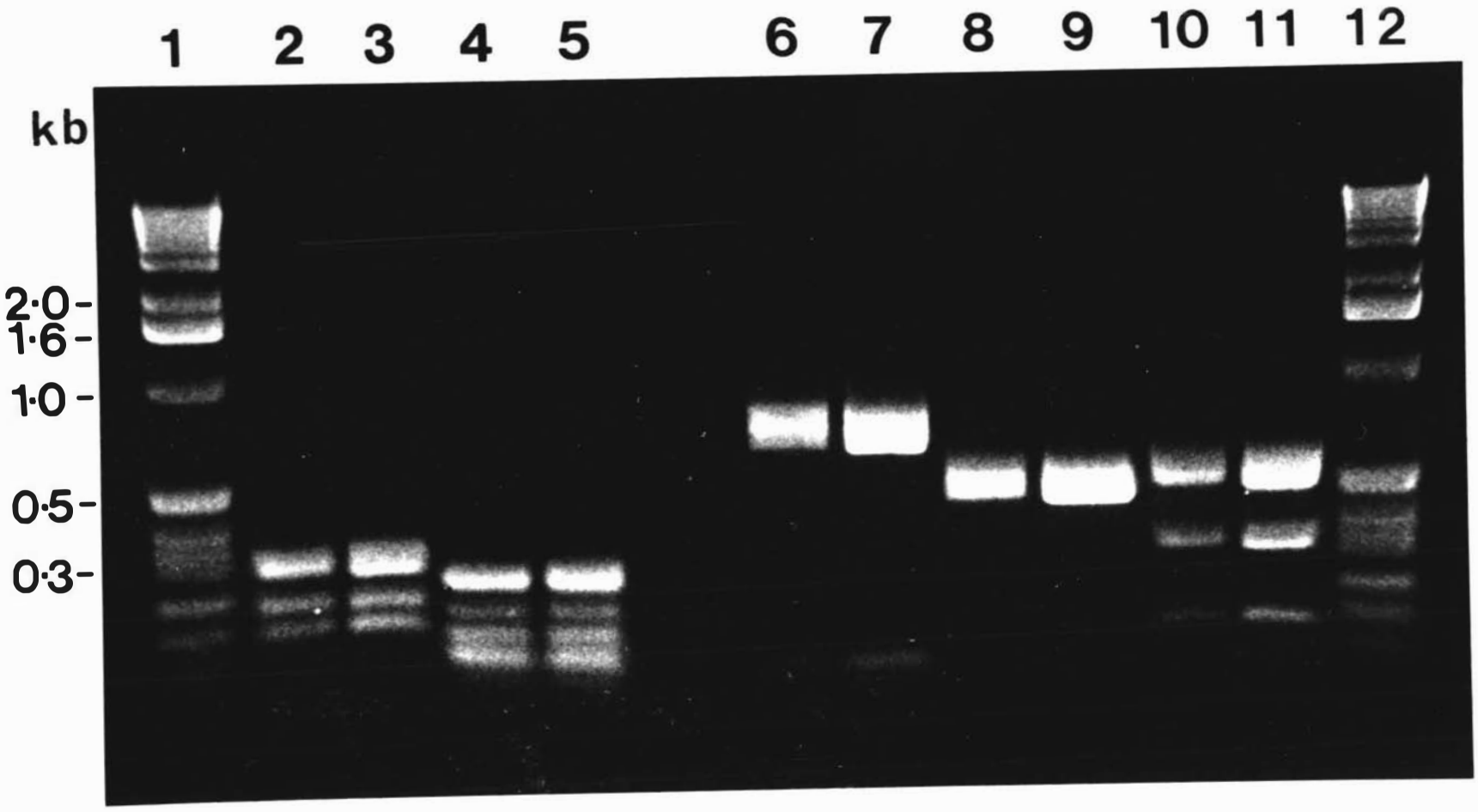

A
B
C
1 2 3 4 5 6 7 8
2 3 4 5 6 7 8
2 3 4 5 6 7 8
kb
6·1
5·1
4·1
3·1
2·0
1·6
1·0
0·5

Figure 16. Digests of the 1500 bp inserts from two λgt10-cDNA clones (lanes 2 and 8 in Figure 15).

Lanes 1 & 12:	BRL 1 kb ladder
Lane 2:	Clone 2 digested with *Alu*I
Lane 3:	Clone 8 digested with *Alu*I
Lane 4:	Clone 2 digested with *Hae*III
Lane 5:	Clone 8 digested with *Hae*III
Lane 6:	Clone 2 digested with *Hpa*II
Lane 7:	Clone 8 digested with *Hpa*II
Lane 8:	Clone 2 digested with *Rsa*I
Lane 9:	Clone 8 digested with *Rsa*I
Lane 10:	Clone 2 digested with *Sau*3AI
Lane 11:	Clone 8 digested with *Sau*3AI


1
2
3
4
5
6
7
8
9
10
11
12
kb
2·0
1·6
1·0
0·5
0·3

the insert. The fragment excised by *Hin*cII is 600 bp or 900 bp depending on the orientation of the insert. Cloning of the purified insert into pGEM3Zf+ yielded only clones with the same orientation as the original. Cloning the purified insert into pUC18 and pUC19 gave clones in both orientations. Of six clones analysed, the three in pUC19 had the insert in the original orientation, while pUC18 had them in the reverse orientation, as shown in Figure 17. The only difference between pUC18 and pUC19 is in the orientation of the cloning cassette. Therefore, by inserting into pUC18 and pUC19 in opposite orientations with respect to the cloning cassette, the insert maintains the same orientation with respect to the rest of the plasmid. As the cloning cassette lies in the *lacZ* (β-galactosidase) gene, it seems likely that it is the orientation of the insert with respect to this gene which is important. It is not clear why clones containing inserts in the opposite orientation within the *lacZ* gene were not isolated. This question was not pursued as the pUC18 clones contained the insert in the desired orientation within the cloning cassette.

Exonuclease III digestion of the plasmids containing the 1500 bp insert in either orientation yielded serial deletions of the insert from one end of the linearized plasmid. This is shown by the decreasing fragment sizes in Figure 18. Self-ligation of these fragments produced plasmids which yielded sequence data completely covering both strands of the 1500 bp insert (see Figure 21, Section 3.4).

The sequence of the 1500 bp fragment indicated that it corresponded to the 5' end of the human transferrin cDNA, through to base 1510. Thus it was contiguous with the previously sequenced 300 bp fragment, joining it at an *Eco*RI site.

3.2.4 Screening for the 3' end of horse transferrin cDNA

When the library was screened extensively with a probe comprising the 430 bp at the 3' end of the human transferrin cDNA, in order to find the 3' fragment of the horse transferrin cDNA, no positive clones were found.

Figure 17. *Hin*cII digests of plasmids containing the 1500 bp fragment of horse transferrin cDNA.

Lanes 1 - 3: pUC18 clones
Lanes 4 - 6: pUC19 clones
Lane 7: BRL 1 kb ladder

There is a *Hin*cII site approximately 600 bp from one end of the 1500 bp insert. Another *Hin*cII site is present in the cloning cassette and therefore lies at one end of the insert. The fragment excised by *Hin*cII is therefore 600 bp or 900 bp depending on the orientation of the insert. Where two inserts are present in the same orientation, two fragments are excised; one of 1500 bp and the other of 600 bp or 900 bp, as in lanes 2 and 4. The digests reveal that the insert is present in pUC18 in one orientation and in the opposite orientation in pUC19.

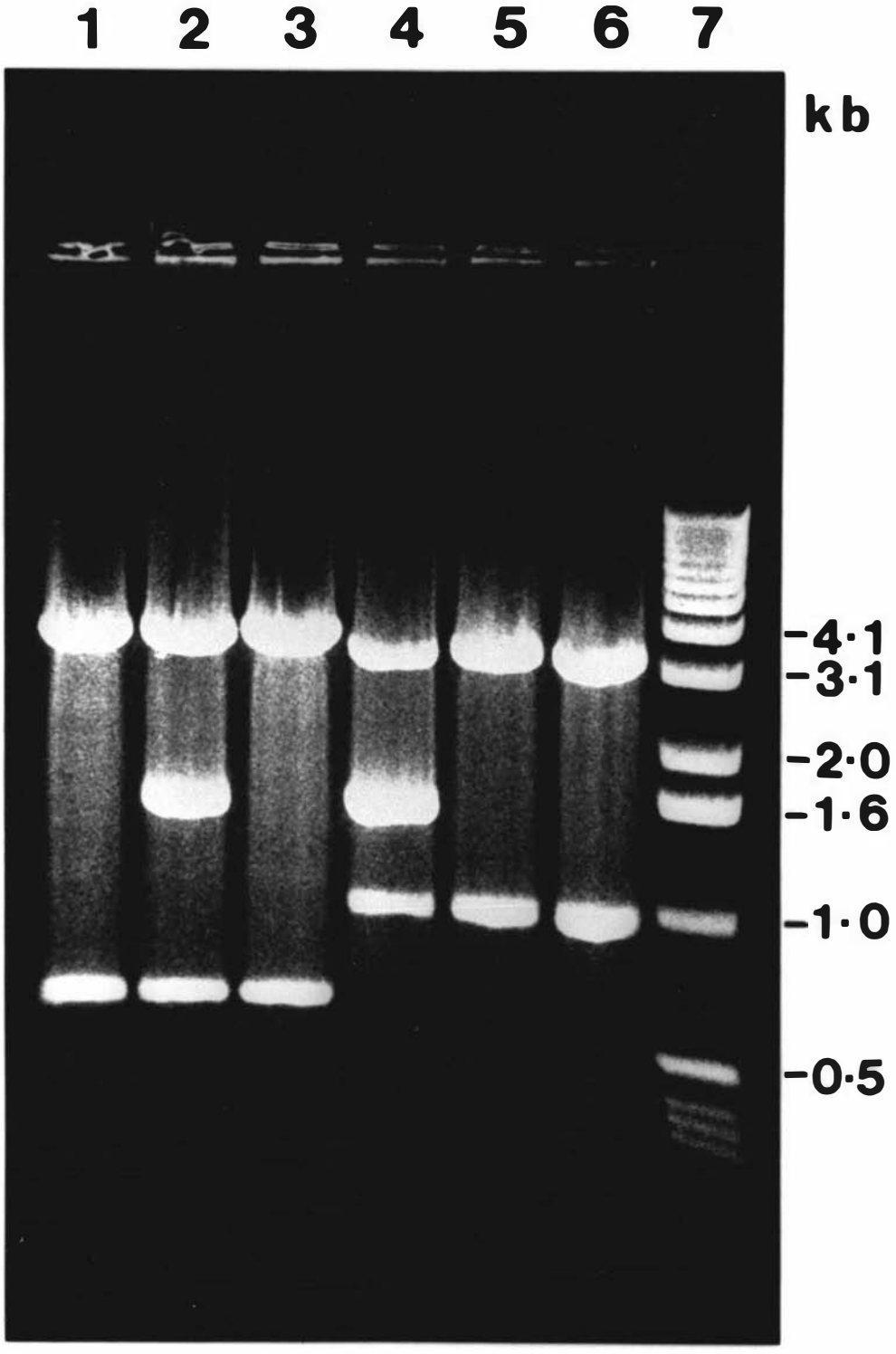
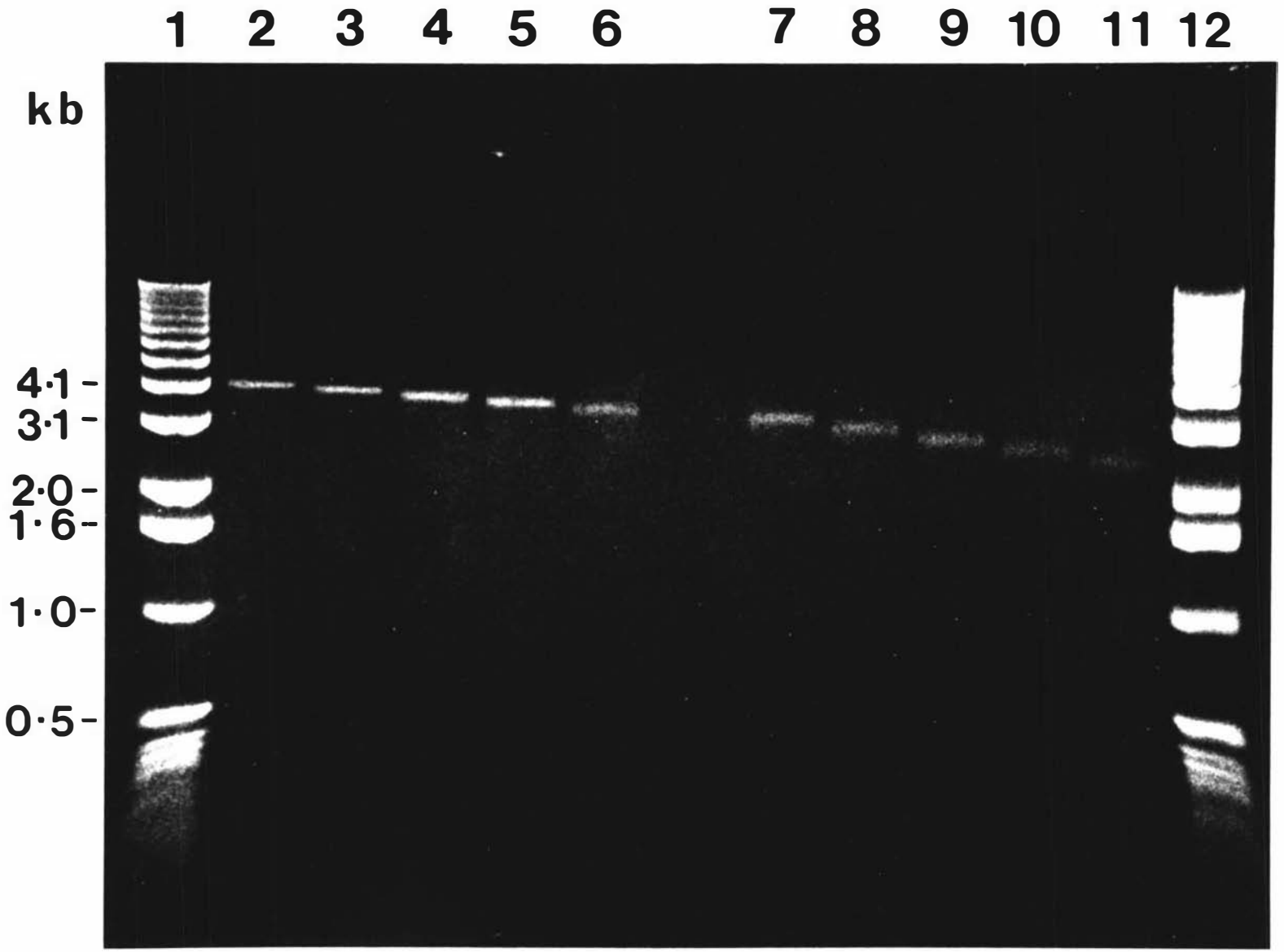

1
2
3
4
5
6
7
kb
4·1
3·1
2·0
1·6
1·0
0·5

Figure 18. Products of the exonuclease digestion of pUC18 carrying the 1500 bp fragment of horse transferrin cDNA.

Lanes 1 & 12: BRL 1 kb ladder

Lanes 2 - 11: Aliquots removed from the exonuclease reaction at 30 s intervals.

The plasmid was digested with exonuclease III after linearization with *Sma*I and *Sph*I. The first aliquot was taken after 30 s and the last after 5 min.


1 2 3 4 5 6 7 8 9 10 11 12
kb
4·1
3·1
2·0
1·6
1·0
0·5

3.3 HORSE TRANSFERRIN cDNA 3' REGION

As the remainder of the horse transferrin cDNA, the 3' end, could not be isolated from the cDNA library, it was instead obtained using PCR.

3.3.1 Amplification of the 3' region

PCR designed to amplify the 3' end of the horse transferrin cDNA, as described in Section 2.2.6, gave a product of the expected size of 800 bp (Figure 19A, lanes 2 & 3). An unexpected 480 bp product was also observed. The positive control from the GeneAmp kit produced the expected 500 bp product (Figure 19A, lane 4). A Southern blot of the PCR products is illustrated in Figure 19B. The 800 bp product in lanes 2 and 3 hybridized to the human transferrin cDNA probe, but the 480 bp product did not. The hybridization signal appears as a smear which is most intense at the position of the 800 bp fragment. The smear is probably due to incomplete PCR products smaller than 800 bp, and larger products synthesized directly from the cDNA template, and therefore extending beyond the opposite primer.

The 800 bp product was isolated, and digested with *Eco*RI and *Bam*HI. The *Eco*RI digest gave fragments of the expected sizes of approximately 300 and 500 bp, as shown in Figure 20, lane 3, confirming the identity of the product. The 300 bp fragment was equivalent to the insert in the 300 bp clones isolated from the library, whereas the size of the 500 bp fragment was predicted by homology with the human transferrin sequence. *Bam*HI did not appear to cut the 800 bp fragment, indicating that the *Bam*HI site at the extreme 3' end, introduced via the PCR primers (Section 2.2.6), was likely to be the only *Bam*HI site in the fragment. This site should therefore be useful for cloning the fragment into pGEM3Zf+. A small quantity of pUC18 DNA was included in the digests to check that the enzymes were cutting efficiently. This is evident in Figure 20 as faint bands at 2.7 kb in lanes 3 and 4.

3.3.2 Sequencing of the 3' region

The sequence of the 800 bp PCR product included the PCR primer derived from sequence at the 3' end of the 1500 bp λgt10 clone, the sequence of the 300 bp λgt10 clone, and approximately 500 bp of sequence homologous to the 3' end of the human transferrin cDNA,

Figure 19. Amplification of the 3' region of the horse transferrin cDNA by PCR.

A. Electrophoresis of the products of the PCR.

B. Southern blot of the gel in A, hybridized with the human transferrin cDNA.

Lane 1:	BRL 1 kb ladder
Lanes 2 & 3:	PCR products from the amplification of the 3' region of the horse transferrin cDNA.
Lane 4:	PCR positive control (500 bp)

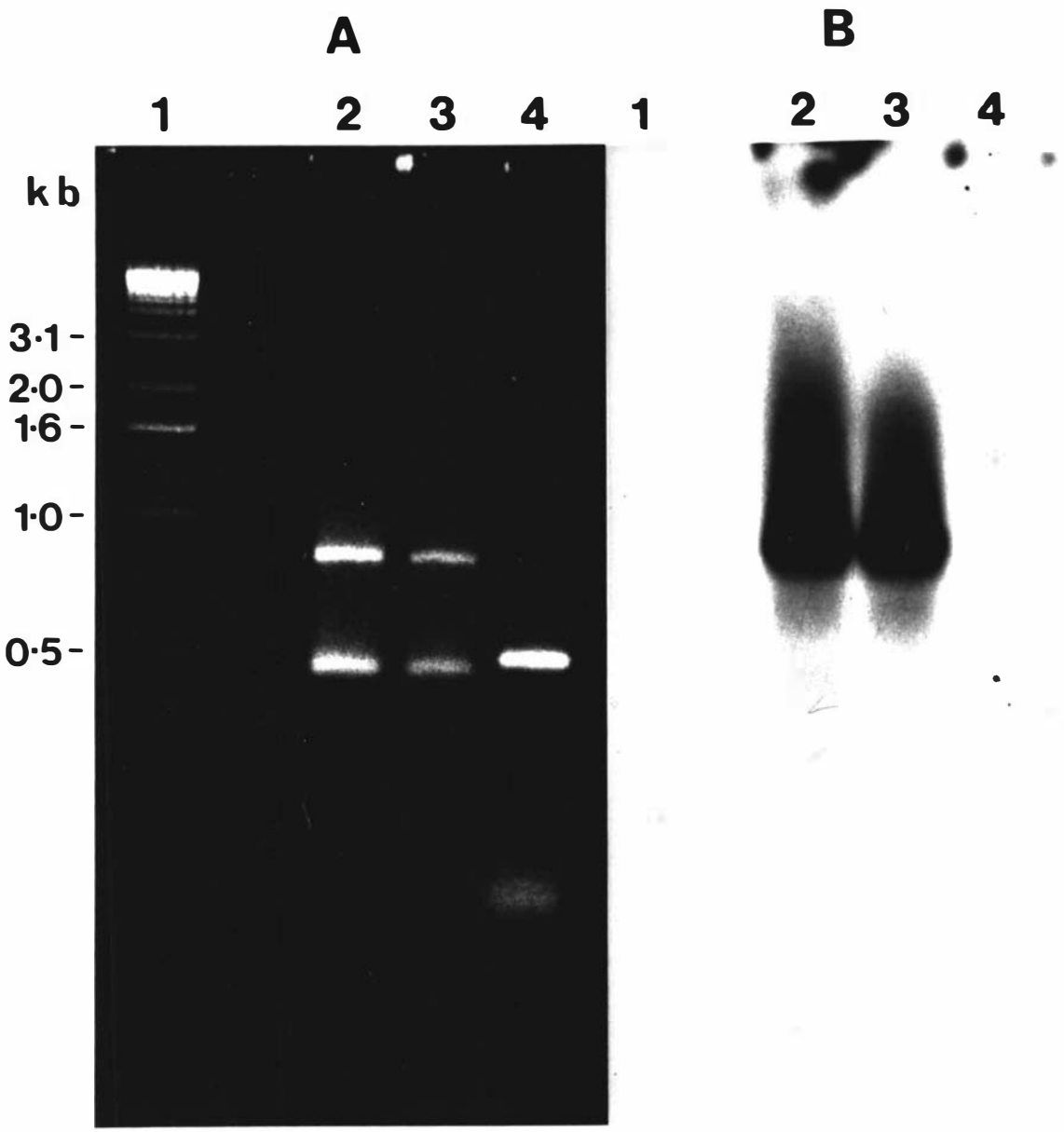
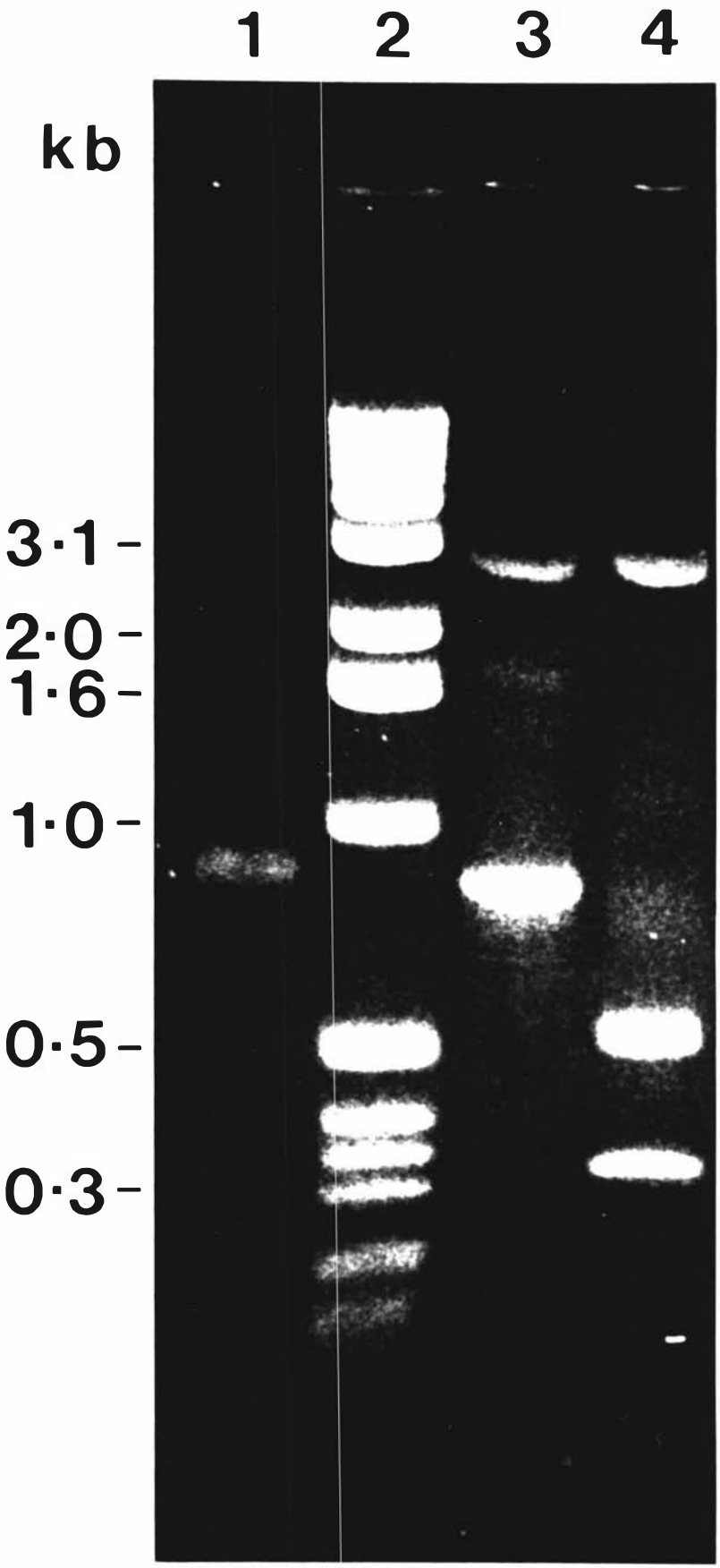

A
B
1
2
3
4
1
2
3
4
kb
3·1
2·0
1·6
1·0
0·5

Figure 20. Restriction enzyme digests of the 800 bp PCR product.

Lane 1:	Undigested 800 bp fragment.
Lane 2:	BRL 1 kb ladder
Lane 3:	*Bam*HI digest of the 800 bp fragment.
Lane 4:	*Eco*RI digest of the 800 bp fragment.

The faint 2.7 kb fragments in lanes 3 and 4 are pUC18 DNA included in the digests as a positive control of the enzyme activity.


1
2
3
4
kb
3·1
2·0
1·6
1·0
0·5
0·3

including a poly(A) tail. This confirmed that the 1500 and 300 bp fragments do join at an *Eco*RI site. Sequence data were obtained from two other clones containing the 500 bp *Eco*RI/*Bam*HI fragment at the 3' end of the cDNA, from separate PCR reactions (Figure 21, Section 3.4). The sequence from the second 500 bp clone differed from the first and third sequences each at one position. One of these differences was a cytosine/thymine substitution at position 1962 of the complete sequence, which is in the 3rd position of a codon and would not produce a substitution in the amino acid sequence. The other is an insertion/deletion of a thymine in the 3' untranslated region, adjacent to the poly(A) tail. These differences could be due to PCR errors or could be differences between the D and F_1 alleles. The sequence reported for this 500 bp region is the consensus of the sequences of the three clones.

3.4 HORSE TRANSFERRIN cDNA AND PROTEIN SEQUENCES

The horse transferrin cDNA sequence was obtained by combining sequence data from a number of clones, as shown in Figure 21.

3.4.1 Features of the cDNA sequence

The horse transferrin cDNA sequence and its predicted translation product are shown in Figure 22. The sequence consists of 2305 bp plus a poly(A) tail. It includes 24 bp 5' to the start codon ATG. The stop codon TAA is followed by a 3' untranslated region of 160 bp. A polyadenylation signal AATAAA is located at position 2277-2282.

3.4.2 Features of the amino acid sequence

The predicted translation product is a polypeptide of 706 amino acids, including a putative signal peptide of 19 residues. The calculated molecular weight of the mature peptide chain is 76,104. Two possible glycosylation sites are present: NST at positions 515-517 and NGS at 638-640.

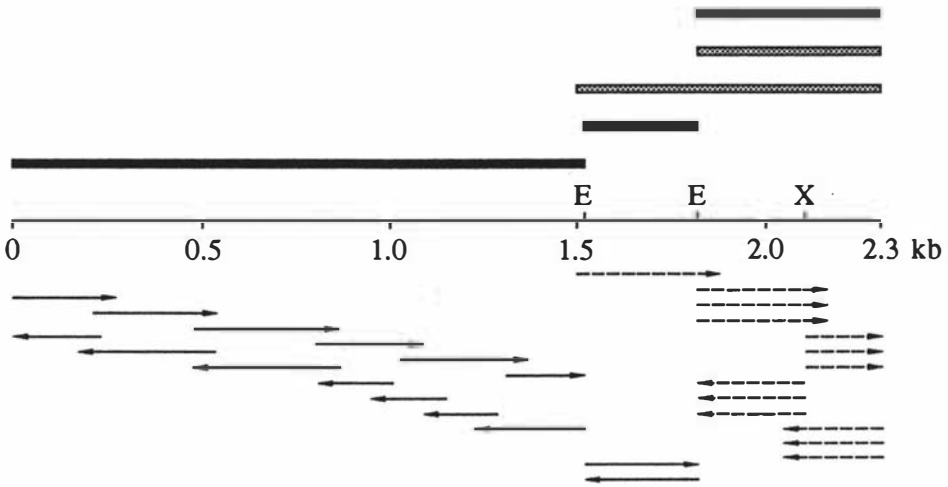


Figure 21. Horse transferrin cDNA clones and sequencing strategy. Inserts from λgt10 clones are represented by black bars. Cloned PCR products are represented by grey bars. Black arrows represent sequence data obtained from λgt10 clones and subclones produced by exonuclease digestion of the 1500 bp fragment. Dashed arrows represent sequence obtained from cloned PCR products and subclones produced by restiction enzyme digestion. E=*Eco*RI site. X=*Xho*I site.

```
   1 GCGCTCCGCCGCAGACCAGGGAGGATGAGGCTCGCCATCCGCGCCCTGCTGGCCTGCGCG   60
   1                          M  R  L  A  I  R  A  L  L  A  C  A      12
  61 GTCCTGGGGCTGTGTCTGGCGGAGCAAACTGTGAGATGGTGCACCGTCTCAAATCATGAG  120
  13 V  L  G  L  C  L  A  E  Q  T  V  R  W  C  T  V  S  N  H  E      32
 121 GTCAGTAAGTGCGCCAGTTTCCGCGACAGTATGAAAAGCATTGTTCCTGCTCCTCCTCTT  180
  33 V  S  K  C  A  S  F  R  D  S  M  K  S  I  V  P  A  P  P  L      52
 181 GTCGCCTGTGTGAAGAGAACCTCCTACCTGGAGTGCATCAAGGCCATTGCGGATAACGAA  240
  53 V  A  C  V  K  R  T  S  Y  L  E  C  I  K  A  I  A  D  N  E      72
 241 GCGGATGCTGTGACGTTGGATGCAGGTTTGGTGTTCGAAGCTGGCCTCTCCCCCTACAAC  300
  73 A  D  A  V  T  L  D  A  G  L  V  F  E  A  G  L  S  P  Y  N      92
 301 CTGAAGCCTGTAGTGGCAGAGTTCTATGGGTCCAAAACTGAGCCACAAACCCACTATTAT  360
  93 L  K  P  V  V  A  E  F  Y  G  S  K  T  E  P  Q  T  H  Y  Y     112
 361 GCTGTGGCCGTGGTGAAGAAGAACAGCAATTTCCAGCTGAACCAGCTCCAAGGCAAGAAG  420
 113 A  V  A  V  V  K  K  N  S  N  F  Q  L  N  Q  L  Q  G  K  K     132
 421 TCCTGCCACACGGGCCTTGGCAGGTCTGCTGGGTGGAACATCCCCATTGGCTTACTTTAT  480
 133 S  C  H  T  G  L  G  R  S  A  G  W  N  I  P  I  G  L  L  Y     152
 481 TGGCAATTGCCTGAGCCACGTGAATCTCTTCAGAAAGCAGTGTCCAATTTCTTCGCGGGC  540
 153 W  Q  L  P  E  P  R  E  S  L  Q  K  A  V  S  N  F  F  A  G     172
 541 AGCTGCGTACCCTGTGCAGATCGGACAGCTGTCCCCAACCTGTGTCAACTGTGTGTGGGG  600
 173 S  C  V  P  C  A  D  R  T  A  V  P  N  L  C  Q  L  C  V  G     192
 601 AAAGGGACAGACAAGTGTGCCTGCTCCAACCACGAACCATACTTTGGCTACTCAGGTGCC  660
 193 K  G  T  D  K  C  A  C  S  N  H  E  P  Y  F  G  Y  S  G  A     212
 661 TTCAAGTGCCTGGCGGATGGCGCTGGGGACGTGGCCTTTGTCAAGCATTCAACAGTATTG  720
 213 F  K  C  L  A  D  G  A  G  D  V  A  F  V  K  H  S  T  V  L     232
 721 GAGAACCTGCCACAAGAGGCTGACAGAGACGAGTATCAGCTGCTCTGCAGGGACAACACC  780
 233 E  N  L  P  Q  E  A  D  R  D  E  Y  Q  L  L  C  R  D  N  T     252
 781 CGGAAGTCAGTGGATGAATACAAGGACTGCTACCTGGCCAGCATCCCTTCCCATGCCGTT  840
 253 R  K  S  V  D  E  Y  K  D  C  Y  L  A  S  I  P  S  H  A  V     272
 841 GTGGCCCGAAGTGTGGACGGCAAGGAGGACTTGATCTGGGGGCTTCTCAACCAGGCCCAG  900
 273 V  A  R  S  V  D  G  K  E  D  L  I  W  G  L  L  N  Q  A  Q     292
 901 GAACATTTTGGCACAGAAAAATCTAAAGACTTCCATCTCTTCAGCTCCCCTCATGGGAAG  960
 293 E  H  F  G  T  E  K  S  K  D  F  H  L  F  S  S  P  H  G  K     312
 961 GACCTGCTGTTTAAGGACTCTGCCCTTGGCTTTTTAAGGATTCCCCCTGCGATGGACACC 1020
 313 D  L  L  F  K  D  S  A  L  G  F  L  R  I  P  P  A  M  D  T     332
1021 TGGCTGTACCTGGGATATGAGTATGTCACTGCTATTCGGAATCTAAGGGAAGATATACGC 1080
 333 W  L  Y  L  G  Y  E  Y  V  T  A  I  R  N  L  R  E  D  I  R     352
1081 CCAGAAGTCCCAAAGGACGAATGCAAGAAGGTGAAGTGGTGTGCAATAGGCCACCATGAG 1140
 353 P  E  V  P  K  D  E  C  K  K  V  K  W  C  A  I  G  H  H  E     372
1141 AAGGTCAAGTGTGACGAGTGGAGTGTAAACAGTGGAGGGAACATAGAGTGTGAGTCAGCA 1200
 373 K  V  K  C  D  E  W  S  V  N  S  G  G  N  I  E  C  E  S  A     392
1201 CAGTCCACTGAAGACTGCATTGCCAAGATTGTGAAAGGAGAAGCTGATGCCATGAGTTTG 1260
 393 Q  S  T  E  D  C  I  A  K  I  V  K  G  E  A  D  A  M  S  L     412
1261 GATGGAGGTTTCATCTACATAGCGGGCAAGTGTGGTCTGGTGCCTGTCCTGGCAGAGAAC 1320
 413 D  G  G  F  I  Y  I  A  G  K  C  G  L  V  P  V  L  A  E  N     432
1321 TATGAAACTAGGTCTGGCTCTGCATGTGTGGACACACCAGAGGAAGGGTATCATGCTGTG 1380
 433 Y  E  T  R  S  G  S  A  C  V  D  T  P  E  E  G  Y  H  A  V     452
1381 GCCGTGGTTAAGTCCTCATCAGATCCTGACCTCACCTGGAACTCTCTGAAAGGCAAGAAG 1440
 453 A  V  V  K  S  S  S  D  P  D  L  T  W  N  S  L  K  G  K  K     472
1441 TCCTGTCACACTGGAGTAGATAGAACCGCCGGCTGGAACATCCCCATGGGCCTGCTCTAT 1500
 473 S  C  H  T  G  V  D  R  T  A  G  W  N  I  P  M  G  L  L  Y     492
1501 AGCGAGATCAAGCACTGTGAATTCGATAAATTTTTCCGTGAAGGCTGTGCCCCTGGGTAT 1560
 493 S  E  I  K  H  C  E  F  D  K  F  F  R  E  G  C  A  P  G  Y     512
1561 AGGCGAAATTCCACCCTCTGCAATCTGTGTATTGGCTCGGCAAGTGGTCCAGGAAGGGAG 1620
 513 R  R  N  S  T  L  C  N  L  C  I  G  S  A  S  G  P  G  R  E     532
1621 TGTGAACCCAACAACCATGAGAGATACTATGGTTACACAGGGGCTTTCAGGTGTCTGGTT 1680
 533 C  E  P  N  N  H  E  R  Y  Y  G  Y  T  G  A  F  R  C  L  V     552
```

```
1681  GAGAAGGGAGATGTGGCCTTTGTGAAACACCAGACTGTCGAACAGAACACTGATGGACGT  1740
 553  E  K  G  D  V  A  F  V  K  H  Q  T  V  E  Q  N  T  D  G  R    572
1741  AACCCTGACGATTGGGCTAAGGATCTGAAGAGTGAAAACTTTAAGCTGCTATGTCCTGAT  1800
 573  N  P  D  D  W  A  K  D  L  K  S  E  N  F  K  L  L  C  P  D    592
1801  GGCACTAGGAAGTCTGTGACTGAATTCAAAAGCTGCTACCTAGCCCGAGCCCCGAATCAT  1860
 593  G  T  R  K  S  V  T  E  F  K  S  C  Y  L  A  R  A  P  N  H    612
1861  GCTGTAGTCTCACGGAAAGAAAAGGCAGCTTGTGTTTGCCAGGAGTTACACAACCAGCAG  1920
 613  A  V  V  S  R  K  E  K  A  A  C  V  C  Q  E  L  H  N  Q  Q    632
1921  GCTTCGTATGGAAAAAATGGAAGTCACTGCCCAGACAAGTTTTGTTTGTTCCAGTCAGCC  1980
 633  A  S  Y  G  K  N  G  S  H  C  P  D  K  F  C  L  F  Q  S  A    652
1981  ACCAAGGACCTTCTGTTCAGGGATGACACACAATGTTTGGCCAACCTTCAGCCCACAACA  2040
 653  T  K  D  L  L  F  R  D  D  T  Q  C  L  A  N  L  Q  P  T  T    672
2041  ACATATAAAACCTACTTAGGAGAAAAGTATCTCACGGCGGTGGCTAACCTGAGACAGTGC  2100
 673  T  Y  K  T  Y  L  G  E  K  Y  L  T  A  V  A  N  L  R  Q  C    692
2101  TCAACCTCGAGACTGCTGGAAGCCTGCACTTTCCACAGAGTTTAAAGCCCAAGAGGTGGG  2160
 693  S  T  S  R  L  L  E  A  C  T  F  H  R  V                     706
2161  GTCACCACCAGATGGAGATGGGAGCTCACGTGACCCATAAGCTTCCCTCTGCTCTCACTG  2220
2221  GCCTGAGTGGTTTGTCTGCCTTCACAGTTTGGTGGTGGCACCTCTGCAGAACATAAAATA  2280
2281  AAAATTATTGTTGGTTTTATCTTTTAAAAA                                 2310
```

Figure 22. Horse transferrin cDNA sequence and its predicted translation product. The nucleotide sequence is numbered from the first nucleotide sequenced. The amino acid sequence is numbered from the first residue in the signal peptide. The start codon ATG, stop codon TAA and polyadenylation signal AATAAA are underlined. The glycosylation signals NST and NGS, and the signal peptide are written in bold text.

3.4.3 Duplicated structure of horse transferrin

Both the cDNA sequence and its translation product show the duplicated structure characteristic of the transferrin family. The cDNA sequence can be divided into two sections which are 51% identical. Likewise the protein sequence halves are 43% identical. Figure 23 shows an alignment of the N- and C-terminal halves of the amino acid sequence.

3.4.4 Comparison with other transferrins

The horse transferrin cDNA sequence is 80% identical to human transferrin cDNA. The amino acid sequence of horse transferrin is 73% identical to that of human transferrin. Further comparisons of horse transferrin to other members of the transferrin family are presented in the discussion (Section 4.2.2).

3.5 POLYMORPHISM OF HORSE TRANSFERRIN

The positions of the two polymorphic sites in the horse transferrin cDNA and protein, predicted by combining the data of Chung and McKenzie (1985) with the horse transferrin sequence, are shown in Figure 24. These lie in the regions equivalent to exon 12 and exon 15 in human transferrin. As the intron/exon pattern appears to be conserved among chicken and human serum transferrin genes and the mouse lactoferrin gene, it is likely that the same pattern is present in the horse transferrin gene. Therefore, the names exon 12 and 15 will be used for the regions of the horse transferrin cDNA and gene equivalent to exons 12 and 15 in the human transferrin gene.

3.5.1 Exon 12

A polymorphic *Sau*3AI site was predicted in exon 12 of the horse transferrin gene based on the data of Chung and McKenzie (1985). Figure 25 shows two Southern blots of genomic DNA from ten horses of different transferrin phenotypes, cut with *Sau*3AI and hybridized with an exon 12 probe (Section 2.3.1). The probe hybridized to a 0.9 kb fragment when the alleles D, F_1, F_2, H_2 or * were present, and to a 1 kb fragment when O or R alleles were present. These

```
  1 MRLAIRALLACAVLGLCLAEQTVRWCTVSNHEVSKCASFRDSMKSIVPAP 50
       .||: :: .       . ..|:||.:::||  ||..:.       | ..
350 ...DIRPEVPKD......ECKKVKWCAIGHHEKVKCDEWS......VNSG 384

 51 PLVACVKRTSYLECIKAIADNEADAVTLDAGLVFEAGLSPYNLKPVVAEF 100
    . :.| .  |  :||  |...||||:.||:|::: ||  .:.| ||:||
385 GNIECESAQSTEDCIAKIVKGEADAMSLDGGFIYIAG..KCGLVPVLAEN 432

101 YGSKTE......PQTHYYAVAVVKKNS..NFQLNQLQGKKSCHTGLGRSA 142
    |:.:.:      |:. |.||||||..|  :: :| |.||||||||::|.|
433 YETRSGSACVDTPEEGYHAVAVVKSSSDPDLTWNSLKGKKSCHTGVDRTA 482

143 GWNIPIGLLYWQLPEPRESLQKAVSNFFAGSCVPCADRTAVPNLCQLCVG 192
    |||||:||||.:::...      ....|| ::|.|.  |.  ..||.||:|
483 GWNIPMGLLYSEIKHC......EFDKFFREGCAPGYRRN..STLCNLCIG 524

193 KGTD...KCACSNHEPYFGYSGAFKCLADGAGDVAFVKHSTVLEN..... 234
    .:.:   .|...|||.|:||.|||:||.:  |||||||| || :|
525 SASGPGRECEPNNHERYYGYTGAFRCLVE.KGDVAFVKHQTVEQNTDGRN 573

235 ...LPQEADRDEYQLLCRDNTRKSVDEYKDCYLASIPSHAVVARSVDGKE 281
       ::.: ..:::.|||.|.|||||.|:|.||||. |.||||.|  .:|.
574 PDDWAKDLKSENFKLLCPDGTRKSVTEFKSCYLARAPNHAVVSR..KEKA 621

282 DLIWGLLNQAQEHFGTEKS...KDFHLFSSPHGKDLLFKDSALGFLRIPP 328
    . : . |:..|. :|.: |   ..| || |: .|||||:|.. .: .:.|
622 ACVCQELHNQQASYGKNGSHCPDKFCLFQSA.TKDLLFRDDTQCLANLQP 670

329 AMDTWLYLGYEYVTAIRNLRE.............. 349
    . . . ||| .|:||: |||:
671 TTTYKTYLGEKYLTAVANLRQCSTSRLLEACTFHRV 706
```

Figure 23. Duplication in the amino acid sequence o f horse transferrin. This alignment gives the maximum identity of 43% between the halves. Vertical lines mark identical amino acids. Single and double dots between the sequences mark moderately and highly conservative substitutions, respectively. Dots in the sequences mark gaps introduced to align the sequences. This alignment was produced using the Gap program of the GCG package (Devereux *et al.*, 1984).

Exon 12:

```
                                S  S  S D/G P  D  L  T  W
 449    Y  H  A  V  A  V  V  K  S  S  S  D  P  D  L  T  W    465
1367  GGTATCATGCTGTGGCCGTGGTTAAGTCCTCATCAGATCCTGACCTCACCTGG  1419
        5'-ATGCTGTGGCCGTGGTTAAG-3'                  Primer 12/1


      N     L  K
 466  N  S  L  K  G  K  K  S  C  H  T  G  V  D  R  T  A  G   483
1420  AACTCTCTGAAAGGCAAGAAGTCCTGTCACACTGGAGTAGATAGAACCGCCGG  1472


 484   W  N  I  P  M  G  L  L  Y  S  E  I  K  H  C  E  F    500
1473  CTGGAACATCCCCATGGGCCTGCTCTATAGCGAGATCAAGCACTGTGAATTCG  1525
            3'-GGGGTACCCGGACGAGATAT-5'              Primer 12/2
```

Exon 15:

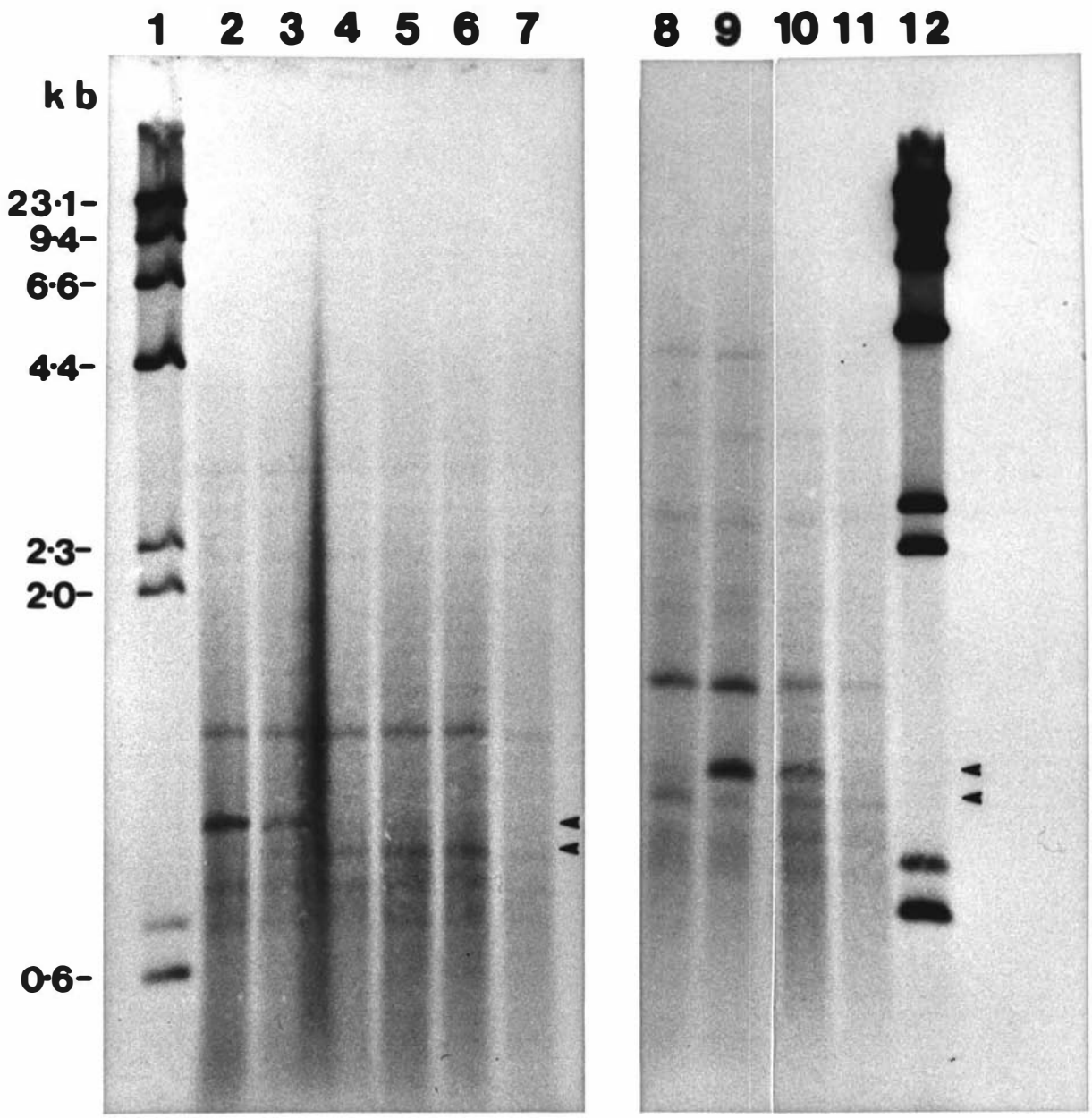
```
                                   K  S E/G N  F
 572    R  N  P  D  D  W  A  K  D  L  K  S  E  N  F  K  L    588
1736  GACGTAACCCTGACGATTGGGCTAAGGATCTGAAGAGTGAAAACTTTAAGCTG  1788
        5'-ACCCTGACGATTGGGCTAAG-3'                  Primer 15/1


 589  L  C  P  D  G  T  R  K  S  V  T  E  F  K  S  C  Y  L    606
1789  CTATGTCCTGATGGCACTAGGAAGTCTGTGACTGAATTCAAAAGCTGCTACCT  1841


 607   A  R  A  P  N  H  A  V  V  S  R  K  E  K  A  A  C  V   624
1842  AGCCCGAGCCCCGAATCATGCTGTAGTCTCACGGAAAGAAAAGGCAGCTTGTG  1894
        3'-CTCGGGGCTTAGTACGACAT-5'                  Primer 15/2


 625    C  Q  E  L  H  N  Q  Q    632
1895  TTTGCCAGGAGTTACACAACCAGCAG  1920
      3'-CGGTCCTCAATGTGTTGGTC-5'                    Primer 15/3
```

Figure 24. Polymorphic sites in exons 12 and 15. The regions of the horse transferrin cDNA sequence equivalent to exons 12 and 15 of the human transferrin gene are shown, together with the predicted translation products of these regions. The peptides shown by Chung and McKenzie (1985) to differ between the D and R variants, are aligned above the translation product. The sequences of the primers used for PCR amplification of these regions are shown below the cDNA sequence. The *Sau*3AI site in exon 12 which was predicted to be polymorphic is shown in bold type. The nucleotides which were predicted to be polymorphic are underlined.

Figure 25. Southern blots of horse genomic DNA digested with *Sau*3AI and hybridized with an exon 12 fragment of the horse transferrin gene. Lanes 2 - 11 contain DNA from horses with the transferrin phenotypes stated below. Polymorphic bands are marked by arrows. The probe hybridizes to a 0.9 kb fragment when the D, F_1, F_2, H_2 or * alleles are present and to a 1 kb fragment when the O or R alleles are present.

Lanes 1 & 12:	λ DNA digested with *Hin*dIII, and end-labelled with ^{32}P (size standard)
Lane 2:	OR
Lane 3:	DR
Lane 4:	DD
Lane 5:	DH_2
Lane 6:	DF_1
Lane 7:	DF_2
Lane 8:	F_2H_2
Lane 9:	DO
Lane 10:	F_1R
Lane 11:	D*

Electrophoresis was in 1% agarose gels, run at 50 V for 12 h. Approximately 6 µg of DNA was loaded per lane. Hybridization was at 60°C, and the final wash was at 60°C in 1x SSC/0.1% SDS.


1 2 3 4 5 6 7
8 9 10 11 12
kb
23·1-
9·4-
6·6-
4·4-
2·3-
2·0-
0·6-

fragments are marked by arrows in Figure 25. This confirmed the presence of the *Sau*3AI site in the D allele, and its absence in the R allele.

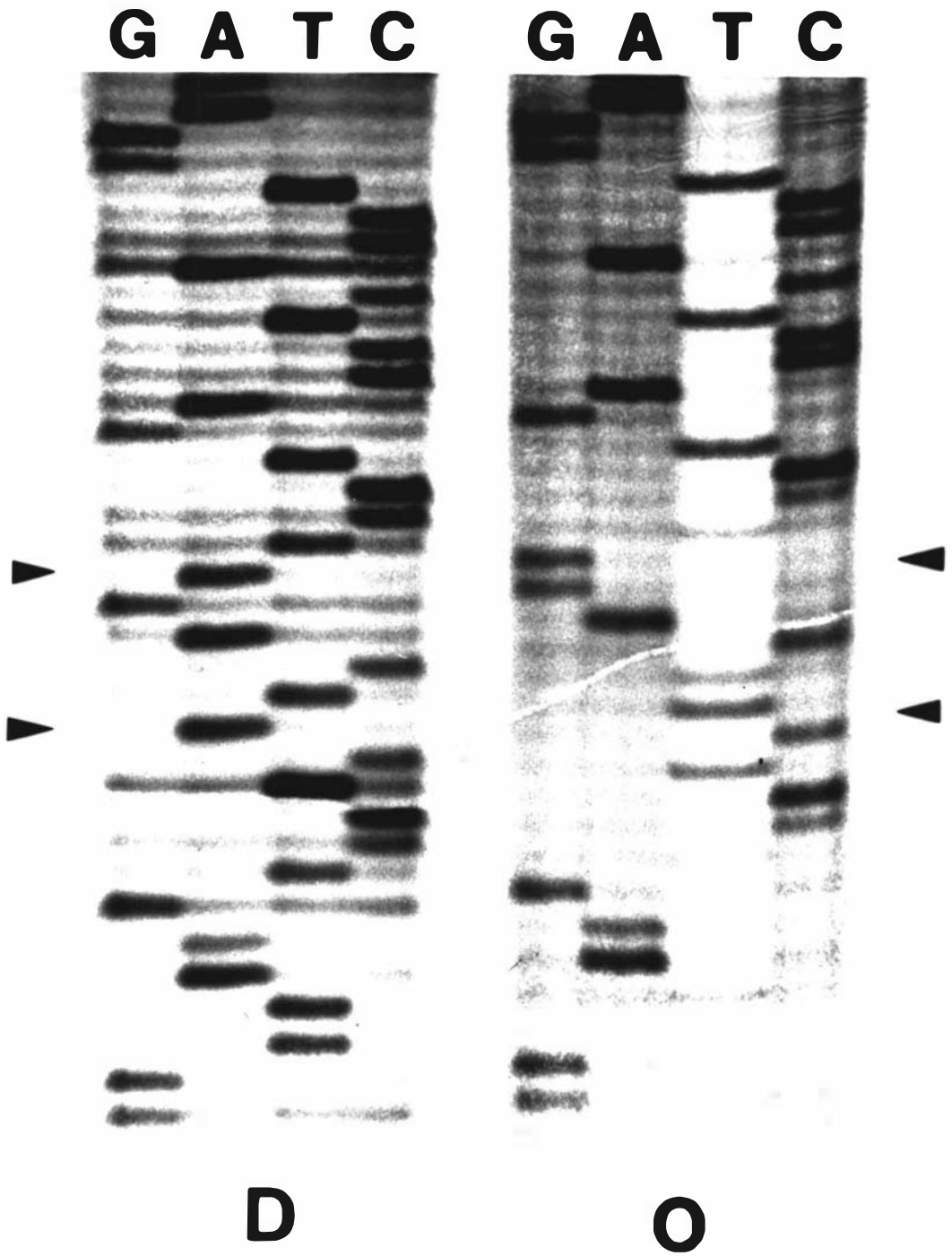
The remaining 0.1 kb fragment in the D, F_1, F_2, H_2 and * alleles was not visible, probably due to its small size. The 1 kb fragment characteristic of O and R alleles consistantly gave a substantially stronger hybridization signal than the 0.9 kb fragment characterising the D, F_1, F_2, H_2 and * alleles. The probe also hybridizes to other fragments, present in all samples, including a 1.3 kb fragment which gives a fairly intense signal. These additional fragments may be due to sequences in the horse genome which are similar to exon 12 of the transferrin gene, such as the equivalent region of the lactoferrin gene, or perhaps a transferrin pseudogene. Alternatively, they could be a result of using PCR to produce the probe. Amplification of the probe by PCR gave products in addition to the one of the required size. Although the probe was purified by isolation from an agarose gel, some contaminating PCR products may still have been present, and their hybridization would produce the extra bands.

As the *Sau*3AI site is present in all of D, F_1, F_2, H_2 and *, the sequences must all be the same at this site, i.e. the same as in the horse transferrin cDNA sequence reported above. The exon 12 region of D alleles from DD, DO and DR samples were sequenced to confirm this. The sequences of the O and R alleles at this site were also determined by sequencing clones from DO, DR and OR samples. In both O and R alleles, an adenine was substituted for a guanine (Figure 26), creating an aspartate to glycine substitution in the protein. This confirms the sequence of the R variant determined by Chung and McKenzie (1985).

In summary, the adenine at position 1403 in the D, F_1, F_2, H_2 and * alleles is replaced by a guanine in O and R. Consequently, the aspartate residue at position 460 in the D, F_1, F_2, H_2 and * variants is replaced by a glycine in the O and R variants (see Table 2, Section 3.5.6).

The DNA sequence of an O allele from a DO individual had an additional synonymous substitution in the cDNA, as shown in Figure 26. This change may have been O-typical, specific to the individual or a PCR artefact. It was not studied further as it had no affect on the protein sequence. As its authenticity was not confirmed, it will not be considered in the discussion.

Figure 26. Polymorphic positions in the sequence of exon 12 of the horse transferrin gene. Arrows mark the polymorphic bands. On the left is a D-type sequence from a DD sample. On the right an O-type sequence from a DO sample. The upper arrow marks an A/G substitution at position 1403 which produces an aspartate/glycine substitution at position 460 in the amino acid sequence. The lower arrow marks a synonymous A/T substitution which was not confirmed in a second sample and could therefore be a PCR artefact.


G A T C
G A T C
D
O

3.5.2 Exon 15

A guanine/adenine substitution was predicted in the sequence of exon 15 based on the glycine/glutamate substitution observed by Chung and McKenzie (1985), which distinguished the D and R variants of horse transferrin. Attempts to confirm the nucleotide substitution by direct sequencing of PCR products gave inconclusive results, as outlined below. Sequencing of cloned PCR products did confirm the predicted difference between the D and R alleles, as well as revealing other differences between the two alleles. In addition, the five other alleles sequenced were found to be identical or very similar to either the D or R allele.

PCR using primers 15/1 and 15/2, as described in Section 2.3.3 and Figure 24, gave products of the expected size (124 bp) for ten samples from horses with the phenotypes DD, DF_1, DF_2, DH_2, DO, DR, D*, F_1R, F_2H_2 and OR. Direct sequencing of these products using one of the amplification primers for sequencing did not give very clear sequence data, but indicated that 9 of the 10 samples were identical to the equivalent region of the complete horse transferrin cDNA sequence, denoted D-type, while the OR sample was different (Figure 27A). These results made little sense as the DR, and perhaps DO and F_1R, samples were expected to appear heterozygous. It was concluded that the results were artefactual, and were due to a bias in the amplification process, such that in DO, DR and F_1R samples, D and F_1 were amplified in preference to O and R. This results from the primer sequences being based on the D-type sequence, and indicates further polymorphism in at least one of the primer regions.

When PCR was used to amplify exon 15 using primers 15/1 and 15/3 (Section 2.3.3 & Figure 24), 9 of the 10 samples gave products of the correct size. The sequences of these 9 samples were identical to the D-type sequence. The sequences of two of the samples, produced by direct sequencing of PCR products using primer 15/2, are shown in Figure 27B. No heterozygosity was observed in the DR sample which was expected to show a polymorphism. The tenth sample, from an individual of phenotype OR, didn't yield a PCR product of the correct size, even when the annealing temperature was lowered to 40°C, and non-specific products were appearing. These results indicated that the PCR was only amplifying D-type sequences, probably due to differences in the primer-binding regions of the D- and R-type alleles.

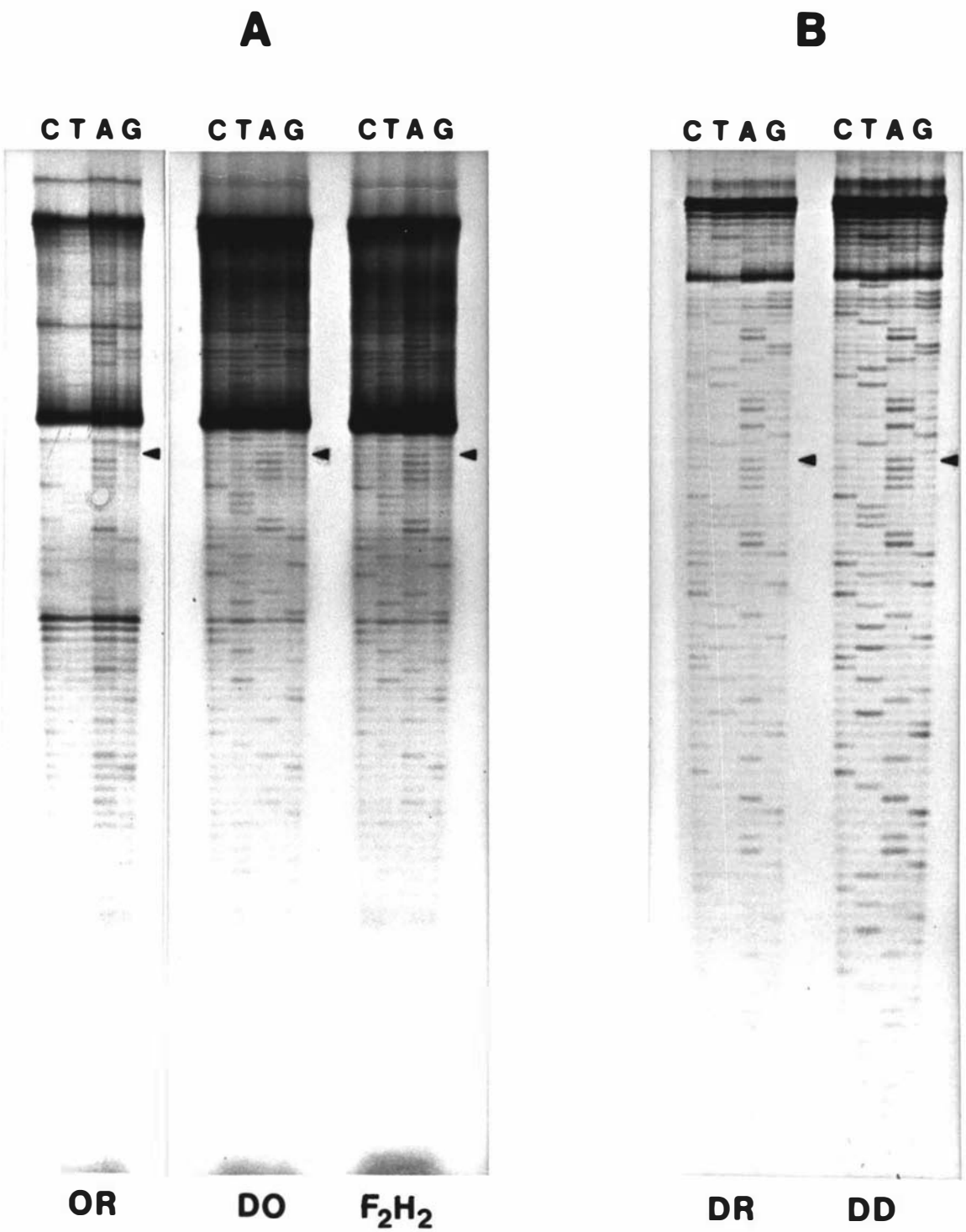
To circumvent these problems with amplification and direct sequencing, PCR products from the primer pair 15/1 and 15/2 were cloned and sequenced. Clones were produced from samples

Figure 27. Direct sequencing of PCR products from exon 15 of the horse transferrin gene.

A. Direct sequencing of PCR products amplified using primers 15/1 and 15/2 and sequenced using primer 15/2.

B. Direct sequencing of PCR products amplified using primers 15/1 and 15/3 and sequenced using primer 15/2.

Arrows mark the positions which were predicted to be polymorphic. The phenotypes of the samples are shown below the sequence. Note that the DO and F_2H_2 sequences in A appear to be identical, whereas the OR sample is different, but the data is not good enough to show this clearly. In B, the sequence can be read more easily, but the DD and DR sequences reveal no differences. It was not possible to sequence an OR sample using a nested primer as there was no PCR product when primers 15/1 and 15/3 were used with the DNA of phenotype OR.

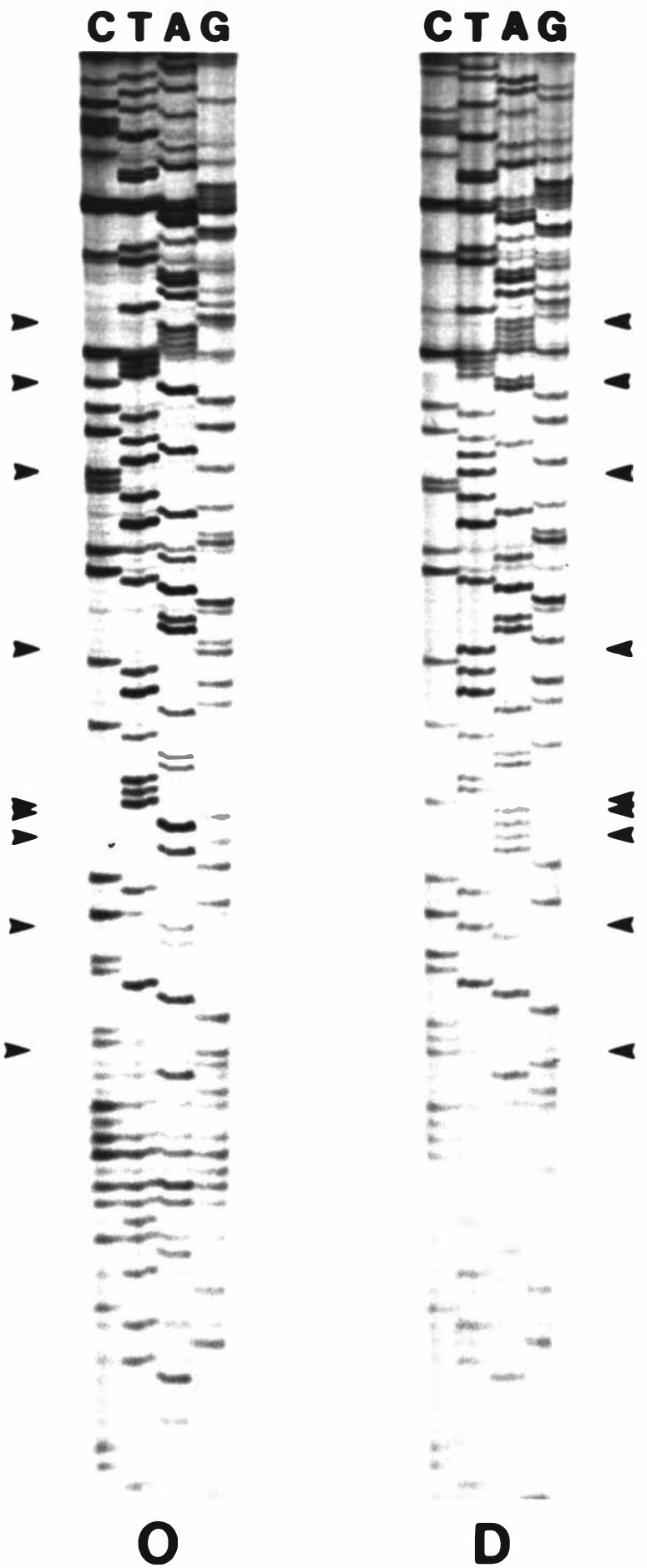

A
B
CTAG
CTAG
CTAG
CTAG
CTAG
OR
DO
F2H2
DR
DD

of six phenotypes, specifically DO, DR, D*, F_1R, F_2H_2, OR. Fortunately the bias in the PCR was not great enough to prevent O or R clones from being isolated from DO, DR and F_1R heterozygotes.

Sequences of clones from D* and F_2H_2 phenotypes were all identical to the D-type sequence shown in Figure 28, indicating that the D, F_2, H_2 and * alleles are identical in the region sequenced. Six clones were sequenced for each of the D* and F_2H_2 samples to reduce the probability that clones with different sequences were present but not observed. The OR sample gave a single sequence type (denoted O-type, shown in Figure 28) from all 6 clones sequenced. DR and F_1R samples produced clones of both D-type and R-type alleles. The R-type differed from the O-type at a single position which was the base adjacent to the 3' primer (primer 15/2). No O-types were isolated from the sample with phenotype DO, but a hybrid D/O (or D/R) was found. This is thought to be a PCR artefact indicating the presence of O-type alleles in this individual. As the DR and F_1R samples yielded an R-type sequence which differed from the only sequence observed in the OR sample, either there are two slightly different R allele sequences, one which is the same as the O allele in the region sequenced, or all six clones from the OR sample were O alleles. It is possible that two variants differing only slightly in their amino acid sequence would have identical electrophoretic mobility, and therefore both be labelled as R variants. It is unlikely, though possible, that six identical clones would be selected at random from a population containing equal quantities of two types of clone. It is also possible that the population of OR clones contained a disproportionate number of O allele clones.

In contrast to the single base substitution expected, the O- and R-type sequences differed from the D-type in 9 positions. Five of these base substitutions cause substitutions in the amino acid sequence. Three of the nucleotide substitutions are in the third position of a codon and do not alter the amino acids encoded. The ninth substitution is at the last base before the 3' primer and is in the first position in the codon. This is the site which differed between the O and R alleles. As the rest of the codon was not sequenced, it is not known what amino acids would be encoded in the O and R variants. As the effect of this substitution on the amino acid sequence is not known, it will not be considered in the discussion. Therefore, the exon 15 sequence for O and R variants will be treated as identical, and referred to as R-type. The polymorphisms observed in exon 15 are summarised in Table 2.

Figure 28. Polymorphic positions in the sequence of exon 15 of the horse transferrin gene. The polymorphic positions are marked by arrows. The sequence on the left is from an OR sample. The sequence on the right is from a DR sample. There are nine base substitutions distinguishing the two sequences. Five of these produce substitutions in the amino acid sequence, while three are synonymous substitutions. The effect of the ninth base substitution is not known as the remainder of its codon was not sequenced as it was replaced by part of the PCR primer 15/2. The polymorphism marked by the uppermost arrows is at position 1775 in the cDNA sequence and the bottom arrows mark a polymorphism at position 1846. The arrows second from top mark a substitution which creates a polymorphic *Pvu*II site and the substitution marked by the fifth pair of arrows produces a polymorphic *Eco*RI site. The non-coding strand was sequenced. The lane labels have been reversed (C/G, A/T) so that the sequence of the coding strand is obtained by reading top to bottom. The results are summarized in Table 2.


C T A G
C T A G
O
D

3.5.3 Detection of single stranded conformation polymorphisms

Further polymorphisms between transferrin types D, F_1 and F_2 were detected using the SSCP method. Transferrin cDNA was amplified by PCR using cDNA from individuals of phenotypes DF_1, DF_2 and F_1F_1. Restriction digests produced the fragments shown in Figure 29, which were electrophoresed to produce SSCP's. Most of the coding sequence (92%) was examined to find differences between D and F_1, while only the 3' 624 bp was used for studying the F_2 allele.

Figure 30A shows the 303 bp *Eco*RI fragments from DF_1 and DF_2 samples which gave positive results in a SSCP gel, together with the F_1F_1 sample as a negative control. The two bands observed in the control sample correspond to the two strands of DNA, while the positive result is conveyed by the presence of three bands. SSCP analysis of a heterozygous sample can reveal four bands, two strands for each of two alleles. When two of these four strands have the same electrophoretic mobility, only three bands will appear, as in the DF_1 and DF_2 samples in Figure 30A.

The *Sau*96AI fragments from a DF_1 sample are shown in Figure 30B. Of these, the 528 bp fragment gave a positive result, while the 411 bp and 269 bp fragments were negative.

The 528 bp *Sau*96I fragment and the 303 bp *Eco*RI fragment both indicated polymorphisms between D and F_1. The *Eco*RI fragment of F_2 was the same as for F_1. The 528 bp fragment did not consistantly appear polymorphic, while the 303 bp one was consistent. Polymorphisms were not readily detected in gels run at room temperature.

3.5.4 Sequence polymorphism in the 528 bp *Sau*96AI fragment

Sequencing of clones containing the 528 bp *Sau*96I fragment from a DF_1 sample revealed two single base substitutions, in two different clones, i.e. each clone differed from the horse transferrin cDNA sequence by one base pair. As only two alleles can be present in an individual, one of the substitutions had to be a PCR artefact (assuming the original sequence was correct). To distinguish the real and artefactual sequences, a 310 bp *Hin*fI fragment which included the two putative polymorphic sites was cloned from a separate PCR. Sequencing the resultant clones confirmed one of the polymorphisms and the original sequence, while failing to show the other polymorphism. Therefore the latter one must have been an artefact.

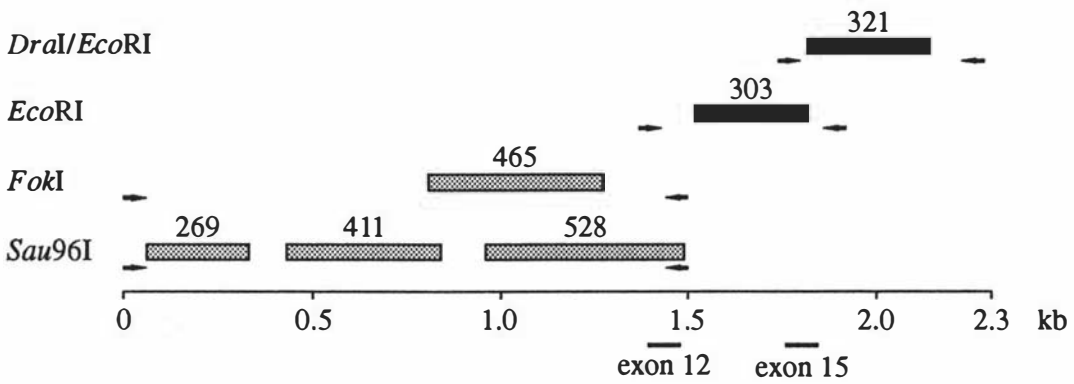


Figure 29. Fragments of horse transferrin cDNA screened for single stranded conformation polymorphisms. Fragments shown in black were produced for DF_1, DF_2 and F_1F_1 samples. Fragments shown in grey were only produced for the DF_1 sample. Fragment sizes in base pairs are shown above fragments. Arrows represent PCR primers. Restriction enzymes used to produce fragments are shown at left. Single stranded conformation polymorphisms were observed for the 528 and 303 bp fragments. The remaining fragments gave negative results. The positions of exons 12 and 15 are shown for reference.

Figure 30. Single stranded conformation polymorphisms.

A. The 303 bp *Eco*RI fragment. The phenotypes of the samples are given below.

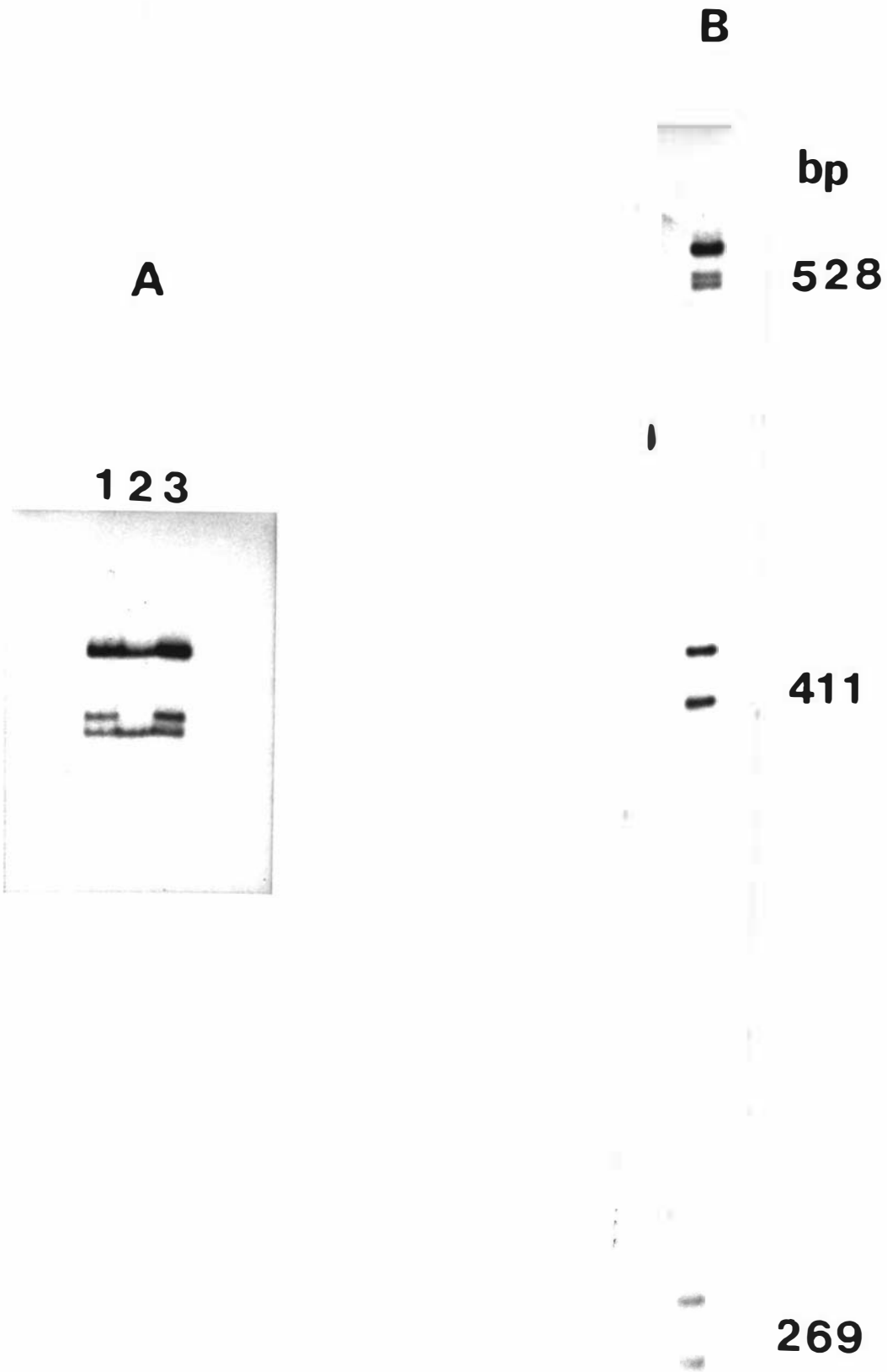
Lane 1: DF_1

Lane 2: F_1F_1

Lane 3: DF_2

The presence of three bands in lanes 1 and 3 indicates a difference between D and F_1 and between D and F_2. The F_1 and F_2 fragments appear to be the same. The two bands in lane 2 indicate an absence of polymorphism as was expected for the F_1F_1 sample.

B. *Sau*96I fragments from a DF_1 sample. The sizes of the fragments in base pairs is shown. The 528 bp fragment is polymorphic as indicated by the presence of three bands.


A
1 2 3
B
bp
528
411
269

The genuine polymorphism located in this fragment was a cytosine/thymine substitution at position 1333, as shown in Figure 31. This results in a serine/proline substitution at position 437 in the protein (Table 2). These results do not determine which is the D form and which is F_1.

3.5.5 Sequence polymorphism in the 303 bp *Eco*RI fragment

Sequencing of clones containing the 303 bp *Eco*RI fragment enabled two types of sequence to be identified. The D-type sequence was found in the DF_1 and DF_2 samples, whereas the F-type was found in DF_1, DF_2 and F_1F_1 samples. Therefore F_1 and F_2 are identical in this region. The two types differed by two single base substitutions, as illustrated in Figure 32. One was a cytosine/thymine substitution at position 1656, which was a synonymous mutation. The other was a cytosine/thymine substitution at position 1738 which replaces the cysteine at position 572 in D with an arginine in F_1 and F_2 (Table 2).

3.5.6 Summary of polymorphisms analysed

The nature and positions of the polymorphisms identified in horse transferrin alleles are summarised in Table 2. A total of 13 different bases in the cDNA were variable, producing amino acid substitutions at 8 positions. A ninth amino acid substitution is likely at position 608, indicated by a base substitution in the first codon position. However, as the sequence of the rest of that codon is not known, the identity of the replacement amino acid is not certain. The positions of the polymorphisms in the cDNA and amino acid sequences of horse transferrin are depicted in Figure 33.

There are two indications that these are not all the polymorphisms which occur between the variants of horse transferrin. Firstly, further polymorphisms must exist as those found were not sufficient to distinguish all the variants. Secondly, the behaviour of the PCR when amplifying exon 15 provides further evidence for polymorphism within the primer regions (Section 3.5.2).

Figure 31. Polymorphism in the sequence of the 528 bp *Sau*96I fragment. The sequences of two clones from a DF_1 sample are shown. It is not known which is D and which is F_1. The arrows mark the position of a C/T substitution at position 1333 of the cDNA sequence. This produces a proline/serine substitution at position 437 in the amino acid sequence.

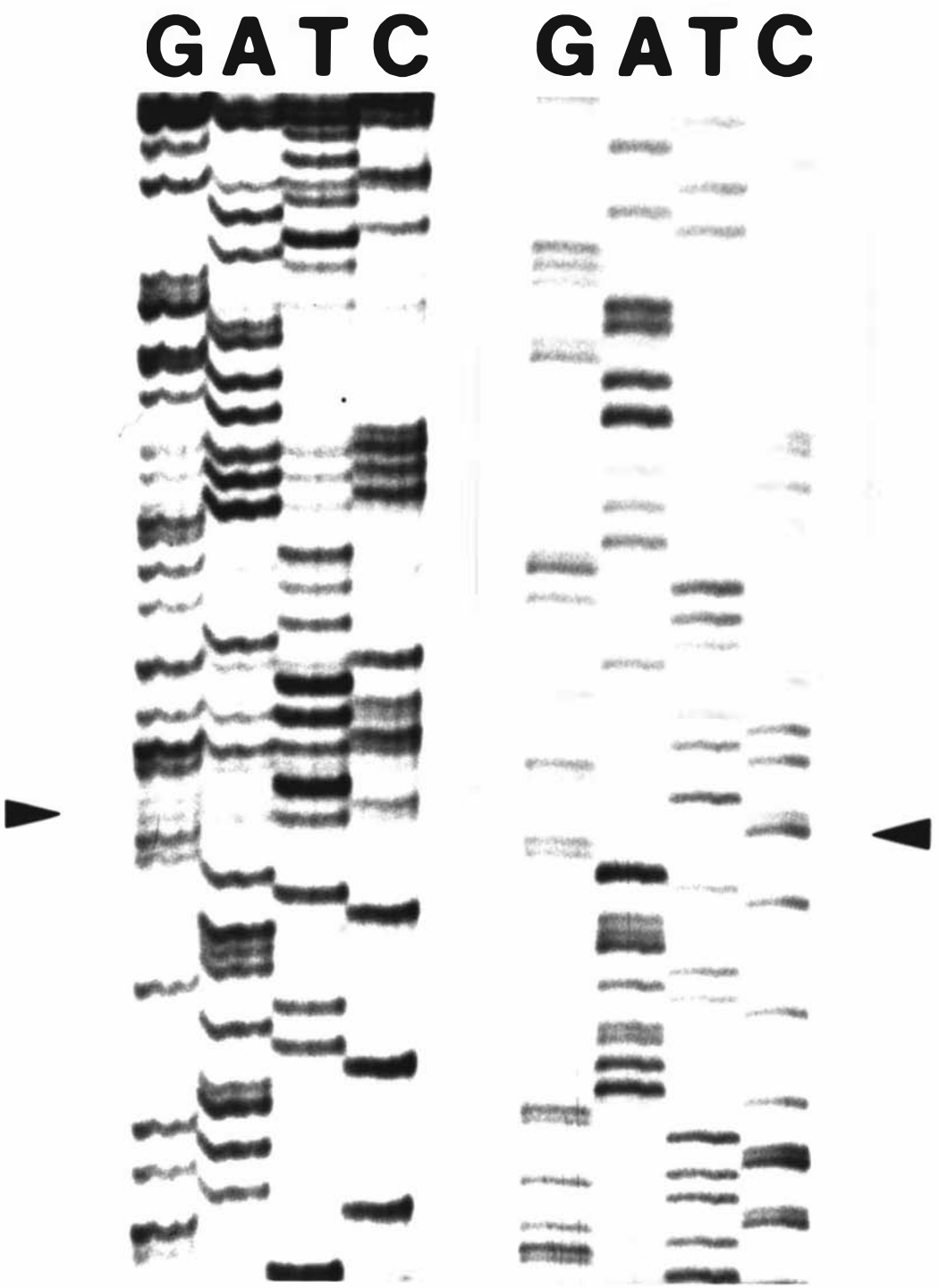
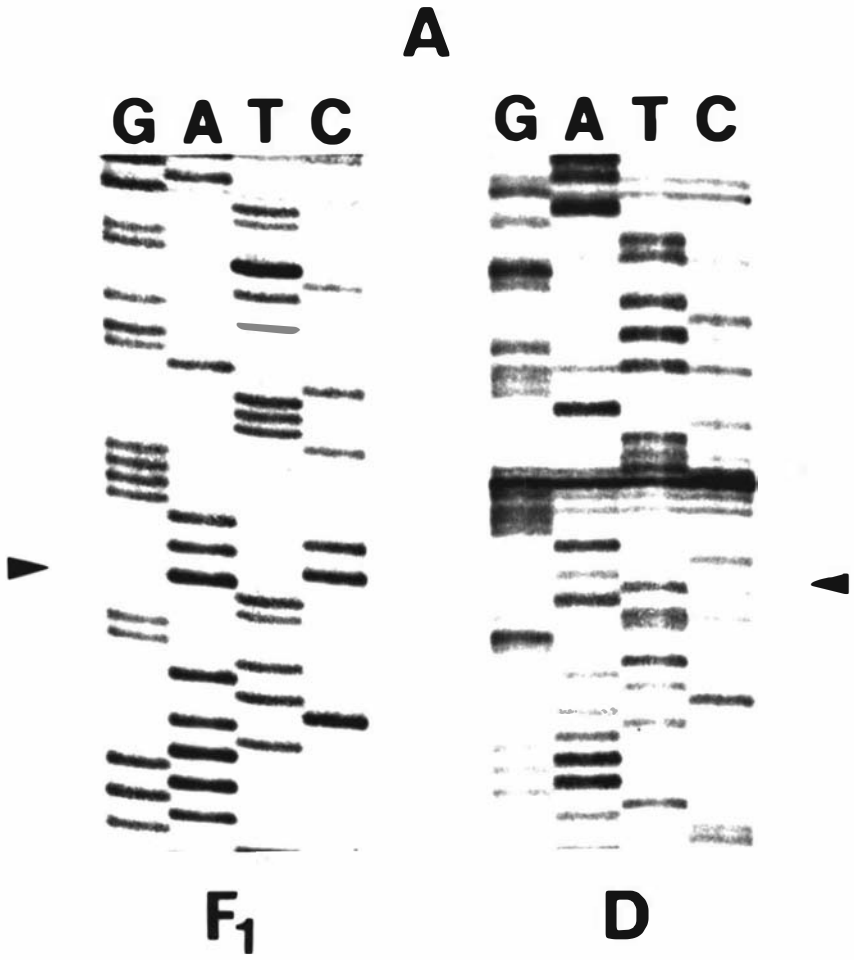

G A T C
G A T C

Figure 32. Polymorphisms in the sequence of the 303 bp *Eco*RI fragment. The sequences on the left are F_1 sequences from a DF_1 sample. On the right are D sequences from a DF_2 sample. Arrows mark the positions of polymorphic bases.

A. The cytosine at position 1656 in F_1 is replaced by a thymine in the D allele. This is a synonymous substitution.

B. The cytosine at position 1738 in F_1 is replaced by a thymine in D. As a result, the arginine at position 572 in the amino acid sequence is replaced by a cysteine.


A
G A T C
G A T C
F1
D

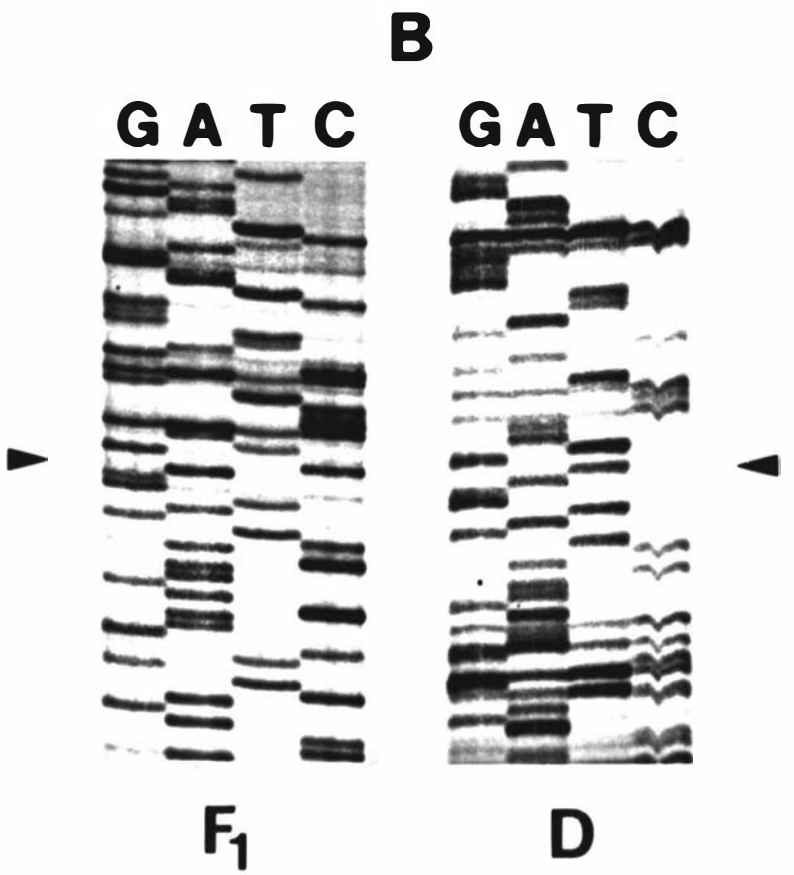

B
G A T C
G A T C
F1
D

Table 2

Polymorphisms in the cDNA and Protein Sequences of Horse Transferrin

Transferrin variants	Change in codon	Change in amino acid	Position in cDNA sequence	Position in protein sequence	Fragment sequenced
D/F_1 (a)	TCT	S	1333	437	SSCP 528
D/F_1 (a)	CCT	P			
D,F_1,F_2,H_2,*	GAT	D	1403	460	exon 12
O,R	GGT	G			
D	TAT	Y	1656	544	SSCP 303
F_1,F_2	TAC	Y			
D	TGT	C	1738	572	SSCP 303
F_1,F_2	CGT	R			
D,F_1,F_2,H_2,*	GAA	E	1775	584	exon 15
O,R	GGA	G			
D,F_1,F_2,H_2,*	AAG	K	1783	587	exon 15
OR	CAG	Q			
D,F_1,F_2,H_2,*	TGT	C	1794	590	exon 15
O,R	TGC	C			
D,F_1,F_2,H_2,*	TCT	S	1813	597	exon 15
O,R	GCT	A			
D,F_1,F_2,H_2,*	TTC	F	1827	601	exon 15
O,R	TTT	F			
D,F_1,F_2,H_2,*	AAA	K	1828	602	exon 15
O,R	GAG	E			
D,F_1,F_2,H_2,*	AAA	K	1830	602	exon 15
O,R	GAG	E			
D,F_1,F_2,H_2,*	TAC	Y	1837	605	exon 15
O,R	AAC	N			
D/F_1	CGA	R	1846	608	exon 15
F_2,H_2,* (b)	C??	R?			
O (b)	G??	G?			
R (b)	A??	R?			

(a) A nucleotide substitution was observed between clones from a heterozygous DF_1 sample. It is not known which form is D and which is F_1.

(b) Only the first nucleotide of this codon was sequenced, as the other two were in the primer sequence. The amino acids shown are correct if there are no other changes in the codon.

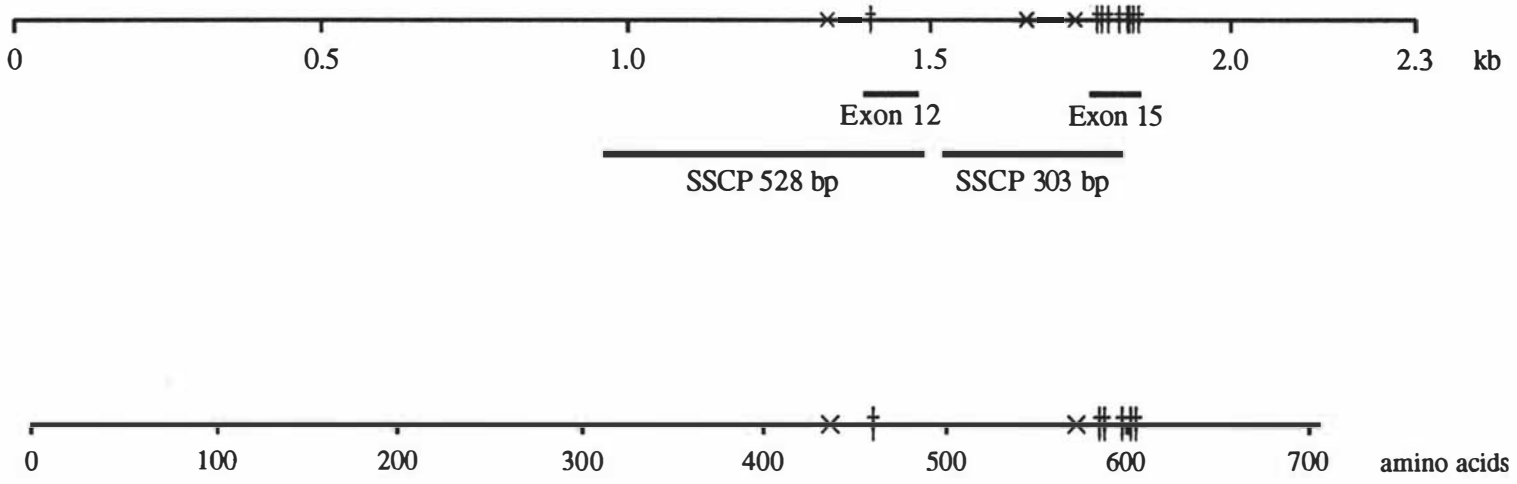


Figure 33. Positions of polymorphisms in the cDNA and amino acid sequences of horse transferrin. Differences between the D and R groups of transferrins are represented by a dagger. Differences between the D and F_1 variants are represented by a cross. Those substitutions shown in the nucleotide sequence but not in the amino acid sequence are synonymous substitutions. The positions of the fragments which were sequenced to identify these sequence polymorphisms are also shown.

CHAPTER 4: DISCUSSION AND CONCLUSIONS

4.1 EXPERIMENTAL CONSIDERATIONS

4.1.1 Quality of the cDNA library

No full length transferrin cDNAs were isolated from the horse liver cDNA library, despite the fact that the cDNA synthesis reaction produced cDNAs of a large enough size (Figure 9). None of the clones isolated at random from the library had internal *Eco*RI sites (Figure 13), nor did the transferrin clones. The 300 bp transferrin clones had *Eco*RI ends which were not derived from linkers (i.e. they had GAATTCG and TGAATTC, rather than the GAATTCC and GGAATTC of the linkers), and must therefore have been *Eco*RI sites within the transferrin cDNA. These facts indicate that that the *Eco*RI methylation step did not work efficiently, if at all. Consequently, any full length transferrin cDNAs were cleaved twice by *Eco*RI, and only fragments were incorporated into the library. The control methylation of λ DNA did confer protection from digestion with *Eco*RI (Figure 10). Possibly the cDNA sample contained residues from the synthesis reaction which impaired the methylase enzyme. For example, *Eco*RI methylase is inhibited by $MgCl_2$, a component of cDNA synthesis buffers.

The prevalence of the 300 bp transferrin cDNA fragment over the 5' and 3' fragments (Section 3.2) can be explained by the fact that it is the central *Eco*RI fragment in the complete cDNA, and that the *Eco*RI digest must have had a greater efficiency than the linker ligation. Therefore there would be more clones containing the central fragment with two *Eco*RI ends, than either of the end fragments. It is not clear why the linker ligation might not have worked efficiently.

The lack of 3' fragments in the library is difficult to explain. cDNA synthesis begins at the poly(A) tail, so generally the 3' end is the easiest to isolate. The possibility of reduced homology between horse and human transferrins in this part of the molecule was considered, but screening at lower stringency did not reveal positive clones. The homology in the 3' region (bases 1823-2305) between horse and human transferrin cDNAs is lower than in the rest of the molecule (76 % compared to 80 %) but this difference is not sufficient to explain why this region could not be isolated from the library. It is possible that the library did not include a clone containing the 3' fragment of horse transferrin, due to poor ligation of linkers.

4.1.2 Size of horse transferrin cDNA

The size of the horse transferrin cDNA at 2305 bp (Figure 22) was considerably smaller than the 2800 bp predicted by Northern blotting (Figure 7). Even with the addition of the remainder of the 5' untranslated region (another 24 bp if it is the same size as the 48 bp in human transferrin, Adrian *et al.*, 1986), and the poly(A) tail (100-300 bp), the size would still only be around 2630 bp. The discrepancy between this figure and the 2800 bases of the mRNA determined by Northern blotting is probably due to experimental error. The size estimate would not be very accurate as the only size standards used were the two rRNA bands.

4.1.3 Polymerase chain reaction

The PCR is a very powerful technique and has the potential to save time by avoiding more circuitous approaches. However it does have limitations which must be considered. PCR using *Taq* polymerase has a fidelity problem. Misincorporations have been reported to occur at frequencies from 1/400 to 1/4000 (Higuchi, 1989). This is not a problem if the starting population of target DNA is large, and the products are not going to be cloned, as the consensus of the products will be correct. Thus, in this study, the fidelity problem was avoided when PCR products were used for direct sequencing, for SSCP analysis and as probes for Southern hybridization. Sequencing of cloned products requires that clones generated from separate PCRs are sequenced. This precaution was taken in most cases in this work, sometimes through sequencing clones carrying the same allele from two different individuals. Clones were only sequenced from a single PCR when confirming that a transferrin allele was identical to another allele sequenced previously. In these cases several clones from the one reaction were sequenced.

Another type of artefact can be created by 'jumping PCR' (Paabo *et al.*, 1989) in which the product is a hybrid molecule generated from more than one template e.g. two different alleles in a heterozygous sample. Two clones of this type were detected during sequencing of exon 15 clones from DO and DR samples (Section 3.5.2).

Successful use of the PCR depends largely on the choice of primers. The importance of the 3' end of the primer was highlighted during this study. One of the exon 12 primers (12/2) had 4 G's at the 3' end (Figure 24), which may have allowed the primer to anneal readily to non-

target sequences. As a result, when amplifying genomic DNA, several products were observed in addition to the one of the expected size. Even after isolation of the desired product from an agarose gel prior to cloning, several incorrect clones were sequenced. Thus it appeared that the specificity of the primer was diminished by the presence of a G- (or GC) rich 3' end.

In using PCR to amplify DNA for studying variation between alleles, it was necessary to try to choose primers from conserved regions of the sequence, under the assumption that regions which were conserved between species would also be conserved between alleles within a species. Two of the exon 15 primers were chosen with this factor in mind, but the third primer (15/3) sequence was from the extreme 3' end of the exon (Figure 24), where the sequence is variable between species. This may explain why the PCR failed to amplify the O and R variants when this primer was used. This result indicated further polymorphism between the D-group (D, F_1, F_2, H_2 and *) and R-group (O and R) within this primer region.

4.1.4 Direct sequencing of PCR products

Direct sequencing of PCR products was attempted during analysis of the exon 15 region. Sequencing using one of the amplification primers as a sequencing primer produced rather poor results (Figure 27A). By using three primers, two for amplification and the third for priming the sequencing reaction, the results were much improved (Figure 27B). This is likely to be due to an increase in specificity, as non-specific PCR products would not be sequenced. This approach was based on work of Engelke *et al.* (1988). However other researchers achieve direct sequencing of PCR products using amplification primers, eg. Cawthon *et al.* (1990).

4.1.5 Single stranded conformation polymorphisms

The SSCP technique proved to be a quick and easy way to detect polymorphisms. However the technique does have its limitations; not every polymorphism is detected. Two different fragments may run in the same position by chance, or a minor difference in sequence may not necessarily produce a substantial conformational change. The chances of observing a polymorphism decrease with increasing size of the fragment analysed. Hayashi (1991) estimated a probability of observing a mobility shift in at least one strand as 99% for 100-300 bp fragments, and 89% for 300-450 bp fragments.

In the work presented here, the 528 bp *Sau*96AI fragment only once showed a mobility shift in one strand (Figure 30). Attempts to reproduce this result failed, yet a single base substitution was detected by sequencing. The fragment size may have been too large for reliable use of the SSCP technique.

4.1.6 Blood typing

The horse blood samples from which DNA was extracted were assigned transferrin types at the EBTRC. The polyacrylamide gel electrophoresis used for transferrin typing does not eliminate the possibility of two different variants having the same electrophoretic mobility, and therefore appearing to be the same variant. As the work presented here relied on blood typing for identifying transferrin alleles, it is possible that the amino acid sequence reported as being characteristic of a particular variant might not be correct for all samples. Consequently, contradictory results are possible. One possible example of this situation is the last nucleotide sequenced in the exon 15 region. This was cytosine in all D-type alleles sequenced (D, F_1, F_2, H_2 and *), adenine in the R alleles from F_1R and DR samples, and guanine in all 6 clones from the OR sample. It is possible, though improbable, that the 6 OR clones were all O alleles, otherwise the apparent R variant is actually two different variants.

4.2 HORSE TRANSFERRIN cDNA AND PROTEIN SEQUENCES

4.2.1 Horse transferrin

The residues 20-29 of the horse transferrin amino acid sequence (Figure 22) match the first ten amino acids of the mature protein, determined by Chung and McKenzie (1985) by peptide sequencing. Therefore the derived protein sequence includes a signal peptide of 19 amino acids. The two small peptides, produced by tryptic and chymotryptic digests of the D variant of horse transferrin, which were sequenced by Chung and McKenzie (1985), correspond to positions 457-469 and 582-586 in the translation product, as shown in Figure 24. The sequence reported here has an insertion of a serine at position 467 compared to Chung and McKenzie's first peptide. Chung and McKenzie (1985) predicted that their second peptide (KSENF/KSGNF) was equivalent to residues 614-618 (CSGNF) of human transferrin. This is a region of low homology between horse and human transferrins, and the sequence of the

equivalent region in the horse transferrin (CPDKF) is quite different from the peptide sequence, so their prediction must be incorrect. The C-terminal residue of horse transferrin is valine, in agreement with the results of Stratil *et al.* (1984).

Peptides from the D variant of horse transferrin have been sequenced by Dr. Chris Moore and Geoff Patterson of the Department of Chemistry and Biochemistry, Massey University (personal communication). Their data confirm the horse transferrin primary structure shown in Figure 22, with two exceptions. A cysteine residue was present at position 572 in place of arginine. This confirms the arginine/cysteine substitution identified between the D and F_1 variants of horse transferrin, and indicates that part of the complete sequence (those residues derived from the 300 bp clone from the cDNA library) is from an F_1 allele. As the complete sequence was derived from several clones all from a DF_1 heterozygote, the sequence may be a combination of D and F_1 alleles. The second difference between the peptide sequencing data and the sequence derived from cDNA sequencing was an apparent deletion of asparagine 515. As this residue is present in a NST sequence, one of the signals for N-linked glycosylation of asparagine, its failure to be detected by peptide sequencing is probably due to that fact that it is glycosylated.

The molecular weight of the mature horse transferrin calculated from the amino acid sequence to be 76,104 (Section 3.4.2) does not include the carbohydrate chains. Stratil *et al.* (1984) determined that the carbohydrate content of horse transferrin was 2.58-5.16%. Combining these figures gives a molecular weight of approximately 78,000-80,000 for horse transferrin, which is consistent with the results of Penhallow *et al.* (1991), Hudson *et al.* (1973) and Stratil *et al.* (1984) (Section 1.10).

4.2.2 Comparison of horse transferrin to other transferrin family sequences

The decision to use the human transferrin cDNA as a probe for isolating the horse transferrin cDNA was based on the assumption that there was substantial homology between the two cDNA sequences. Comparison of the cDNA sequences confirmed this assumption when it revealed 80% identity (Section 3.4.4).

A multiple alignment of the horse transferrin amino acid sequence with the sequences of other members of the transferrin family is shown in Figure 34. The sequences were aligned using

```
Horse TF      1 MRLAIRALLACAVLGLCLA--E-QTVRWCTVSNHEVSKCASFRDSMKSIV
Pig TF          -------------------VAQ-KTVRWCTISNQEANKCSSFRENMSKAV
Human TF        MRLAVGALLVCAVLGLCLAVPD-KTVRWCAVSEHEATKCQSFRDHMKSVI
Rabbit TF       MRLAAG-LLACAALGLCLAVTE-KTVRWCAVNDHEASKCANFRDSMKKVL
Human LF        MKLVFLVLLFLGALGLCLAGRRRRSVQWCAVSQPEATKCFQWQRNMRKV-
Mouse LF        MRLLIPSLIFLEALGLCLA--KATTVRWCAVSNSEEEKCLRWQNEMRKV-
Pig LF          MKLFIPALLFLGTLGLCLAAPK-KGVRWCVISTAEYSKCRQWQSKIRR--
Bovine LF       MKLFVPALLSLGALGLCLAAPR-KNVRWCTISQPEWFKCRRWQWRMKKL-
Chicken TF      MKLILCTVLSLGIAAVCFAAPPKSVIRWCTISSPEEKKC----NNLRDLT
Xenopus TF      MDFSLRVALCLSMLALCLAIQKEKQVRWCVKSNSELKKC----KDLVDTC
Human MTF       MRGPSGALWLLLALRTVLGG---MEVRWCATSDPEQHKCGNMSEAFRE--
                                          C         C

Horse TF     48 PA-PPLVACVKRTSYLECIKAIADNEADAVTLDAGLVFEAGLSPYNLKPV
Pig TF          KN-GPLVSCVKKSSYLDCIKAIRDKEADAVTLDAGLVFEAGLAPYNLKPV
Human TF        PSDGPSVACVKKASYLDCIRAIAANEADAVTLDAGLVYDAYLAPNNLKPV
Rabbit TF       PEDGPRIICVKKASYLDCIKAIAAHEADAVTLDAGLVHEAGLTPNNLKPV
Human LF        --RGPPVSCIKRDSPIQCIQAIAENRADAVTLDGGFIYEAGLAPYKLRPV
Mouse LF        --GGPPLSCVKKSSTRQCIQAIVTNRADAMTLDGGTMFDAGKPPYKLRPV
Pig LF          ---TNPMFCIRRASPTDCIRAIAAKRADAVTLDGGLVFEA--DQYKLRPV
Bovine LF       --GAPSITCVRRAFALECIPGIAEKKADAVTLDGGMVFEAGRDPYKLRPV
Chicken TF      QQERISLTCVQKATYLDCIKAIANNEADAISLDGGQVFEAGLAPYKLKPI
Xenopus TF      KNKEIKLSCVEKSNTDECSLLFRKTMQMQFVWTGGDVYKGSLQPYNLKPI
Human MTF       AGIQPSLLCVRGTSADHCVQLIAAQEADAITLDGGAIYEAGKEH-GLKPV
                        C         C               *

Horse TF     97 VAEFYGSKTEPQTHYYAVAVVKKNSNFQLNQLQGKKSCHTGLGRSAGWNI
Pig TF          VAEFYGQKDNPQTHYYAVAVVKKGSNFQWNQLQGKRSCHTGLGRSAGWII
Human TF        VAEFYGSKEDPQTFYYAVAVVKKDSGFQMNQLRGKKSCHTGLGRSAGWNI
Rabbit TF       VAEFYGSKENPKTFYYAVALVKKGSNFQLNELQGKKSCHTGLGRSAGWNI
Human LF        AAEVYGTERQPRTHYYAVAVVKKGGSFQLNELQGLKSCHTGLRRTAGWNV
Mouse LF        AAEVYGTKEQPRTHYYAVAVVKNSSNFHLNQLQGLRSCHTGIGRSAGWKI
Pig LF          AAEIYGTEENPQTYYYAVAVVKKGFNFQ-NQLQGRKSCHTGLGRSAGWNI
Bovine LF       AAEIYGTKESPQTHYYAVAVVKKGSNFQLDQLQGRKSCHTGLGRSAGWII
Chicken TF      AAEIYEHTEGSTTSYYAVAVVKKGTEFTVNDLQGKNSCHTGLGRSAGWNI
Xenopus TF      MAENYGSHTETDTCYYAVAVVKKSSKFTFDELKDKKSCHTGIGKTAGWNI
Human MTF       VGEVYD--QEVGTSYYAVAVVRRSSHVTIDTLKGVKSCHTGINRTVGWNV
                         *                          C *   * **

Horse TF    147 PIGLL--Y--WQLPE-PRESLQKAVSNFFAGSCVPCADRTAVPN-LCQLC
Pig TF          PMGLL--Y--DQLPE-PRKPIEKAVASFFSSSCVPCADPVNFPK-LCQQC
Human TF        PIGLL--Y--CDLPE-PRKPLEKAVANFFSGSCAPCADGTDFPQ-LCQLC
Rabbit TF       PIGLL--Y--CDLPE-PRKPLEKAVASFFSGSCVPCADGADFPQ-LCQLC
Human LF        PTGTLRPF--LNWTG-PPEPIEAAVARFFSASCVPGADKGQFPN-LCRLC
Mouse LF        PIGTLRPY--LNWNG-PPASLEEAVSKFFSKSCVPGAQKDRFPN-LCSSC
Pig LF          PIGLLRRF--LDWAG-PPEPLQKAVAKFFSQSCVPCADGNAYPN-LCQLC
Bovine LF       PMGILRPY--LSWTE-SLEPLQGAVAKFFSASCVPCIDRQAYPN-LCQLC
Chicken TF      PIGTLLHWGAIEWEGIESGSVEQAVAKFFSASCVPGATIEQ--K-LCRQC
Xenopus TF      IIGLLLERKLLKWAGPDSETWRNAVSKFFKASCVPGAKE---PK-LSQLC
Human MTF       PVGYLVESGRLSVMGC---DVLKAVSDYFGGSCVPGAGETSYSESLCRLC
                          C                     C  C          C  C
```

```
Horse TF     191 VGK--GTDKCACSNHEPYFGYSGAFKCLADGAGDVAFVKHSTVLENLPQE
Pig TF           AGK--GAEKCACSNHEPYFGYAGAFNCLKEDAGDVAFVKHSTVLENLPDK
Human TF         PG-------CGCSTLNQYFGYSGAFKCLKDGAGDVAFVKHSTIFENLANK
Rabbit TF        PG-------CGCSSVQPYFGYSGAFKCLKDGLGDVAFVKQETIFENLPSK
Human LF         AGT--GENKCAFSSQEPYFSYSGAFKCLRDGAGDVAFIRESTVFEDLSDE
Mouse LF         AGT--GANKCASSPEEPYSGYAGALRCLRDNAGDVAFTRGSTVFEELPNK
Pig LF           IGK--GKDKCACSSQEPYFGYSGAFNCLHKGIGDVAFVKESTVFENLPQK
Bovine LF        KGE--GENQCACSSREPYFGYSGAFKCLQDGAGDVAFVKETTVFENLPEK
Chicken TF       KGD--PKTKCA--RNAPYSGYSGAFHCLKDGKGDVAFVKHTTVNENAPDL
Xenopus TF       AGI--KEHKCSRSNNEPYYNYAGAFKCLQDDQGDVAFVKQSTVPEEF---
Human MTF        RGDSSGEGVCDKSPLERYYDYSGAFRCLAEGAGDVAFVKHSTVLENTDGK
                          C C        *     C

Horse TF     239 ADR--------DEYQLLCRDNTRKSVDEYKDCYLASIPSHAVVARSVDGK
Pig TF           ADR--------DQYELLCRDNTRRPVDDYENCYLAQVPSHAVVARSVDGQ
Human TF         ADR--------DQYELLCLDNTRKPVDEYKDCHLAQVPSHTVVARSMGGK
Rabbit TF        DER--------DQYELLCLDNTRKPVDEYEQCHLARVPSHAVVARSVDGK
Human LF         AER--------DEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSVNGK
Mouse LF         AER--------DQYKLLCPDNTWKPVTEYKECHLAQVPSHAVVSRSTNDK
Pig LF           ADR--------DKYELLCPDNTRKPVEAFRECHLARVPSHAVVARSVNGK
Bovine LF        ADR--------DQYELLCLNNSRAPVDAFKECHLAQVPSHAVVARSVDGK
Chicken TF       ND----------EYELLCLDGSRQPVDNYKTCNWARVAAHAVVARD-DNK
Xenopus TF       --H--------KDYELLCPDNTRKSIKEYKNCNLAKVPAHAVLTRGRDDK
Human MTF        TLPSWGQALLSQDFELLCRDGSRADVTEWRQCHLARVPAHAVVVRA-DTD
                                    C             C       *

Horse TF     281 EDLIWGLLNQAQEHFGTEKSKDFHLFSSPHG------KDLLFKDSAL-GF
Pig TF           EDSIWELLNQAQEHFGRDKSPDFQLFSSSHG------KDLLFKDSAN-GF
Human TF         EDLIWELLNQAQEHFGKDKSKEFQLFSSPHG------KDLLFKDSAH-GF
Rabbit TF        EDLIWELLNQAQEHFGKDKSGDFQLFSSPHG------KNLLFKDSAY-GF
Human LF         EDAIWNLLRQAQEKFGKDKSPKFQLFGSPSGQ-----KDLLFKDSAI-GF
Mouse LF         EEAIWELLRQSQEKFGKKQASGFQLFASPSGQ-----KDLLFKESAI-GF
Pig LF           ENSIWELLYQSQKKFGKSNPQEFQLFGSPGQQ-----KDLLFRDATI-GF
Bovine LF        EDLIWKLLSKAQEKSGKNKSRSFQLFGSPPGQ-----RDLLFKDSAL-GF
Chicken TF       VEDIWSFLSKAQSDFGVDTKSDFHLFGPPGKKD-PVLKDFLFKDSAI-ML
Xenopus TF       SKDIIEFLQEAQK---TQECKLFRLPGMGKGSNFQGQRSESYSPPIFYGQ
Human MTF        GGLIFRLLNEGQRLFSHEGSS-FQMFSS---EAY-GQKDLLFKDSTS-EL

Horse TF     324 LRIPPAMDTWLYLGYEYVTAIRNLREDIRPEVPKDECKKVKWCAIGHHEK
Pig TF           LKIPSKMDSSLYLGYQYVTALRNLREEISPDSSKNECKKVRWCAIGHEET
Human TF         LKVPPRMDAKMYLGYEYVTAIRNLREGTCPEAPTDECKPVKWCALSHHER
Rabbit TF        FKVPPRMDANLYLGYEYVTAVRNLREGICPDPLQDECKAVKWCALSHHER
Human LF         SRVPPRIDSGLYLGSGYFTAIQNLRKS--EEEVAARRARVVWCAVGEQEL
Mouse LF         VRVPQKVDVGLYLTFSYTTSIQNLNKK--QQDVIASKARVTWCAVGSEEK
Pig LF           LKIPSKIDSKLYLGLPYLTAIQGLRET--AAEVEARQAKVVWCAVGPEEL
Bovine LF        LRIPSKVDSALYLGSRYLTTLKNLRET--AEEVKARYTRVVWCAVGPEEQ
Chicken TF       KRVPSLMDSQLYLGFEYYSAIQSMRKDQLTPSPREN--RIQWCAVGKDEK
Xenopus TF       FSVPR---SRLF------QCIQALKEGVKEDDSAAQ-VKVRWCTQSKAEK
Human MTF        VPIATQT-YEAWLGHEYLHAMKGL---LCDPNRLPPY--LRWCVLSTPEI
                                              C       C     C
```

```
Horse TF    374 VKCDE-----WSVNSGGNIECESAQSTEDCIAKIVKGEADAMSLDGGFIY
Pig TF          QKCDA-----WSINSGGKIECVSAENTEDCIAKIVKGEADAMSLDGGYIY
Human TF        LKCDE-----WSVNSVGKIECVSAETTEDCIAKIMNGEADAMSLDGGFVY
Rabbit TF       LKCDE-----WSVTSGGLIECESAETPEDCIAKIMNGEADAMSLDGGYVY
Human LF        RKCNQ-----WSGLSEGSVTCSSASTTEDCIALVLKGEADAMSLDGGYVY
Mouse LF        RKCDQ-----WNRDSRGRVTCISFPTTEDCIVAIMKGDADAMSLDGGYIY
Pig LF          RKCRQ-----WSSQSSQNLNCSLASTTEDCIVQVLKGEADAMSLDGGFIY
Bovine LF       KKCQQ-----WSQQSGQNVTCATASTTDDCIVLVLKGEADALNLDGGYIY
Chicken TF      SKCDR-----WSVVSNGDVECTVVDETKDCIIKIMKGEADAVALDGGLVY
Xenopus TF      TKCDD-----WTTISGGAIECTEASTAEECIVQILKGDADAVTLDGGYMY
Human MTF       QKCGDMAVAFRRQRLKPEIQCVSAKSPQHCMERIQAEQVDAVTLSGEDIY
                  C                 C       C             *

Horse TF    419 IAGK-CGLVPVLAENYETRSGSA----CVDTPE--EGYHAVAVVKSSSDP
Pig TF          IAGK-CGLVPVLAENYKT-EGEN----CVNTPE--KGYLAVAVVKKSSGP
Human TF        IAGK-CGLVPVLAENYNK--SDN----CEDTPE--AGYFAVAVVKKSA-S
Rabbit TF       IAGQ-CGLVPVLAENYES--TD-----CKKAPE--EGYLSVAVVKKSN-P
Human LF        TACK-CGLVPVLAENYKSQQSSDPDPNCVDRPV--EGYLAVAVVRR-SDT
Mouse LF        TAGK-CGLVPVLAENQKSSKSNGLD--CVNRPV--EGYLAVAAVRR-EDA
Pig LF          TAGK-CGLVPVLAENQKSRQSSSSD--CVHRPT--QGYFAVAVVRK-ANG
Bovine LF       TAGK-CGLVPVLAENRKSSKHSSLD--CVLRPT--EGYLAVAVVKK-ANE
Chicken TF      TAGV-CGLVPVMAERYDD------ESQCSKTDERPASYFAVAVARK--DS
Xenopus TF      TAGL-CGLVPVMGEYYDQDDLTPCQRSCSQAKG---VYYAVAIVKKGTQ-
Human MTF       TAGKKYGLVPAAGEHYAPEDSSN------------SYYVVAVVRRDSSH
                     C                     C        *

Horse TF    462 DLTWNSLKGKKSCHTGVDRTAGWNIPMGLLYSE----IKHCEF----DKF
Pig TF          DLNWNNLKGKKSCHTAVDRTAGWNIPMGLLYNK----INSCKF----DQF
Human TF        DLTWDNLKGKKSCHTAVGRTAGWNIPMGLLYNK----INHCRF----DEF
Rabbit TF       DINWNNLEGKKSCHTAVDRTAGWNIPMGLLYNR----INHCRF----DEF
Human LF        SLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQ----TGSCKF----DEY
Mouse LF        GFTWSSLRGKKSCHTAVDRTAGWNIPMGLLANQ----TRSCKF----NEF
Pig LF          GITWNSVRGTKSCHTAVDRTAGWNIPMGLLVNQ----TGSCKF----DEF
Bovine LF       GLTWNSLKDKKSCHTAVDRTAGWNIPMGLIVNQ----TGSCAF----DEF
Chicken TF      NVNWNNLKGKKSCHTAVGRTAGWVIPMGLIHNR----TGTCNF----DEY
Xenopus TF      -VSWSNLRGVKTCHTAVGRTAGWNIPVGLITSE----TANCDF----ASY
Human MTF       AFTLDELRGKRSCHAGFGSPAGWDVPVGALIQRGFIRPKDCDVLTAVSEF
                          C *   * **                   C

Horse TF    504 FREGCAPGYRRN---STLCNLCIGSAS---GPGRECEPNNHERYYGYTGA
Pig TF          FGEGCAPGSQRN---SSLCALCIGSER---APGRECLANNHERYYGYTGA
Human TF        FSEGCAPGSKKD---SSLCKLCMGSGL------NLCEPNNKEGYYGYTGA
Rabbit TF       FRQGCAPGSQKN---SSLCELCIGPS--------VCAPNNREGYYGYTGA
Human LF        FSQSCAPGSDPR---SNLCALCIGDEQ---GENK-CVPNSNERYYGYTGA
Mouse LF        FSQSCAPGADPK---SNLCALCIGDEK---GENK-CAPNSKERYQGYTGA
Pig LF          FSQSCAPGSQPG---SNLCALCVGNDQ---GVDK-CVPNSNERYYGYTGA
Bovine LF       FSQSCAPGADPK---SRLCALCAGDDQ---GLDK-CVPNSKEKYYGYTGA
Chicken TF      FSEGCAPGSPPN---SRLCQLCQGSGG---IPPEKCVASSHEKYFGYTGA
Xenopus TF      VGESCAPGSDVK---SNLCALCIGDPEKLSEREKKCSPSASEAYYGYSGA
Human MTF       FNASCVPVNNPKNYPSSLCALCVGDEQ----GRNKCVGNSQERYYGYRGA
                    C             C  C             C          *
```

```
Horse TF     548 FRCLVEK-GDVAFVKHQTVEQNTDGRNPDDWAKDLKSENFKLLCPDGTRK
Pig TF           FRCLVEK-GDVAFVKDQVVQQNTDGKNKDDWAKDLKQMDFELLCQNGARE
Human TF         FRCLVEK-GDVAFVKHQTVPQNTGGKNPDPWAKNLNEKDYELLCLDGTRK
Rabbit TF        FRCLVEK-GDVAFVKSQTVLQNTGGRNSEPWAKDLKEEDFELLCLDGTRK
Human LF         FRCLAENAGDVAFVKDVTVLQNTDGNNNEAWAKDLKLADFALLCLDGKRK
Mouse LF         LRCLAEKAGNVAFLKDSTVLQNTDGKNTEEWARNLKLKDFELLCLDDTRK
Pig LF           FRCLAENAGDVAFVKDVTVLDNTNGQNTEEWARELRSDDFELLCLDGTRK
Bovine LF        FRCLAEDVGDVAFVKNDTVWENTNGESTADWAKNLNREDFRLLCLDGTRK
Chicken TF       LRCLVEK-GDVAFIQHSTVEENTGGKNKADWAKNLQMDDFELLCTDGRRA
Xenopus TF       FRCLVEK-GQVGFAKHTTVFENTDGKNPAGWAKDLKSEDFELLCPDGSRA
Human MTF        FRCLVENAGDVAFVRHTTVFDNTNGHNSEPWAAELRSEDYELLCPNGARA
                   C                                           C

Horse TF     597 SVTEFKSCYLARAPNHAVVSRKEKAA-CVCQELHNQQASYGKNGSHCPDK
Pig TF           PVDNAENCHLARAPNHAVVARDDKVT-CVAEELLKQQAQFGRHVTDCSSS
Human TF         PVEEYANCHLARAPNHAVVTRKDKEA-CVHKILRQQQHLFGSNVTDCSGN
Rabbit TF        PVSEAHNCHLAKAPNHAVVSRKDKAA-CVKQKLLDLQVEYGNTVADCSSK
Human LF         PVTEARSCHLAMAPNHAVVSRMDKVE-RLKQVLLHQQAKFGRNGSDCPDK
Mouse LF         PVTEAKNCHLAIAPNHAVVSRTDKVE-VLQQVVLDQQVQFGRNGQRCPGE
Pig LF           PVTEAQNCHLAVAPSHAVVSRKEKAA-QVEQVLLTEQAQFGRYGKDCPDK
Bovine LF        PVTEAQSCHLAVAPNHAVVSRSDRAA-HVKQVLLHQQALFGKNGKNCPDK
Chicken TF       NVMDYRECNLAEVPTHAVVVRPEKAN-KIRDLLERQEKRFGVNGSE-KSK
Xenopus TF       PVTDYKRCNLAEVPAHAVVTLPDKRE-QVAKIVVNQQSLYGRKGFQ-KDI
Human MTF        EVSQFAACNLAQIPPHAVMVRPDTNIFTVYGLLDKAQDLFGDDHN--KNG
                         C       *           C                   C

Horse TF     646 FCLFQSAT---KDLLFRDDTQCLANLQPTTTYKTYLGEKYLTAVANLR--
Pig TF           FCMFKSNT---KDLLFRDDTQCLARVG-KTTYESYLGADYITAVANLR--
Human TF         FCLFRSET---KDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLR--
Rabbit TF        FCMFHSKT---KDLLFRDDTKCLVDLRGKNTYEKYLGADYIKAVSNLR--
Human LF         FCLFQSET---KNLLFNDNTECLARLHGKTTYEKYLGPQYVAGITNLK--
Mouse LF         FCLFQSKT---KNLLFNDNTECLAKIPGKTTSEKYLGKEYVIATERLK--
Pig LF           FCLFRSET---KNLLFNDNTECLAQLQGKTTYEKYLGSEYVTAIANLK--
Bovine LF        FCLFKSET---KNLLFNDNTECLAKLGGRPTYEEYLGTEYVTAIANLK--
Chicken TF       FMMFESQN---KDLLFKDLTKCLFKVREGTTYKEFLGDKFYTVISNLK--
Xenopus TF       FQMFQSTG--GKDLLFKDSTQCLLEIPSKTTMQEFLGDKYHTAVTSLN--
Human MTF        FKMFDSSNYHGQDLLFKDATVRAVPVGEKTTYRGWLGLDYVAALEGMSSQ
                  C                   C
Horse TF     691 QCSTSRL-------LEAC--TFHRV---------
Pig TF           KCSTSKL-------LEAC--TFHSAKNPRVETTT
Human TF         KCSTSSL-------LEAC--TFRRP---------
Rabbit TF        KCSTSRL-------LEAC--TFHKH---------
Human LF         KCSTSPL-------LEAC--EFLRK---------
Mouse LF         QCSSSPL-------LEAC--AFLTQ---------
Pig LF           QCSVSPL-------LEAC--AFMMR---------
Bovine LF        KCSTSPL-------LEAC--AFLTR---------
Chicken TF       TCNPSDI-------LQMC--SFLEGK--------
Xenopus TF       KCSTSNEASWLPAQFHSCMKIYIMVDCPL--SII
Human MTF        QCSGAAAPAPGAPLLPLLLPALAAR---LLPPAL
                  C               C
```

Figure 34. Multiple alignment of amino acid sequences of members of the transferrin family. Residues which bind iron and (bi)carbonate are indicated by *. Positions of cysteine residues which form disulphide bonds are marked by C. Numbering is for horse transferrin. TF= serum transferrin. LF=lactoferrin. MTF=melanotransferrin. The pig transferrin sequence is incomplete at the N-terminal. Sequences were aligned using Clustal V (Higgens *et al.*, 1992).

References and Genbank/EMBL Databank accession numbers for the sequences are as follows:

Pig TF - Baldwin and Weinstock (1988), X12386

Human TF - Yang *et al.* (1984), M12530

Rabbit TF - Banfield *et al.* (1991), X58533

Human LF - Rey *et al.* (1990), X53961

Mouse LF - Pentecost and Teng (1987), J02737

Pig LF - Alexander *et al.* (1992), M81327

Bovine LF - Goodman and Schanbacher (1991), M63502

Chicken TF - Jeltsch and Chambon (1982), X02009

Xenopus TF - Moskaitis *et al.* (1990), X54530

Human MTF - Rose *et al.* (1986), M12154

Clustal V (Higgens *et al.*, 1992). The positioning of the N-terminal residues of pig transferrin has been altered as the alignment was distorted by the absence of the signal sequence. Cysteine residues which are known to form disulphide bonds in at least one of the proteins are marked 'C'. Many of these cysteines are conserved in the proteins shown. Residues in the iron binding site which are ligands of the ferric ion or (bi)carbonate ion, which are also highly conserved, are marked with an asterisk.

The alignment does not include all the sequences known to be homologous to serum transferrin. The transferrin of the insect *Manduca sexta* has been omitted. Its maximum identity to other transferrin family sequences is 29%, which makes it difficult to align the *Manduca* transferrin correctly with the other sequences. Rat hemiferrin has also been excluded as it only consists of 216 residues, and is therefore homologous to only a part of the other transferrins. The human and chicken B-lym transforming gene proteins have weak homology to a small part of serum transferrin, and were therefore not included. The sequences which were not included in the alignment have also been excluded from the sequence comparisons described below.

Horse transferrin is much like other members of the transferrin family. Its preprotein consists of 706 residues, while other reported vertebrate serum transferrins and lactoferrins have preproteins of 694-717 residues. Horse transferrin has a signal sequence of 19 residues, while those of other transferrins are 18 or 19 residues. In addition, horse transferrin has the internally duplicated structure (Figure 23) shared by other serum transferrins, lactoferrins and human melanotransferrin.

The matrix in Table 3 shows sequence identity in the transferrin family by pairwise comparisons. The figures represent the percentage of identical amino acids, calculated using the Gap program of the GCG package (Devereux *et al.*, 1984). Horse transferrin is most similar to other mammalian serum transferrins (human, pig, rabbit), less similar to mammalian lactoferrins (human, mouse, bovine, pig), and still less similar to chicken and *Xenopus* transferrins and human melanotransferrin.

The mammalian serum transferrins share a large proportion of identical amino acids (72-79%), and similarly the lactoferrins have 64-73% in common. Identities between mammalian lactoferrins and serum transferrins are lower; they have 57-63% identical amino acids. Chicken transferrin is 49-54% identical to the mammalian serum transferrins and lactoferrins, and *Xenopus* shares 45-49% of amino acids with chicken transferrin and mammalian

Table 3

Amino Acid Sequence Identity among Members of the Transferrin Family

	Horse TF	Pig TF	Human TF	Rabbit TF	Human LF	Mouse LF	Bovine LF	Pig LF	Chicken TF	*Xenopus* TF
Pig TF	73.7									
Human TF	73.4	71.5								
Rabbit TF	72.8	73.1	79.0							
Human LF	61.9	62.0	61.8	62.3						
Mouse LF	56.7	59.0	57.5	57.9	70.2					
Bovine LF	62.9	61.5	61.6	61.2	69.4	63.6				
Pig LF	62.2	61.8	61.5	61.4	71.0	64.7	73.3			
Chicken TF	52.8	53.7	52.4	52.6	52.7	49.3	53.6	51.4		
Xenopus TF	48.6	48.3	46.1	46.8	46.7	47.4	45.4	45.2	47.1	
Human MTF	44.6	42.3	43.9	44.3	42.7	41.0	41.9	42.2	41.4	41.2

Identity among members of the transferrin family is expressed as percentage of identical amino acids, calculated using the GAP program of the GCG package (Devereux *et al.*, 1984). TF=transferrin. LF=lactoferrin. MTF=melanotransferrin.

serum transferrins and lactoferrins. Human melanotransferrin is only 41-45% identical to the other sequences.

The relationships between members of the transferrin family are illustrated by the phylogenetic tree in Figure 35. This was constructed using the Clustal V package (Higgens *et al.*, 1991). This tree shows horse transferrin to be most closely related to the other mammalian serum transferrins. The mammalian serum transferrins form a monophyletic group, as do the lactoferrins. The tree shown is unrooted as sequence data doesn't give the position of the root.

Alignment of the horse transferrin amino acid sequence with related sequences (Figure 34) revealed several insertions/deletions which were specific to particular groups of sequences: mammalian serum transferrins; lactoferrins; mammalian serum transferrins plus lactoferrins; serum transferrins plus lactoferrins. These insertions/deletions are consistent with the relationships illustrated in Figure 35. There are a few cases of insertions or deletions which must have occurred more than once, independently, if the relationships shown in Figure 35 are correct. An example is the deletion between positions 554 and 555 in horse transferrin. In the lactoferrin sequences this residue is either an alanine or a valine and in melanotransferrin it is an alanine. In the serum transferrins of *Xenopus*, chicken and mammals it is deleted. The equivalent residue in the N-lobe is alanine in seven of the eleven sequences. It appears that an alanine at this position is likely to be the ancestral form, whereas the deletion is a derived characteristic. According to the relationships shown in Figure 35 and Table 3, the occurrence of a deletion in the mammalian, chicken and *Xenopus* serum transferrins indicates that this deletion must have occurred independently three times. As it is present specifically in serum transferrins, it may have some functional significance, the nature of which is not clear.

The positions of the insertions/deletions in the transferrin family sequences were located on the 3-dimensional structure of human lactoferrin (Anderson *et al.*, 1989) in order to predict whether they might have some functional significance. Most were located on external loops of the molecule, where they would probably be unlikely to influence iron binding. However, such external features could affect interaction of the transferrins with other molecules, such as their respective receptors. One notable example is a 2 residue deletion present in all four lactoferrins, in the region joining the two lobes. This region is discussed further in Section 4.2.6.

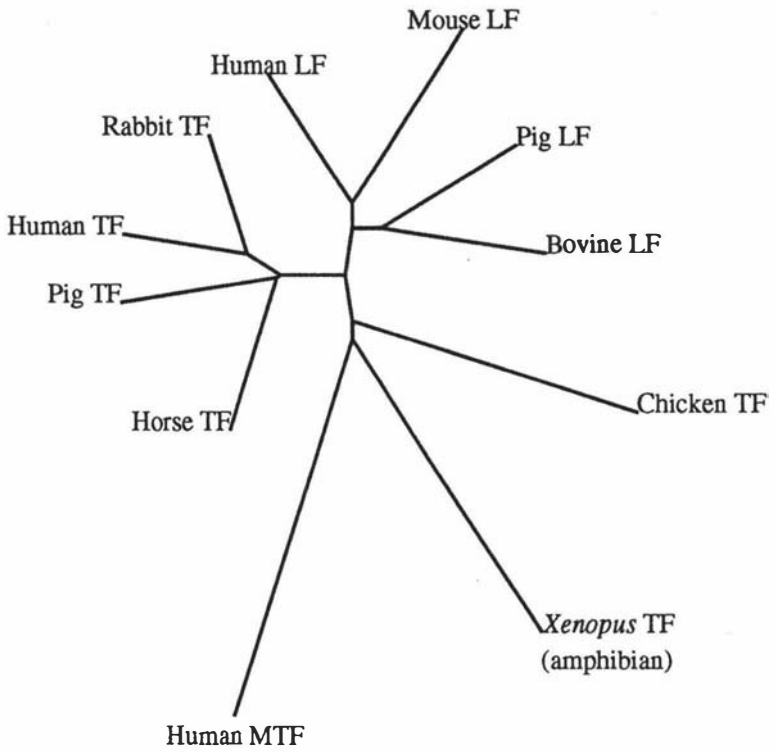


Figure 35. Phylogenetic tree of members of the transferrin family. The tree was constructed by analysing the amino acid sequences shown in Figure 34, using Clustal V (Higgens *et al.*, 1991). TF=transferrin. LF=lactoferrin. MTF=melanotransferrin.

The transferrin family relationships revealed by comparison of the derived horse transferrin protein sequence to other members of the transferrin family (Figure 35, Table 3) are consistent with the model proposed for transferrin family evolution (Yang *et al.*, 1985; Bowman *et al.*, 1987), illustrated in Figure 4, and described in Section 1.9. The grouping of lactoferrins and mammalian serum transferrins is consistent with the idea that the gene duplication which produced lactoferrin occurred after the divergence of birds and mammals. The gene duplication producing human melanotransferrin appears to have occurred before the bird-mammal divergence, suggesting that melanotransferrin is likely to be present in birds and mammals, and perhaps all vertebrates, although it has so far only been detected in humans.

The alignment of transferrin family sequences reveals that residues important for transferrin function are generally conserved between the different proteins, notably those residues involved in iron binding, disulphide bonds, and to a lesser degree glycosylation.

4.2.3 Glycosylation

The sixteen transferrin family proteins studied by Spik *et al.* (1988) each had one to four N-linked glycan chains. The signal sequences for N-linked glycosylation are NXS and NXT, where X can be any amino acid. These sites are not always glycosylated. Two N-linked glycosylation signals are present in the horse transferrin sequence (Figure 22, Section 3.4.2). This is consistent with the results of Spik *et al.* (1988) and Stratil *et al.* (1984) who determined independently that horse transferrin has 1-2 glycans.

Glycosylation sites occur at equivalent positions in other transferrins. Pig and rabbit transferrins (Baldwin and Weinstock, 1988; Bailey *et al.*, 1988) have an NSS glycosylation signal in an equivalent position to the NST at position 515 in horse. The asparagine residue is glycosylated in rabbit transferrin, but it is not known whether it is in pig transferrin. The horse transferrin NGS glycosylation site at position 638 is replaced by NVT in human transferrin which is glycosylated (MacGillivray, 1982), and NGS in human lactoferrin and chicken ovotransferrin, which are not glycosylated (Metz-Boutigue *et al.*, 1984).

4.2.4 Iron and anion binding residues

The 8 residues known to be iron ligands in rabbit transferrin and human lactoferrin (Bailey *et al.*, 1988; Anderson *et al.*, 1989) occur in equivalent positions in horse transferrin (Figure 34). Likewise, the eight residues which bind the (bi)carbonate anion (Sarra *et al.*,1990; Anderson *et al.*, 1989) are conserved in horse transferrin.

The importance of these residues to the function of the transferrins is illustrated by the way they are highly conserved among most of the proteins. The sixteen iron and anion binding residues are completely invariant in the mammalian serum transferrins and lactoferrins, and in chicken transferrin. In *Xenopus* transferrin two of the sixteen residues are different. A threonine replaces the iron-binding aspartate in the N-lobe. This change has been introduced into human lactoferrin, where it does not inhibit iron binding, but does result in iron being more readily released (Catherine Day, personal communication). Therefore, this substitution probably does not prevent iron binding in *Xenopus*. The other change in *Xenopus* transferrin is the substitution of a lysine for the anion-binding arginine in the N-lobe. As these are both basic amino acids, this conservative change is unlikely to interfere with anion binding. It is likely to be the positive charge on the residue in this position which is essential for binding the anion. The iron and anion-binding residues are less well conserved in human melanotransferrin, particularly in the C-lobe, which may be indicative of functional differences. It has been suggested that melanotransferrin does not bind iron in the C-lobe (Baker *et al.*, 1987).

4.2.5 Disulphide bonds

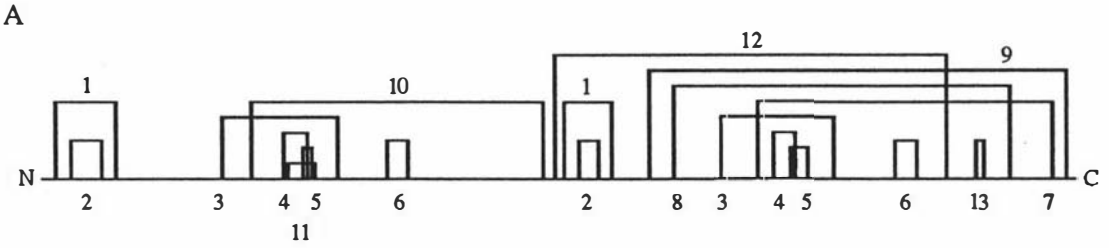
Horse transferrin has 37 cysteine residues, 36 of which are in positions where there are cysteine residues in pig, human and rabbit transferrins (Figure 34). Many of these 36 cysteines are also present in other transferrin family proteins. Previous studies have shown that many disulphide bonds are conserved among human transferrin, human lactoferrin and chicken ovotransferrin (Metz-Boutigue *et al.*, 1984), and that rabbit transferrin and human lactoferrin have very similar 3-dimensional structures (Bailey *et al.*, 1988; Anderson *et al.*, 1989). This information, combined with sequence similarities, suggests that horse transferrin is likely to have the same general structure as other transferrins, and to share many of the same disulphide bonds. Therefore, it should be acceptable to predict the disulphide bonds present in horse and other transferrins using sequence data and known bonding patterns.

Horse transferrin and pig transferrin are likely to have 18 disulphide bonds (Figure 36), which are equivalent to 18 of the 19 present in human transferrin (Metz-Boutigue *et al.*, 1984; MacGillivray *et al.*, 1982, 1983) and rabbit transferrin (Bailey *et al.*, 1988). These include the 6 N-lobe and 9 C-lobe bonds which are present in chicken transferrin (Williams *et al.*, 1982a). The lactoferrins probably also have the set of 15 'fundamental' bonds found in chicken transferrin, with the addition of some of the 'extra' bonds which occur in the mammalian transferrins (Figure 36). Human melanotransferrin and *Xenopus* transferrin probably have some of the same disulphide bonds as chicken transferrin, but others must be missing due to the absence of some cysteine residues. There are also some extra cysteines present in these proteins, which have the potential to form alternative bonds, or modified versions of the conserved ones.

4.2.6 Structure of the region joining the two lobes

In rabbit transferrin the region bridging the two lobes is linear, and each end of it is anchored to its respective lobe by a disulphide bond, specifically bond 10 in the N-lobe and bond 12 in the C-lobe (Figure 36 & Bailey *et al.*, 1988). These two bonds are present in human transferrin, which is therefore likely to have a similar joining peptide. In both rabbit and human transferrins, there are seven residues between the cysteines forming bonds 10 and 12, which make up the joining peptide. These seven residues include two prolines, which would ensure that the joining region was linear as proline residues disrupt α-helix. The joining peptide of human lactoferrin is a three turn α-helix, which lacks bonds 10 and 12 (Anderson *et al.*, 1989). Bovine, murine and porcine lactoferrins are likely to have the same joining peptide as human lactoferrin, as they share the absence of bonds 10 and 12, the absence of proline residues and also share a two residue deletion which is not present in the serum transferrins. Horse and pig transferrins have the cysteine residues required for bond 12, but lack bond 10, and have proline residues in the joining region. Therefore, they probably have a linear joining peptide of the same length as rabbit transferrin, but differing in that it is not disulphide bonded to the N-lobe. It is possible that the pig transferrin could have some α-helix in this region as only one proline is present. In chicken transferrin bonds 10 and 12 are absent, and the presence of prolines suggests a linear joining peptide. The transferrin family proteins apparently have joining peptides which take on a variety of structures as shown diagrammatically in Figure 37. The structure of the joining peptides of human melanotransferrin and *Xenopus* transferrin are not as

A



B

	N lobe	C lobe
Horse TF	1-6, 11	1-9, 12, 13
Pig TF	1-6, 11	1-9, 12, 13
Human TF	1-6, 10, 11	1-9, 12, 13
Rabbit TF	1-6, 10, 11	1-9, 12, 13
Human LF	1-6	1-9, 13
Mouse LF	1-6	1-9, 13
Pig LF	1-6, 11	1-9, 13
Bovine LF	1-6, 11	1-9, 13
Chicken TF	1-6	1-9

Figure 36. Disulphide bonding in the transferrins. A. Disulphide bonds present in human and rabbit transferrins, redrawn from Metz-Boutigue *et al.* (1984), and numbered according to Bailey *et al.* (1988). The transferrin sequence is represented as a line with N- and C-terminals labelled. The disulphide bonds are depicted joining positions within the sequence. Bonds numbered 1-6 are represented in both the N- and C- terminal halves, whereas bonds 7-13 are only present in one of the halves. B. Table showing which of the disulphide bonds illustrated in the diagram above are present in other members of the transferrin family. Some or all of the bonds in human lactoferrin, chicken ovotransferrin and rabbit and human serum transferrins are known (Anderson *et al.*, 1989, Metz-Boutigue *et al,* 1984, Williams, 1982a, Bailey *et al.*, 1988, MacGillivray *et al.*, 1982, 1983); the remainder are predicted from the amino acid sequences.

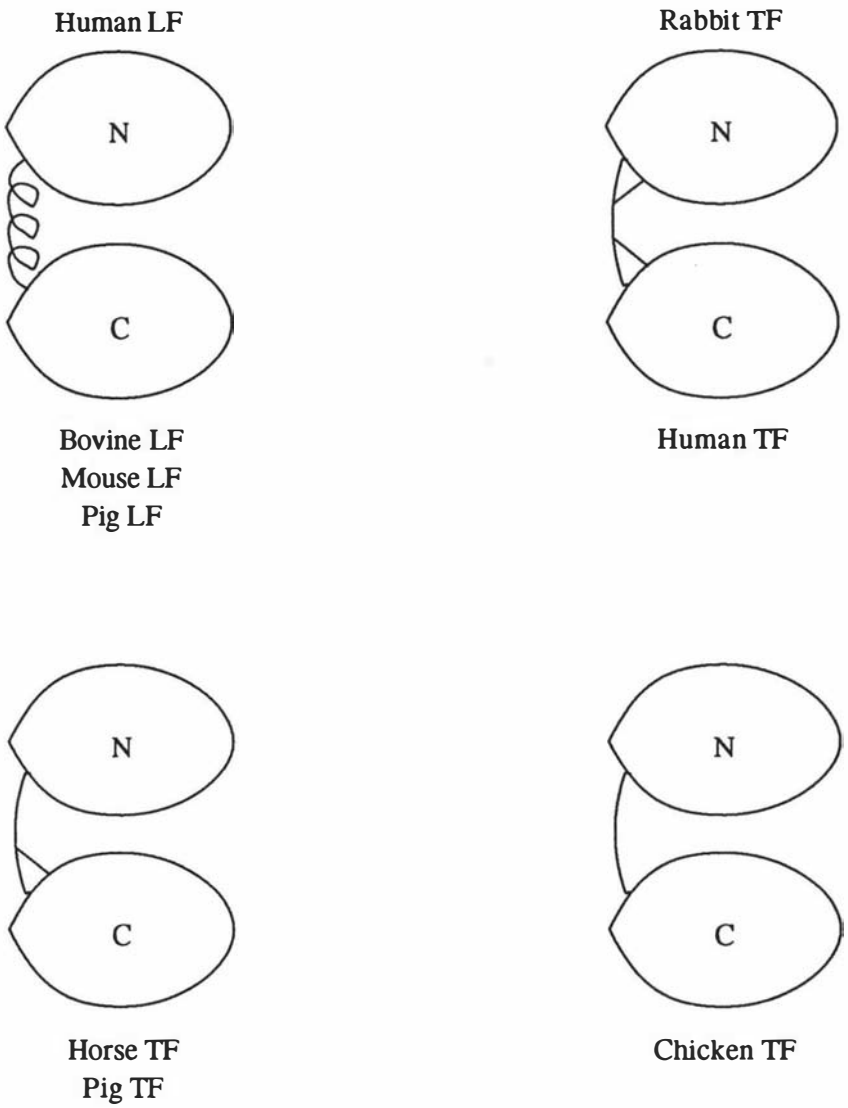


Figure 37. Structure of the region joining the two lobes in serum transferrins and lactoferrins. In human lactoferrin this region is a 3-turn α-helix (Anderson *et al.*, 1989). A similar structure is likely in the other lactoferrins indicated, according to similarities in primary structure. The two lobes of rabbit transferrin are joined by a linear region of 7 amino acids, each end of which is anchored to its respective lobe by a disulphide bond (Bailey *et al.*, 1988). Human transferrin is likely to have a similar structure. In horse and pig transferrins, the joining peptide is likely to be similar to that of rabbit transferrin, except that the disulphide bond to the N-lobe is absent. Chicken transferrin is also likely to have a linear joining peptide, but has neither of the disulphide bonds present in rabbit transferrin. The proteins named above the diagrams are known to have the structure shown. Those named below are predictions based on the amino acid sequences, particularly the presence of cysteine and proline residues. TF=serum transferrin, LF=lactoferrin.

easy to predict due to deletions near this region and low sequence homology with the proteins whose structures are known.

The structure of the joining peptide may affect protein function by influencing interactions between the lobes, and the overall shape and flexibility of the protein. Such differences could affect the specificity of the binding of transferrins and lactoferrins to their respective receptors. Anderson *et al.* (1989) suggested that differences in the connecting peptide could affect the relative orientation of the two lobes. This was based on the observation that the rotation required to superimpose the human lactoferrin C-lobe on the N-lobe was 180°, whereas in rabbit transferrin it was 167°, and the joining peptides of these two proteins differ substantially as described above. This now seems unlikely to be true because the bovine and human lactoferrins, which probably have very similar joining peptides, have a 10° difference in the rotation required to superimpose the two lobes (Baker *et al.*, 1991). The pronounced difference between the joining peptides of the mammalian serum transferrins and lactoferrins, and the conservation within each group, does suggest some functional importance, but the significance remains unknown.

4.3 POLYMORPHISMS

4.3.1 Positions and types of amino acid substitutions

The amino acid substitutions in horse transferrin all occur at sites which are variable between species or between proteins in the transferrin family (Figure 38). At these positions, several different amino acids occur among the different proteins, generally involving moderately conservative changes. Therefore their complete conservation does not appear to be crucial to protein function. Two of the polymorphisms in horse transferrin (S/P at position 437 and D/G at 460) occur at positions which are deleted in human and rabbit transferrins. In addition, deletions occur in the N-lobe of horse transferrin at positions equivalent to the D/G substitution at position 460 and the C/R substitution at position 572. The 5 exon 15 substitutions, however, are in a region that has no deletions among any of the proteins in Figure 38, nor in the N-lobe of any of these proteins.

The eight amino acid substitutions identified were a mixture of different types (Table 2), most being moderately conservative changes. For example, the P/S, D/G, E/G, K/Q, S/A and K/E

Horse TF D/F_1	427	PVLAENYETR**P**GSA----CVD	443
Horse TF D/F_1		PVLAENYETR**S**GSA----CVD	
Pig TF		PVLAENYKT-EGEN----CVN	
Human TF		PVLAENYNK--SDN----CED	
Rabbit TF		PVLAENYES--TD-----CKK	
Human LF		PVLAENYKSQQSSDPDPNCVD	
Mouse LF		PVLAENQKSSKSNGLD--CVN	
Pig LF		PVLAENQKSRQSSSSD--CVH	
Bovine LF		PVLAENRKSSKHSSLD--CVL	
Chicken TF		PVMAERYDD------ESQCSK	
Xenopus TF		PVMGEYYDQDDLTPCQRSCSQ	
Human MTF		PAAGEHYAPEDSSN-------	
Horse N-lobe	95	PVVAEFYGSKTE---------	106

Horse TF D	457	SSS**D**PDLTWNSLKGKKSCHTGVDRTAGWN	485
Horse TF R		SSS**G**PDLTWNSLKGKKSCHTGVDRTAGWN	
Pig TF		KSSGPDLNWNNLKGKKSCHTAVDRTAGWN	
Human TF		KSA-SDLTWDNLKGKKSCHTAVGRTAGWN	
Rabbit TF		KSN-PDINWNNLEGKKSCHTAVDRTAGWN	
Human LF		R-SDTSLTWNSVKGKKSCHTAVDRTAGWN	
Mouse LF		R-EDAGFTWSSLRGKKSCHTAVDRTAGWN	
Pig LF		K-ANGGITWNSVRGTKSCHTAVDRTAGWN	
Bovine LF		K-ANEGLTWNSLKDKKSCHTAVDRTAGWN	
Chicken TF		K--DS**N**VNWNNLKGKKSCHTAVGRTAGWV	
Xenopus TF		KGTQ--VSWSNLRGVKTCHTAVGRTAGWN	
Human MTF		RDSSHAFTLDELRGKRSCHAGFGSPAGWD	
Horse N-lobe	119	KNS--NFQLNQLQGKKSCHTGLGRSAGWN	145

Horse TF D	562	HQTVEQNTDG**C**NPDDWAKDLKS**E**NF**K**LLCPDGTRK**S**VTEF**K**SC**Y**LA	607
Horse TF F_1		HQTVEQNTDG**R**NPDDWAKDLKS**E**NF**K**LLCPDGTRK**S**VTEF**K**SC**Y**LA	
Horse TF R		DLKS**G**NF**Q**LLCPDGTRK**A**VTEF**E**SC**N**LA	
Pig TF		DQVVQQNTDGKNKDDWAKDLKQMDFELLCQNGAREPVDNAENCHLA	
Human TF		HQTVPQNTGGKNPDPWAKNLNEKDYELLCLDGTRKPVEEYANCHLA	
Rabbit TF		SQTVLQNTGGRNSEPWAKDLKEEDFELLCLDGTRKPVSEAHNCHLA	
Human LF		DVTVLQNTDGNNNEAWAKDLKLADFALLCLDGKRKPVTEARSCHLA	
Mouse LF		DSTVLQNTDGKNTEEWARNLKLKDFELLCLDDTRKPVTEAKNCHLA	
Pig LF		DVTVLDNTNGQNTEEWARELRSDDFELLCLDGTRKPVTEAQNCHLA	
Bovine LF		NDTVWENTNGESTADWAKNLNREDFRLLCLDGTRKPVTEAQSCHLA	
Chicken TF		HSTVEENTGGKNKADWAKNLQMDDFELLCTDGRRANVMDYRECNLA	
Xenopus TF		HTTVFENTDGKNPAGWAKDLKSEDFELLCPDGSRAPVTDYKRCNLA	
Human MTF		HTTVFDNTNGHNSEPWAAELRSEDYELLCPNGARAEVSQFAACNLA	
Horse N-lobe	228	HSTVLEN--------LPQEADRDEYQLLCRDNTRKSVDEYKDCYLA	265

Figure 38. Positions of the horse transferrin sequence polymorphisms with respect to interspecies, interprotein and interlobe variation of transferrins. Polymorphic residues are shown in bold type. Numbering is according to horse transferrin sequence. Residues 562-579 of horse transferrin R were not sequenced. The first set of sequences correspond to part of the 528 bp SSCP fragment; the second set to exon 12; and the third set is part of the 303 bp SSCP fragment and exon 15 combined. TF=transferrin. LF=lactoferrin. MTF=melanotransferrin.

substitutions have scores of 0 or 1 in Dayhoff's log odds matrix for 250 PAMs (accepted point mutations per 100 residues) (Dayhoff *et al.*, 1978), indicating substitutions which occur with moderate frequency. The two most potentially disruptive changes are the cysteine/arginine substitution between D and F_1 (position 572), and the asparagine/tyrosine between the D- and R-groups (position 605), which have Dayhoff's scores of -4 and -2 respectively.

The horse transferrin sequence shown in Figure 22 has an odd number of cysteine residues. Comparison with other transferrins indicates that the cysteine at position 625 is likely to be unpaired. Therefore the substitution of a cysteine in the D variant for an arginine in F_1 and F_2 produces the potential for an extra disulphide bond. Consideration of the positions of the unpaired cysteines on the 3-dimensional structure of human lactoferrin (Anderson *et al.*, 1989) indicates that a bond between the two is unlikely.

4.3.2 Polymorphism and electrophoretic mobility

The SSCP analysis of DF_1 cDNA revealed nucleotide differences encoding only two amino acid substitutions. The sequences of these two variants were expected to be fairly similar due to the similarity of their electrophoretic mobilities. The substitutions observed were serine for proline, both of which carry a neutral charge, and cysteine for arginine, which replaces a neutral residue in D for a positively charged one in F_1. This observation accounts for the difference in electrophoretic mobility between the two variants. The greater negative charge on the D variant would cause its increased mobility toward the cathode (Figure 5). No differences were found between F_1 and F_2, but only part of the F_2 allele was studied. As F_1 and F_2 variants are difficult to distinguish by electrophoresis, there may be only a single amino acid difference between them.

In comparison, the six amino acid differences which distinguish the D, F_1, F_2, H_2, * and O, R groups of variants correspond well with the large differences in electrophoretic mobility between these groups (Figure 5). However, the observed substitutions do not involve charge changes which would produce the mobility differences. The variable residues in the D-type variants (D, F_1, F_2, H_2 and *) include two positively and two negatively charged residues, while the R-type variants (O and R) have a single negative residue. The net charge difference is 1, with the R-type variants being more negative. Therefore, neither the size nor polarity of charge difference is in agreement with differences in electrophoretic mobility. This can be explained

by the proposal that other substitutions distinguish the two groups (Section 4.3.4), and that these undefined substitutions may account for the differences in electrophoretic mobility.

4.3.3 Number of substitutions

Between the D and R variants, there are at least 6 amino acid substitutions in a sequence of 706 amino acids (Table 2). Is 6 amino acid substitutions in 706 (0.8%) residues a high substitution rate for variants within a species? A useful comparison could be the globins as these are the best known mammalian proteins. Two hundred variants of human haemoglobin have been analysed, most of which are single amino acid substitutions in the α or β chains. A few β-globin variants occur which differ from the common variants by 2 amino acids (Weatherall and Clegg, 1976). Therefore, the difference between any two human globin variants must be a maximum of 4 amino acid substitutions. As the human β-globin polypeptides are 146 residues in length, the variants differ by 0.7-2.7%. If the six amino acid substitutions distinguishing the D- and R-groups were the only differences between the two groups, this would not be a high substitution frequency. However, as only two small regions of the amino acid sequence were studied, it is possible that there are many other differences between the two groups. There is evidence for further differences near the part of exon 15 which was sequenced, as described in Section 4.3.4.

The six substitutions distinguishing the D and R variants were in the 29 amino acids sequenced from exon 12, and the 28 from exon 15, a total of 57 residues. If this level of polymorphism (10.5%) were present throughout the whole molecule, the D and R variants of horse transferrin would be very different for variants within a species, in comparison to the 0.7-2.7% sequence variation in the human globins.

The exon 12 and 15 regions were analysed in other mammalian transferrin sequences available, to establish an inter-species level of homology. The exon 15 region of 28 amino acids differed at 7-15 positions (25-54%) in pairwise comparisons of human, porcine, rabbit and horse transferrins (Table 4), while the 29 residues of exon 12 differed at 5-8 positions (14-28%), giving a combined range for the two regions of 13-22 positions in 57 (23-39%). Although the level of polymorphism between the two groups of horse variants (10.5%) is substantially less than between the species in Table 4 (23-29%), it might be close to the level which would occur between closely related species.

Table 4

Intra- and Interspecies Variation in Exons 12 and 15 of Mammalian Serum Transferrins

A.

Exon 12

Horse D	SSSDPDLTWNSLKGKKSCHTGVDRTAGWN
Horse R	SSSGPDLTWNSLKGKKSCHTGVDRTAGWN
Pig	KSSGPDLNWNNLKGKKSCHTAVDRTAGWN
Human	KSA-SDLTWDNLKGKKSCHTAVGRTAGWN
Rabbit	KSN-PDINWNNLEGKKSCHTAVDRTAGWN

Exon 15

Horse D	DLKSENFKLLCPDGTRKSVTEFKSCYLA
Horse R	DLKSGNFQLLCPDGTRKAVTEFESCNLA
Pig	DLKQMDFELLCQNGAREPVDNAENCHLA
Human	NLNEKDYELLCLDGTRKPVEEYANCHLA
Rabbit	DLKEEDFELLCLDGTRKPVSEAHNCHLA

B.

		Exon 12	Exon 15	Exons 12 & 15
Horse D	Horse R	3.4	17.9	10.5
	Pig	17.2	53.6	35.1
	Human	27.6	50.0	38.6
	Rabbit	24.1	35.7	29.8
Horse R	Pig	13.8	46.4	29.8
	Human	27.6	50.0	38.6
	Rabbit	24.1	39.3	31.6
Pig	Human	20.7	46.4	33.3
	Rabbit	13.8	32.1	22.8
Human	Rabbit	25.0	25.0	25.0

A. Amino acid sequences of the D and R variants of horse transferrin in the regions of exons 12 and 15 which were sequenced, aligned with the equivalent regions of serum transferrins from other species.

B. Percentage of residues which differ between pairs of the sequences in A.

The six substitutions in exons 12 and 15 which distinguish the D- and R-groups of horse transferrins seem to indicate a substantial difference between the two groups. However with the limited data available, it is difficult to predict the extent of this difference, or to compare it to levels of polymorphism between variants of other genes. The aspect of this data which really is striking is that members of the two groups differed in all six positions, and no 'hybrid' types were observed.

4.3.4 Further polymorphism between D- and R-type alleles

Given the level of polymorphism between the D- and R-types in the exon 15 region, it is possible that the two groups may differ extensively in other parts of the molecule, eg. in the region which is the N-lobe equivalent to exon 15 (exon 7 in human transferrin). However, not all parts of the molecule are as variable as exon 15, as shown by the single amino acid substitution in exon 12.

There is evidence for further polymorphism around the region of exon 15 which was studied, in the behaviour of the PCR. One primer pair (15/1 and 15/3) amplified only the D-type alleles, and gave no products with O or R. The other pair (15/1 and 15/2) could amplify O and R, but in the presence of D-type alleles, the D-type alleles were amplified preferentially to O and R alleles. As the primers were synthesized according to a D or F_1 sequence, this shows that O and R must differ from D and/or F_1 in the region of at least two of the three primers.

The peptide maps of D and R transferrins produced by Chung and McKenzie (1985) revealed differences in several peptides but only two of these were analysed by sequencing. Some of the other differences observed may have been due to the 4 amino acid substitutions in exon 15 which lay outside of the peptide they sequenced, while others could have been due to polymorphisms in other parts of the molecule.

As selection theoretically has no effect on synonymous mutations in the coding region, we would expect them to occur equally frequently throughout the molecule. The 85 bp region of exon 15 which was sequenced had three synonymous mutations between the D- and R-groups. This level of synonymous mutations between D and R may occur throughout the molecule, indicating extensive differences between the D- and R-groups. In contrast, sequencing of

fragments of D and F_1 alleles (the 303 and 528 bp SSCP fragments) revealed only one synonymous mutation in a total of 831 bp, a much lower level of variability.

4.3.5 Two groups of horse transferrins

The sequencing of exons 12 and 15 identified 6 amino acid substitutions which distinguish the D, F_1, F_2, H_2, * group from the O, R group. This extends the results of other workers who divided the horse transferrin variants into two groups according to electrophoretic mobility, immunological reactivity and serum levels (Kaminski *et al.*, 1981; Didkowski *et al.*, 1981; summarized in Table 1). The 6 amino acid substitutions distinguishing the D, F_1, F_2, H_2, * and O, R variants, being clustered and on an external part of the molecule (see Section 4.3.7) account for the difference in immunoreactivity between the two groups, which was observed by Kaminski *et al.* (1981).

In addition, Bailey *et al.* (1991) detected an RFLP in horse genomic DNA digested with *Msp*I, which gave a 2.4 kb fragment in the D, F_1, F_2, G and H_2 variants, and a 0.5 kb fragment for O and R. However Bailey *et al.* (1991) did not detect RFLPs with *Eco*RI and *Pvu*II. My sequence data for exon 15 reveal polymorphic *Eco*RI and *Pvu*II sites distinguishing the D- and R-groups. The probe used by Bailey *et al.* (1991) did include exon 15, but they had some "difficulties" with their human transferrin cDNA probe, and recommended "using an equine transferrin probe for future studies".

Six amino acid substitutions distinguish the two groups of horse transferrin, with no 'hybrid' types occurring (i.e. those which are D-type at some positions and R-type at others) at least among the variants studied here. The presence of two such distinct groups of variants within a species seems odd. Hybrid variants must once have existed, either as common ancestors to the modern variants, or during the evolution of one group from the other. The current absence of these hybrids is readily explained by their loss through selection or random events. However, recombination between a D-type allele and an R-type allele of horse transferrin would produce novel hybrid alleles. A range of different alleles can be predicted depending on the point of cross-over. Therefore, the existence of two distinct groups of transferrin variants in horses, with an absence of hybrid variants, requires explanation.

The two groups of transferrins appear to extend through most breeds of horse, according to the many reports of frequencies of electrophoretic variants of horse transferrin (Braend, 1964; Gahne, 1966; Kaminski and Urbanska-Nicholas, 1979; Bowling and Clark, 1985). The two groups may also occur in other equid species. Analysis of the transferrins of the donkey and zebra (*Equus burchelli*) revealed a range of variants similar to those of the horse. Singhvi and Khanna (1988) reported the presence of transferrin variants D, F, H, M, O and R in the donkey. Osterhoff (1966) observed in the donkey variants apparently the same as the horse variants H, M, O and R, as well as new variants denoted D_d and I. The transferrins of *E. burchelli* included the horse variants H, J, M, O and R, as well as new variants D_z, G, I and J (Osterhoff, 1966). Interestingly, Przewalski's horse (*E. przewalski*), the species most closely related to the domestic horse (*E. caballus*), has 2-4 variants which appear to be members of the D-group of transferrins, as well as 2-3 novel variants, but had no O or R variants in 226 samples (Putt and Whitehouse, 1983; Bowling and Ryder, 1987). Przewalski's horse and the domestic horse are so closely related that they can produce fertile hybrids (Short *et al.*, 1974), yet their transferrin variants appear to share no more similarity than the domestic horse shares with the donkey and zebra. This unlikely situation may be partially due to poor blood-typing data, and confusion with the nomenclature of horse transferrin types.

The variants which these authors describe as being the same as horse variants have a similar electrophoretic mobility to the equivalent horse variant, but it is not yet known if they are actually identical in amino acid sequence. However it is likely that some of the variants are at least similar across the species. As both donkeys and zebras have variants which appear to be similar to members of the D- and R-groups of horse transferrin, it is likely that these two groups occur in the donkey and zebra. Therefore, the points discussed in the remainder of this section will be based on the assumption that the two groups of horse transferrins, as defined by the 6 amino acid substitutions, occur in the donkey and zebra. The two groups of alleles are likely to be ancient, predating the divergence of the contemporary equid species, an event which is estimated to have occurred 3.9 million years (myr) ago (George and Ryder, 1986). This is consistent with the substantial differences between the two groups.

Assuming the D- and R-groups of equid transferrins have existed for approximately 3.9 myr, there should have been sufficient time for recombination between transferrin alleles to have produced the hybrid types of transferrin. As five of the nucleotide substitutions which produce amino acid substitutions occur within a very small region (85 bp) the frequency of cross-over within this region would be very low. However, the distance between the substitution in exon

12 and those in exon 15 is much greater (approximately 9 kb in human transferrin, Schaeffer *et al.*, 1987) so the cross-over frequency would be expected to be higher. The average map unit size in the human genome is approximately 10^6 bp (Lewin, 1990). Assuming recombination occurs at a comparable frequency in the horse genome, there would be an average of 1% recombination per 10^6 bp per generation in horses. Therefore positions in the nucleotide sequence separated by 9 kb would have a recombination frequency of $9x10^{-5}$ per generation. The average generation time in thoroughbred horses is 11 years (Cunningham, 1991), so the 3.9 myr during which the two groups of transferrins appear to have existed represent approximately 350,000 generations. The linkage disequilibrium which would be expected to exist at present if recombination was occurring at a rate equal to the average for the genome can be estimated using the relationship $d_t=(1-c)^t d_0$ where d_t is the linkage disequilibrium after t generations; c is the recombination frequency; and d_0 is the initial linkage disequilibrium, which has a maximum value of 0.25 (Ayala and Kiger, 1984). Using $c=9x10^{-5}$, t=350,000 and $d_0=0.25$ gives a value for d_t which approaches zero ($5x10^{-15}$) indicating that after 3.9 myr, markers 9 kb apart and initially in total linkage disequilibrium would have virtually reached equilibrium. Therefore, assuming recombination at a frequency close to the average for the genome, the 3.9 myr since the equid divergence would be plenty of time for hybrid D/R-type alleles produced by cross-over in the region between exons 12 and 15 to have reached a significant frequency in equid species.

Recombination occurs at a greater frequency in some regions of the DNA than in others. In the human β-globin gene cluster there are two regions in which recombination is rare, separated by a region described as a recombination hot spot (Kazazian *et al.*, 1983). The 32 kb 5' region, including 3 functional globin genes and a pseudogene, has 5 restriction enzyme sites which are polymorphic in Mediterranean and Asian populations. Of the 32 possible combinations of these polymorphic sites, only 3 are common. Similarly, the 18 kb 3' region, containing the β-globin gene and flanking DNA, has 5 polymorphic restriction sites, of which only 3 combinations are common. Only one of the ten polymorphic restriction sites is in the coding sequence of a functional gene; the others lie in flanking DNA, introns or in the pseudogene. Consequently, the lack of alternative types is unlikely to be due to selection. Therefore, the fact that only three common sequence types occur in each region indicates that recombination is infrequent. In contrast, the 11 kb region between the two segments described above appears to be the site of frequent recombination, as indicated by the occurrence of all 9 combinations of the 3 5' types with the 3 3' types (Kazazian *et al.*, 1983).

The horse transferrin gene, or a part of it, could be a region where recombination occurs very rarely, if at all. The 6 amino acid substitutions distinguishing the D- and R-groups of horse transferrin occur in a region of the gene whose human equivalent spans 9 kb (Schaeffer *et al.*, 1987). The data on human haemoglobin indicate that recombination can occur so infrequently in a region even larger than this size, that recombinant types are not observed. Thus, the two groups of horse transferrins could have evolved independently, then maintained their differences, due to the absence of recombination.

An interesting situation, which in some ways parallels the horse transferrin variants, occurs in the horse α-globins. The horse α-globins are encoded by two genes, α1 and α2, which are approximately 5 kb apart. The two genes differ at one position: residue 60 is lysine in α1 and glutamine in α2. Two haplotypes occur: BI has a tyrosine at position 24 in both α1 and α2, while BII has phenylalanine in both these positions. Other haplotypes occur at lower frequencies, and all these are easily explained as being derivatives of the BI and BII haplotypes. However 'hybrid' haplotypes (eg. α1=Tyr 24, α2=Phe 24) have not been observed in the screening of thousands of horses (Clegg, 1987). Both the BI and BII haplotypes also occur in Przewalski's horse (Ryder *et al.*, 1979) indicating that these haplotypes predate the formation of the domestic horse species. This situation differs from that of transferrin in that it involves two separate genes, but it may be evidence for another region in which recombination is infrequent.

Alternatively, the two distinct groups of horse transferrin variants could be a result of selection against the hybrid types. The ancestral horse transferrin could have been such a hybrid type. Random mutations would have produced separate progenitors of the D- and R-types, each of which were selected in favour of the hybrid type. Each type then accumulated substitutions, and consequently the two types diverged. Possibly, the two types evolved in separate populations and were subsequently brought together by cross-breeding. The two types would have been selectively neutral with respect to each other. At any stage, including at present, any hybrid types formed through recombination would have been selected against, and therefore eliminated. This model could be tested by constructing the hybrid variants, and attempting to determine any functional differences between them and the natural variants.

Generally, synonymous substitutions are observed more frequently in the coding regions of genes than are amino acid-altering substitutions (Kimura, 1983). This is due to selection acting to remove many of the mutations which cause amino acid substitutions. The ratio of

synonymous to non-synonymous substitutions in exon 15 (3:5) is therefore odd, although the data is so limited this cannot be taken too seriously. It does suggest that amino acid substitutions in this region are not subject to elimination by natural selection. Therefore, although the absence of insertions/deletions in this region suggests that its length may be important, it appears that the types of amino acids at many positions in this region is of lesser importance. This suggests that it would be unlikely that hybrid variants would be removed by natural selection.

A low frequency of recombination explains how the two types of horse transferrin maintain their differences, but not why both types are present in three diverse equid species. As the R-type has been lost in Przewalski's horse, the time required for the loss of an allele is short compared to the time since the zebra, domestic horse and donkey species diverged. These species could have retained members of both groups of transferrins simply through random events, or selection could have had a role in maintaining this genetic diversity through mechanisms such as heterozygous advantage or frequency-dependent selection.

I have stated that there are no horse transferrin variants which share characteristics of both the D- and R-types. In fact there are two variants, J and M, which have this potential. They each have an electrophoretic mobility between that of the D- and R-group variants which were studied. The J variant has been assigned to the D-group, and M to the R-group, according to immunological reactivity (Kaminski *et al.*, 1981). Both variants are relatively rare, being found only in a few breeds. It would be interesting to study these variants to determine whether they do have characteristics of both the D- and R-types.

The extent of the difference in amino acid sequence between the D- and R-groups of horse transferrins might be expected to indicate a functional difference between the two groups. Transferrin heterozygosity has been associated with a healthy endometrium in mares (Weitkamp *et al.*, 1991). The FF phenotype has been associated with superior racing performance in thoroughbreds (Osterhoff *et al.*, 1974). It might be expected that these associations between phenotype and function would reflect the two groups of horse transferrins. For example, is the heterozygous advantage in mares due mainly to D-type/R-type heterozygotes? Inspection of the data of Weitkamp *et al.* (1991) shows that the heterozygous advantage is due to DF and OR phenotypes as much as DO, DR, FO and FR. Thus the differences between the two groups do not appear to be significant in this case. The correlation between the F variant and racing performance might be expected to be shared by the other D-

type variants. The data of Osterhoff *et al.* (1974) indicate that D variants may be advantageous compared to O and R regarding racing performance, so perhaps the two groups are reflected in this case, though the data is somewhat limited.

4.3.6 Positions of polymorphisms in the primary structure

All the polymorphisms found in the horse transferrin sequence were in the C-terminal half of the molecule (Figure 33). Is this a consequence of the method of looking for them, or does it represent the real distribution of variation within the molecule? Admittedly, attention was focussed on this half of the molecule, initially due to the findings of Chung and McKenzie (1985) who detected 2 polymorphisms in the amino acid sequence after peptide mapping the whole molecule, i.e. they did not focus on the C-lobe. The SSCP analysis covered almost the whole molecule when using the DF_1 sample, and found two polymorphic fragments, both in the C-terminal half. The SSCP analysis using DF_2 and F_1F_1 samples did concentrate on the 3' end, for reasons of experimental convenience, but did not detect any further variation. Therefore, although experimental methods did focus on the C-terminal half-molecule to some degree, I believe the greater variability in this region is at least partially representative of the real situation. This question can only be resolved by complete sequencing of either the O or R allele and comparison with the D/F_1 sequence already obtained.

Assuming then, that the C-lobe of horse transferrin shows greater variability between alleles than the N-lobe, does this parallel the situation when inter-species comparisons are considered? The C-terminal portion of the molecule does differ more between species than the N-terminal half when serum transferrins are compared, but the opposite is true of lactoferrins (Table 5).

The transferrins are thought to have evolved via a gene duplication, resulting in the duplicated protein structure. The two lobes would initially have been identical in sequence, but differed in their position within the molecule, and may therefore have differed functionally. This situation could result in one lobe (the N-lobe in serum transferrins) being selectively constrained while the other is comparatively free from selective constraints and is able to evolve a little more rapidly, perhaps taking on a slightly modified function.

Table 5

Interspecies Comparisons of Mammalian Serum Transferrins and Lactoferrins

		% Identical Residues			
		N-lobe	C-lobe	Exon 15	Total
Transferrins					
Horse	Human	76.1	70.6	50.0	73.4
Horse	Pig	74.9	72.6	46.4	73.7
Horse	Rabbit	74.6	70.9	64.3	72.8
Human	Pig	71.9	71.1	53.6	71.5
Human	Rabbit	81.9	75.9	75.0	79.0
Pig	Rabbit	73.4	72.9	67.9	73.1
Mean		75.5	72.3	59.5	73.9
Lactoferrins					
Bovine	Human	65.6	73.1	75.0	69.4
Bovine	Mouse	58.8	68.6	75.0	63.6
Bovine	Pig	70.5	76.0	78.6	73.3
Human	Mouse	66.9	73.7	75.0	70.3
Human	Pig	66.6	75.4	71.4	71.0
Mouse	Pig	59.4	70.0	78.6	64.7
Mean		64.6	72.8	75.6	68.7

Percentage identity was calculated using the Gap program of the GCG package. The N and C lobes were divided between residues 357 and 358 in horse transferrin, and in the equivalent position in other sequences. Exon 15 was residues 580-607 of horse transferrin, and the equivalent region of other sequences.

The degree of polymorphism is concentrated particularly in exon 15. In the 85 bp lying between the PCR primers, 9 base substitutions distinguish the D, F_1, F_2, H_2, * alleles from the O, R alleles, producing 5 amino acid differences between the two groups. This region of the molecule does appear to be more variable than the average for the whole molecule (see Table 5). However in the lactoferrins (human, bovine, mouse) exon 15 is, if anything, a little less variable than the average for the whole molecule. This implies that if this region is functionally significant, the significance differs between serum transferrin and lactoferrin.

4.3.7 Positions of polymorphisms in the 3-dimensional structure

The 3-dimensional structures of rabbit transferrin and human lactoferrin have been determined (Bailey *et al.*, 1988; Anderson *et al.*, 1989). The two proteins are very similar in the overall structure, so it is likely that the horse transferrin also has a similar structure. The paper reporting the human lactoferrin structure (Anderson *et al.*, 1989) included a stereo diagram of an α-carbon chain trace with some residues numbered. Less detail was published for rabbit transferrin, so it has been easier to analyse the horse transferrin polymorphisms using human lactoferrin as a model structure, rather than rabbit transferrin.

When the positions of polymorphisms found in horse transferrin were located on the 3-dimensional structure of human lactoferrin (Figure 39), they appear to be clustered in a particular portion of the C-lobe, i.e. the end furthest from the N-lobe. The polymorphisms all also appear to be on external loops of the protein, such that they may be exposed on the surface of the molecule. Interestingly, two amino acid substitutions distinguishing bovine transferrin variants A and D2 (Maeda *et al.*, 1984) are located on external loops of the molecule adjacent to the cluster of polymorphisms in horse transferrin (i.e. at positions equivalent to residues 470 and 554 in horse transferrin). The only amino acid substitution in human transferrin whose position is known lies in the N-lobe, in a position equivalent to residue 321 of horse transferrin. This is close to the part of the N-lobe which is equivalent to the C-lobe region which is polymorphic in horse transferrin (Wang *et al.*, 1967).

Why might the polymorphic residues in horse transferrin all occur in one region of the molecule? As outlined in Section 4.3.6, the distribution of substitution sites in the primary structure could be partly due to experimental bias. This bias would also influence the distribution of these sites in the 3-dimensional structure. However, the folding of the molecule

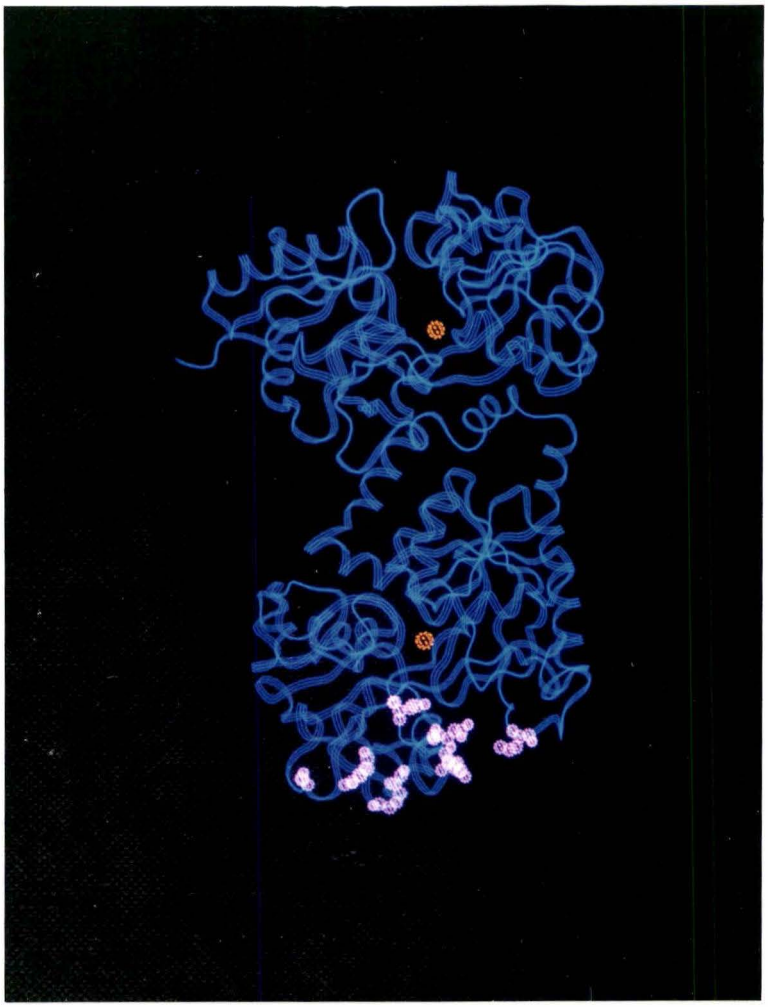


Figure 39. Positions of the horse transferrin amino acid sequence polymorphisms on the 3-dimensional structure of human lactoferrin. Human lactoferrin is represented as a ribbon diagram, with the N-lobe above the C-lobe. The two orange spheres represent iron atoms. Pink spheres represent the side-chain atoms of the residues which are polymorphic in horse transferrin. Note the clustered distribution of the polymorphic residues. Photograph courtesy of Dr. Ted Baker, Department of Chemistry and Biochemistry, Massey University.

brings polymorphic sites which are relatively distant in the primary structure close together in the tertiary structure, which is not readily explained by chance. This, together with the two bovine polymorphic sites in the adjacent region, indicates that this variable region is a genuine observation. This conclusion can only be confirmed by further sequencing to determine the level of polymorphisms throughout the molecule.

The clustered distribution of the polymorphic sites suggests that this variable region of the molecule may form a functional unit. The high variability of this region may indicate that it is functionally unimportant. The residues that are variable are on external parts of the molecule, and provided they are not involved in interactions with other molecules (eg. the transferrin receptor) the observed substitutions might be insignificant. If this is the case, it seems likely that there would be other parts of the molecule which are equally variable. Although both the N- and C-lobes of transferrin are required for receptor binding (Esparza and Brock, 1980; Brown-Mason and Woodworth, 1984), there is no indication that almost the whole molecule interacts with the receptor. Therefore, unless other external parts of transferrin are constrained by the requirements of interacting with other molecules, there must be other parts which are essentially functionless, and might be as variable as the region I have observed. Further sequencing would clarify this.

Alternatively, the polymorphic region could have a function which is favoured by heterozygosity, which explains both the large number of horse transferrin variants and the substantial differences between the two groups. Transferrin heterozygosity has been associated with a healthy endometrium, which is an indicator of fertility in mares (Weitkamp *et al.*, 1991). Weitkamp *et al.* (1991) suggested that this effect was due to the antimicrobial role of transferrin, the health of the endometrium being maintained better through more effective protection from pathogens in transferrin heterozygotes. It is possible that it is the polymorphic region of the molecule which is important in this heterozygous advantage.

It is unlikely that the function of the polymorphic region would be related to binding the transferrin receptor. A receptor binding site is more likely to be a conserved region than a variable one. The region of human transferrin which binds the receptor was shown to be phylogenetically conserved among human, pig , rabbit, horse and dog transferrins, by monoclonal antibody-binding studies (Bartek *et al.*, 1985). In contrast, the sequence of the exon 15 region is more variable between species than the rest of the molecule (Table 5). In addition, Mason *et al.* (1990) used antibodies to determine whether a disulphide bonded loop

in chicken ovotransferrin (residues 570-584) was the receptor binding site. This region is equivalent to residues 590-604 in horse transferrin, which is part of the variable region I have observed. They found this loop to be remote from the receptor binding site.

The part of the variable region encoded by exon 15 has an amino acid sequence which is more variable between species than the average for the molecule. However, it contains a structural feature which is highly conserved. This is the disulphide bonded loop selected for study by Mason *et al.* (1990). There are no insertions/deletions in this region in any of the proteins shown in Figure 38, so that its length is conserved exactly. The equivalent region in the N-lobe of the transferrins (encoded by exon 7) is very similar and is likewise conserved between proteins/species. This conserved structural feature may be have some functional significance.

4.3.8 Interpretation of the polymorphism data

Several differences have been demonstrated between the D- and R-groups of horse transferrin, in a particular region of the molecule. This limited data does not reveal the extent of the differences between the two groups. It may be that there are few differences other than those reported here, in which case the two groups do not differ much. Alternatively, the level of polymorphisms could be equally high throughout the whole molecule, which would negate the observation that the polymorphisms defined a variable region of the molecule. This situation would be clarified by obtaining the sequence of an O or R variant and comparing it to the D/F_1 sequence reported here. If many additional polymorphic positions in the protein sequence were located, but these were grouped in particular regions, this approach might define functional regions of the molecule. For example, a non-variable surface region might be the receptor-binding site.

4.4 CONCLUSIONS

The cDNA and amino acid sequences of horse transferrin show it to be very similar to other mammalian serum transferrins. Conserved features of the amino acid sequence such as iron and anion binding residues and cysteine residues indicate structural and functional similarity to other transferrins.

Three polymorphic nucleotide positions were found by sequencing fragments of the cDNA encoding horse transferrin variants D and F_1. Two of these produced amino acid substitutions distinguishing the D and F_1 variants.

Ten polymorphic nucleotide positions which distinguish the D, F_1, F_2, H_2 and * alleles of horse transferrin from the O and R alleles were found by Southern blotting of genomic DNA and/or by sequencing parts of the exons from these alleles. These nucleotide substitutions resulted in six amino acid sequence polymorphisms.

The number of substitutions distinguishing the D- and R-groups of horse transferrin in the small portion of the coding region which was sequenced indicates substantial differences between the two groups of variants. This is in accordance with published results which reported differences between the two groups in characteristics such as immunological reactivity, electrophoretic mobility, serum levels and RFLPs. The presence of two distinct groups of transferrin variants may indicate a low recombination frequency in all or part of the horse transferrin gene. Two similar groups of transferrin variants may also be present in other equid species such as the donkey and zebra, indicating that the precursors of the two groups diverged over 4 million years ago.

The positions of the eight amino acid substitutions observed among horse transferrin variants were located on the 3-dimensional structure of human lactoferrin. All eight substitutions were located in one part of the C-lobe, on external loops. This may define a region of the molecule in which variability is high due to a lack of functional constraints.

BIBLIOGRAPHY

Abola, J.E., Wood, M.K., Chweh, A., Abraham, D., Pulsinelli, P.D. (1982) Structure of hen ovotransferrin at 5 Å resolution. pp.27-34 in The Biochemistry and Physiology of Iron. Saltman, P., Hegenauer, J., eds. Elsevier Biomedical, New York

Adrian, G.S., Bowman, B.H., Herbert, D.C., Weaker, F.J., Adrian, E.K., Robinson, L.K., Walter, C.A., Eddy, C.A., Riehl, R., Pauerstein, C.J., Yang, F. (1990) Human transferrin: expression and iron modulation of chimeric genes in transgenic mice. J. Biol. Chem. 265: 13344-13350

Adrian, G.S., Korinek, B.W., Bowman, B.H., Yang, F. (1986) The human transferrin gene: 5' region contains conserved sequences which match the control elements regulated by heavy metals, glucocorticoids and acute phase reaction. Gene 49: 167-175

Aisen, P., Leibman, A. (1972) Lactoferrin and transferrin: a comparative study. Biochim. Biophys. Acta 257: 314-323

Aisen, P., Leibman, A., Sia, C.-L. (1972) Molecular weight and subunit structure of hagfish transferrin. Biochemistry 11: 3461-3464

Aisen, P., Listowsky, I. (1980) Iron transport and storage proteins. Annu. Rev. Biochem. 49: 357-393

Aldred, A.R., Dickson, P.W., Marley, P.D., Schreiber, G. (1987) Distribution of transferrin synthesis in brain and other tissues in the rat. J. Biol. Chem. 262: 5293-5297

Aldred, A.R., Howlett G.J., Schreiber, G. (1984) Synthesis of rat transferrin in *Escherichia coli* containing a recombinant bacteriophage. Biochem. Biophys. Res. Commun. 122: 960-965

Alexander, L.J., Levine, W.B., Teng, C.T., Beattie, C.W. (1992) Cloning and sequencing of the porcine lactoferrin cDNA. Animal Genetics 23:251-256

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

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

Archibald, F. (1983) *Lactobacillus plantarum*, an organism not requiring iron. FEMS Microbiol. Lett. 19: 29-32

Arnold, R.R., Cole M.F., McGhee, J.R. (1977) A bactericidal effect for human lactoferrin. Science 197: 263-265

Ashton, G.C. (1965) Cattle serum transferrins: a balanced polymorphism? Genetics 52: 983-997

Auconi, P., Biagini, R., Colarizi, P., Pascali, V. (1982) Transferrin C subtypes in extremely premature newborn infants. Pediatr. Res. 16: 1022-1024

Ayala, F.J., Kiger, J.A. Jr. (1984) Modern Genetics, 2nd edition. Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA.

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

Bailey, E., Lear T.L., Cothran, E.G. (1991) Association of *Msp*I restriction fragment length polymorphisms with transferrin in horses. Animal Genetics 22: 436

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) Molecular structure of serum transferrin at 3.3 Å resolution. Biochemistry 27: 5804-5812

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

Baker E.N., Rumball, S.V., Andeson, B.F. (1987) Transferrins: insights into structure and function from studies on lactoferrin. Trends Biochem. Sci. 12: 350-353

Baldwin, G.S., Bacic, T., Chandler, R., Grego, B., Pedersen, J., Simpson, R.J., Toh, B.H., Weinstock, J. (1990) Isolation of transferrin from porcine gastric mucosa: comparison with porcine serum transferrin. Comp. Biochem. Physiol. 95B: 261-268

Baldwin, G.S., Weinstock, J. (1988) Nucleotide sequence of porcine liver transferrin. Nucleic Acids Res. 16: 8720

Bali, P.K., Aisen, P. (1991) Receptor-modulated iron release from transferrin: differential effects on the N- and C-terminal sites. Biochemistry 30: 9947-9952

Bali, P.K., Zak, O., Aisen, P. (1991) A new role for the transferrin receptor in the release of iron from transferrin. Biochemistry 30: 324-328

Banfield, D.K., Chow, B.K.-C., Funk, W.D., Robertson, K.A., Umelas, T.M., Woodworth, R.C., MacGillivray, R.T.A. (1991) The nucleotide sequence of rabbit liver transferrin cDNA. Biochim. Biophys. Acta 1089: 262-265

Baranov, V.S., Schwartzman, A.L., Gorbunova V.N., Gaitskhoki, V.S., Rubtsov, N.B., Timchenko, N.A., Neifakh, S.A. (1987) Chromosomal localization of ceruloplasmin and transferrin genes in laboratory rats, mice and in man by hybridization with specific DNA probes. Chromosoma 96: 60-66

Bartek, J., Viklicky, V., Stratil, A. (1985) Phylogenetically more conservative epitopes among monoclonal antibody-defined antigenic sites of human transferrin are involved in receptor binding. Br. J. Haematol. 59: 435-441

Bartfeld, N.S., Law J.H. (1990) Isolation and molecular cloning of transferrin from the tobacco hornworm, *Manduca sexta*. J. Biol. Chem. 265: 21684-21691

Beach, R.L., Popiela, H., Festoff, B.W. (1983) The identification of neurotrophic factor as a transferrin. FEBS Lett. 156: 151-156

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

Bellamy, W., Takase, M., Yamauchi, K., Wakabayashi, H., Kawase, K., Tomita, M. (1992) Identification of the bactericidal domain of lactoferrin. Biochim. Biophys. Acta 1121: 130-136

Bernstein, S.E. (1987) Hereditary hypotransferrinemia with hemosiderosis, a murine disorder resembling human atransferrinemia. J. Lab. Clin. Med. 110: 690-705

Bethesda Research Laboratories Catalogue and Reference Guide (1988). Bethesda Research Laboratories Life Technologies, Inc. Gaithersburg, MD.

Bowling, A.T. (1991) An unusual transferrin variant in a thoroughbred 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.

Bowling, A.T., Ryder O.A. (1987). Genetic studies of blood markers in Przewalski's horses. J. of Hered. 78: 75-80.

Bowman, B.H., Adrian, G.S., Yang, F. (1987) Expression of genes encoding the vitamen D binding protein and transferrin. Protides Biol. Fluids Proc. Colloq. 35: 3-12

Bowman, B.H., Yang, F., Adrian, G.S. (1988) Transferrin: evolution and genetic regulation of expression. Adv. Genet. 25: 1-38

Bowman, B.H., Yang, F., Brune, J.L., Naylor, S.L., Barnett, D.R., McGill, J.R., Moore, C.M., Lum, J.B., McCombs, J. (1985) Organization and chromosomal locations of genes encoding human plasma proteins. Protides Biol. Fluids Proc. Colloq. 33: 15-20

Braend, M. (1964) Serum types of Norwegian horses. Nord. Veterinaermed. 16: 363-373

Broad, T.E., Forrest, J.W., Pugh, P.A. (1988) Effect of NaOH on hybridization efficiency of DNA. Trends Genet. 4: 146.

Brock, J.H. (1985) Transferrins. pp.183-262 in Metalloproteins (Part 2). Harrison, P.M., ed. Macmillan, London

Brown, J.P., Hewick, R.M., Hellstrom, I., Hellstrom, K.E., Doolittle, R.F., Dreyer, W.J. (1982) Human melanoma-associated antigen p97 is structurally and functionally related to transferrin. Nature 296: 171-173

Brown-Mason, A., Woodworth, R.C. (1984) Physiological levels of binding and iron donation by complementary half-molecules of ovotransferrin to transferrin receptors of chick reticulocytes. J. Biol. Chem. 259: 1866-1873

Broxmeyer, H.E., DeSousa, M., Smithyman, A., Ralph, P., Hamilton, J., Kurland, J.I., Bognacki, J. (1980) Specificity and modulation of the action of lactoferrin, a negative feedback regulator of myelopoiesis. Blood 55: 324-333

Brunel, F., Ochoa, A., Schaeffer, E., Boissier, F., Guillou, Y., Cereghini, S., Cohen, G.N., Zakin, M.M. (1988) Interactions of DNA-binding proteins with the 5' region of the human transferrin gene. J. Biol. Chem. 263: 10180-10185

Buchta, R. (1991) Ovine lactoferrin: isolation from colostrum and characterization. J. Dairy Res. 58: 211-218

Bullen, J.J., Armstrong, J.A. (1979) The role of lactoferrin in the bactericidal function of polymorphonuclear leucocytes. Immunology 36: 781-791

Cawthon, R.M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P., White, R. (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure and point mutations. Cell 62: 193-201

Chen, L.-H., Bissell, M.J. (1987) Transferrin mRNA level in the mouse mammary gland is regulated by pregnancy and extracellular matrix. J. Biol. Chem. 262: 17247-17250

Chomczynski, P., Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162: 156-159

Chung, M.C.-M., McKenzie, H.A. (1985) Studies on equine transferrin-I. The isolation and partial characterization of the D and R variants. Comp. Biochem. Physiol. 80B: 287-297

Clegg, J.B. (1987) Gene conversions in the horse α-globin gene complex. Mol. Biol. Evol. 4: 492-503

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., Chambon, P. (1979a) Organisation and sequence studies of the 17-piece chicken conalbumin gene. Nature 282: 567-574

Cochet, M., Perrin, F., Gannon, F., Krust, A., Chambon, P. (1979b) Cloning of an almost full-length chicken conalbumin double-stranded cDNA. Nucleic Acids Res. 6: 2435-2452

Coddeville, B., Stratil, A., Wieruszeski, J.-M., Strecker, G., Montreuil, J., Spik, G. (1989) Primary structure of horse serotransferrin glycans. 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

Cunningham, P. (1991) The genetics of thoroughbred horses. Sci. Am. 264: 56-62

Davidson, L.A., Lonnerdal, B. (1986) Isolation and characterization of rhesus monkey milk lactoferrin. Pediatr. Res. 20: 197-201

Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C. (1978) A model of evolutionary change in proteins. pp.345-352 in Atlas of Protein Sequence and Structure. Vol. 5, Suppl. 3. Dayhoff, M.O. ed. National Biomedical Research Foundation, Washington, D.C.

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

Debruyne, V., Montreuil, J., Spik, G. (1984) Crossed immunoaffinity electrophoresis of human transferrin in normal and cirrhotic sera. Protides Biol. Fluids Proc. Colloq. 31: 63-68

Derisbourg, P., Wieruszeski, J.-M., Montreuil, J., Spik, G. (1990) Primary structure of glycans isolated from human leucocyte lactotransferrin. Biochem. J. 269: 821-825

Devereux, J., Haeberli, P., Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395

Diamond, A., Devine, J.M., Cooper, G.M. (1984) Nucleotide sequence of a human *Blym* transforming gene activated in Burkitt's lymphoma. Science 225: 516-519

Didkowski, S., Kaminski, M., Kerjan, P., Tomaszewska-Guszkiewicz, K., Zurkowski, M. (1984) Obtaining of pure transferrins D, M and R from equine serum and determination of transferrin level in relation to phenotype. Acta Physiol. Pol. 35: 529-538

Dierich, A., Gaub, M.-P., LePennec, J.-P., Astinotti, D., Chambon, P. (1987) Cell-specificity of the chicken ovalbumin and conalbumin promoters. EMBO J. 6: 2305-2312

Dorland, L., Haverkamp, J., Vliegenthart, J.F.G., Spik, G., Fournet, B., Montreuil, J. (1979) Investigation by 360-MHz ^{1}H-nuclear-magnetic-resonance spectroscopy and methylation analysis of the single glycan chain of chicken ovotransferrin. Eur. J. Biochem. 100: 569-574

Duguid, J.R., Bohmont, C.W., Liu, N., Tourtellotte, W.W. (1989) Changes in brain gene expression shared by scrapie and Alzheimer disease. Proc. Natl. Acad. Sci. U.S.A. 86: 7260-7264

Ekblom, P., Thesleff, I., Saxen, L., Miettinen, A., Timpl, R. (1983) Transferrin as a fetal growth factor: Acquisition of responsiveness related to embryonic induction. Proc. Natl. Acad. Sci. U.S.A. 80: 2651-2655

Ellison, R.T. III, Giehl, T.J., Laforce, F.M. (1988) Damage to the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. Infect. Immun. 56: 2774-2781

Engelke, D.R., Hoener, P.A., Collins, F.S. (1988) Direct sequencing of enzymatically amplified human genomic DNA. Proc. Natl. Acad. Sci. U.S.A. 85: 544-548

Enns, C.A., Larrick, J.W., Suomalainen, H., Schroder, J., Sussman, H.H. (1983) Co-migration and internalization of transferrin and its receptor on K562 cells. J. Cell Biol. 97: 579-585

Esparza, I., Brock, J.H. (1980) The interaction of bovine transferrin and monoferric transferrin fragments with rabbit reticulocytes. Biochim. Biophys. Acta 624: 479-489

Evans, R.W., Meilak, A., Aitkin, A., Patel, K.J., Wong, C., Garratt, R.C., Chitnavis, B. (1988) Characterization of the amino acid change in a transferrin variant. Biochem. Soc. Trans. 16: 834-835

Evans, R.W., Williams, J., Moreton, K. (1982) A variant of human transferrin with abnormal properties. Biochem. J. 201: 19-26

Farndale, B.M., Anderson, I.L., Moore, C.H. (1991) Identification of an apparent transferrin mutation in a commercial thoroughbred stallion. Animal Genetics 22 (Suppl. 1): 19-20

Fernandez-Pol, J.A., Klos, D.J. (1980) Isolation and characterization of normal rat kidney cell membrane proteins with affinity for transferrin. Biochemistry 19: 3904-3912

Frohman, M.A., Dush, M.K., Martin, G.R. (1988) Rapid production of full length cDNA's from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. U.S.A. 85: 8998-9002

Gahne, B. (1966) Studies on the inheritance of electrophoretic forms of transferrins, albumins, prealbumins and plasma esterases of horses. Genetics 53: 681-694

Gelb, M.H., Harris, D.C. (1980) Correlation of proton release and ultraviolet difference spectra associated with metal binding by transferrin. Arch. Biochem. Biophys. 200: 93-98

George, M. Jr., Ryder, O.A. (1986) Mitochondrial DNA evolution in the genus *Equus*. Mol. Biol. Evol. 3: 535-546

Gilmont, R.R., Coulter, G.H., Sylvester, S.R., Griswold, M.D. (1990) Synthesis of transferrin and transferrin mRNA in bovine sertoli cells in culture and *in vivo*: sequence of partial cDNA clone for bovine transferrin. Biol. Reprod. 43: 139-150

Goding, J.W., Harris, A.W. (1981) Subunit structure of cell surface proteins: Disulphide bonding in antigen receptors, Ly-2/3 antigens, and transferrin receptors of murine T and B lymphocytes. Proc. Natl. Acad. Sci. U.S.A. 78: 4530-4534

Goodman, R.E., Schanbacher, F.L. (1991) Bovine lactoferrin mRNA: sequence, analysis and expression in the mammary gland. Biochem. Biophys. Res. Commun. 180: 75-84

Goubin, G., Goldman, D.S., Luce, J., Neiman, P.E., Cooper, G.M. (1983) Molecular cloning and nucleotide sequence of a transforming gene detected by transfection of chicken B-cell lymphoma DNA. Nature 302: 114-119

Goya, N., Miyazaki, S., Kodate, S., Ushio, B. (1972) A family of congenital atransferrinemia. Blood 40: 239-245

Grigor, M.R., Bennett, B.L., Carne, A., Cowan, P.E. (1991) Whey proteins of the common brushtail possum (*Trichosurus vulpecula*): isolation, characterization and changes in concentration in milk during lactation of transferrin, α-lactalbumin and serum albumin. Comp. Biochem. Physiol. 98B: 451-459

Grigor, M.R., Carne, A., Geursen, A., Flint, D.J. (1988) Effect of extended lactation and diet on transferrin concentrations in rat milk. J. Nutr. 118: 669-674

Grigor, M.R., McDonald, F.J., Latta, N., Richardson, C.L., Tate, W.P. (1990) Transferrin-gene expression in the rat mammary gland. Independence of maternal iron status. Biochem. J. 267: 815-819

Gulamhusein, A.P., Pratten, M.K., Williams, C.A., Beck, F. (1990) The effect of macromolecular rat serum fractions on conceptuses cultured in human serum: role of transferrin. J. Anat. 168: 113-121

Gutteberg, T.J., Rokke, O., Anderson, O., Jorgensen, T. (1989) Early fall of circulating iron and rapid rise of lactoferrin in septicemia and endotoxemia: an early defence mechanism. Scand. J. Infect. Dis. 21: 709-715

Harris, D.C. (1977) Different metal-binding properties of the two sites of human transferrin. Biochemistry 16: 560-564

Harris, H. (1966) Enzyme polymorphisms in man. Proc. R. Soc. Lond. B 164: 298-310.

Harris, W.R. (1983) Thermodynamic binding constants of the zinc-human serum transferrin complex. Biochemistry 22: 3920-3926

Hayashi, K. (1991) PCR-SSCP:a simple and sensitive method for detection of mutations in genomic DNA. PCR Methods and Applications 1: 34-38

Heilmeyer, L. (1966) Die Atransferrinamien. Acta Haematol. 36: 40-49

Higgins, D.G., Bleasby, A.J., Fuchs, R. (1992). Clustal V: improved software for multiple sequence alignment. Comput. Appl. Biosci. 8: 189-191.

Higuchi, R. (1989) Using PCR to engineer DNA. pp. 61-70 in PCR Technology Principles and Applications for DNA Amplification. Erlich, H.A. ed. Stockton Press, New York.

Holmes, D.S., Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193-197.

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

Huebers, H.A., Huebers, E., Finch, C.A., Martin, A.W. (1982) Characterization of an invertebrate transferrin from the crab *Cancer magister* (Arthropoda). J. Comp. Physiol. 148: 101-109

Huebers, H.A., Huebers, E., Finch, C.A., Webb, B.A., Truman, J.W., Riddiford, L.M., Martin, A.W., Massover, W.H. (1988) Iron binding proteins and their roles in the tobacco hornworm, *Manduca sexta* (L.). J. Comp. Physiol. B 158: 291-300.

Huggenvik, J.I., Craven, C.M., Idzerda, R.L., Bernstein, S., Kaplan, J., McKnight, G.S. (1989) A splicing defect in the mouse transferrin gene leads to congenital atransferrinemia. Blood 74: 482-486

Huggenvik, J.I., Idzerda, R.L., Haywood, L., Lee, D.C., McKnight, G.S., Griswold, M.D. (1987) Transferrin messenger ribonucleic acid: Molecular cloning and hormonal regulation in rat Sertoli cells. Endocrinology 120: 332-340

Hutchens, T.W., Magnuson, J.S., Yip, T.-T. (1989) Rapid purification of porcine colostral whey lactoferrin by affinity chromatography on single-stranded DNA-agarose. Characterization, amino acid composition and N-terminal amino acid sequence. Biochim. Biophys. Acta 999: 323-329

Huynh, T.V., Young, R.A., Davis, R.W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. pp.49-78 in DNA Cloning: A Practical Approach Vol. 1. Glover, D.M., ed. IRL Press, Oxford.

Idzerda, R.L., Behringer, R.R., Theisen, M., Huggenvik, J.I., McKnight, G.S., Brinster, R.L. (1989) Expression from the transferrin gene promoter in transgenic mice. Mol. Cell. Biol. 9: 5154-5162

Idzerda, R.L., Huebers, H., Finch, C.A., McKnight, G.S. (1986) Rat transferrin gene expression: Tissue-specific regulation by iron deficiency. Proc. Natl. Acad. Sci. U.S.A. 83: 3723-3727

Irie, S., Minguell, J.J., Tavassoli, M. (1988) Analysis of the microheterogeneity of the glycan chain of rat transferrin. Biochem. Int. 17: 1079-1085

Iwase, H., Hotta, K. (1977) Ovotransferrin subfractionation dependent upon carbohydrate chain differences. J. Biol. Chem. 252: 5437-5443

Jeanpierre, M. (1987) A rapid method for the purification of DNA from blood. Nucleic Acids Res. 15: 9611

Jeltsch, J.-M., Chambon, P. (1982) The complete nucleotide sequence of the chicken ovotransferrin mRNA. Eur. J. Biochem. 122: 291-295

Jeltsch, J.-M., Hen, R., Maroteaux, L., Garnier, J.-M., Chambon, P. (1987) Sequence of the chicken ovotransferrin gene. Nucleic Acids Res. 15: 7643-7645

Jolles, J., Donda, A., Amiguet, P., Jolles, P. (1984) Mare lactotransferrin: purification, analysis and N-terminal sequence determination. FEBS Lett. 176: 185-188

Junankar, P.R., McKenzie, H.A., Shaw, D.C. (1990) Sequence studies of rabbit serum transferrin. Biochem. Int. 21: 243-250

Kaiser, K., Murray, N.E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. pp.1-47 in DNA Cloning: A Practical Approach Vol. 1. Glover, D.M., ed. IRL Press, New York.

Kaminski, M., Didkowski, S., Sykiotis, M. (1981) Polymorphism of transferrin locus in horses: Immunochemical evidence of two structurally different subgroups of the allelic proteins. Comp. Biochem. Physiol. 68B: 505-507

Kaminski, M., Urbanska-Nicolas, H. (1979) Electrophoretic polymorphism of proteins in the blood of horses: Studies of eleven pony breeds or populations. Biochem. Syst. Ecol. 7: 299-237

Kanda, Y., Maeda, M, Yamada, K., Yoshino, Y. (1983) Physiological properties of two types of rat transferrin (A and B). pp.399-400 in Structure and Function of Iron Storage and Transport Proteins. Urushizaki, I., Aisen, P., Listowsky, I., Drysdale, J.W., eds. Elsevier, Ansterdam

Kazazian, H.H. Jr., Chakravarti, A., Orkin, S.H., Antonarkis, S.E. (1983). DNA polymorphisms in the human β-globin gene cluster. pp.137-146 in Evolution of Genes and Proteins. Nei, M., Koehn, R.K., eds. Sinauer Associates, Inc. Sunderland, MA.

Kimura, M. (1983) The Neutral theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kojima, N., Bates, G.W. (1981) The formation of Fe^{3+}-transferrin-CO_3^{2-} via the binding and oxidation of Fe^{2+}. J. Biol. Chem. 256: 12034-12039

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

Kourilsky, F.M., Burtin, P. (1968) Immunochemical difference between iron-saturated and unsaturated human transferrin. Nature 218: 375-377

Kugler, W., Wagner, U., Ryffel, G.U. (1988) Tissue-specificity of liver gene expression: a common liver-specific promoter element. Nucleic Acids Res. 16: 3165-3174

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

Kuhnl, P., Spielmann, W. (1979) A third common allele in the transferrin system, Tf^{C3}, detected by isoelectric focusing. Human Genetics 50: 193-198

Laskey, J., Webb, I., Schulman, H.M., Ponka, P. (1988) Evidence that transferrin supports cell proliferation by supplying iron for DNA synthesis. Exp. Cell Res. 176: 87-95

Le Beau, M.M., Diaz, M.O., Plowman, G.D., Brown, J.P., Rowley, J.D. (1986) Chromosomal sublocalization of the human p97 melanoma antigen. Human Genetics 72: 294-296

Le Beau, M.M., Diaz, M.O., Yang, F., Plowman, G.D., Brown, J.P., Rowley, J.D. (1985) Molecular analysis of transferrin (TF), transferrin receptor (TFR), and p97 antigen genes in the inv(3) and t(3;3) in acute nonlymphocytic leukemia (ANLL). Am. J. Hum. Genet. 37: A31

Leclercq, Y., Sawatzki, G., Wieruszeski, J.-M., Montreuil, J., Spik, G. (1987) Primary structure of the glycans from mouse serum and milk transferrins. Biochem. J. 247: 571-578

Lee, D.C., McKnight, G.S., Palmiter, R.D. (1978a) The action of estrogen and progesterone on the expression of the transferrin gene. J. Biol. Chem. 253: 3494-3503

Lee, M.Y., Huebers, H., Martin, A.W., Finch, C.A. (1978b) Iron metabolism in a spider, *Dugesiella hentzi*. J. Comp. Physiol. 127: 349-354

Leger, D., Campion, B., Decottignies, J.-P., Montreuil, J., Spik, G. (1989) Physiological significance of the marked increased branching of the glycans of human serotransferrin during pregnancy. Biochem. J. 257: 231-238

Lescoat, G., Jouanolle, H., Desvergne, B., Pasdeloup, N., Kneip, B., Deugnier, Y., Guillouzo, A., Brissot, P. (1989) Effects of iron overload on transferrin secretion by cultured fetal rat hepatocytes. Biol. Cell 65: 221-229

Levin, M.J., Tuil, D., Uzan, G., Dreyfus, J.-C., Kahn, A. (1984) Expression of the transferrin gene during development of non-hepatic tissues. Biochem. Biophys. Res. Commun. 122: 212-217

Lewin, B. (1990). Genes IV. Oxford University Press, New York and Cell Press, Cambridge.

Liu, Y., Teng, C.T. (1991) Characterization of estrogen-responsive mouse lactoferrin promotor. J. Biol. Chem. 266: 21880-21885

Lombardi, A., Tramontano, D., Braverman, L.E., Ingbar, S.H. (1989) Transferrin in FRTL5 cells: regulation of its receptor by mitogenic agents and its role in growth. Endocrinology 125: 652-658

Lucero, M.A., Schaeffer, E., Cohen, G.N., Zakin, M.M. (1986) The 5' region of the human transferrin gene: structure and potential regulatory sites. Nucleic Acids Res. 14: 8692

MacGillivray, R.T.A., Mendez, E., Shewale, J.G., Sinha, S.K., Lineback-Zins, J., Brew, K. (1983) The primary structure of human serum transferrin. J. Biol. Chem. 258: 3543-3553

MacGillivray, R.T.A., Mendez, E., Sinha, S.K., Sutton, M.R., Lineback-Zins, J., Brew, K. (1982) The complete amino acid sequence of human serum transferrin. Proc. Natl. Acad. Sci. U.S.A. 79: 2504-2508

Maeda, K., McKenzie, H.A., Shaw, D.C. (1980) Nature of the heterogeneity within genetic variants of bovine serum transferrin. Anim. Blood Groups Biochem. Genet. 11: 63-75

Maeda, K., McKenzie, H.A., Shaw, D.C. (1984) Comparison of bovine serum transferrin A and D_2. I. Amino acid residue differences. Anim. Blood Groups Biochem. Genet. 15: 299-312

Martin, A.W., Huebers, E., Huebers, H., Webb, J., Finch, C.A. (1984) A mono-sited transferrin from a representative deuterostome: the ascidian *Pyura stolonifera* (Subphylum Urochordata). Blood 64: 1047-1052

Mason, A.B., Brown, S.A., Church, W.R. (1990) A highly conserved surface loop in the C-terminal domain of ovotransferrin (residues 570-584) is remote from the receptor-binding site. Biochem. J. 266: 393-398

Masson, P.L., Heremans, J.F., Dive, C.H. (1966) An iron-binding protein common to many external secretions. Clin. Chim. Acta 14: 735-739

Mazurier, J., Metz-Boutigue, M.-H., Jolles, J., Spik, G., Montreuil, J., Jolles, P. (1983) Human lactotransferrin: Molecular, functional and evolutionary comparisons with human serum transferrin and hen ovotransferrin. Experientia 39: 135-141

McCombs, J.L., Teng, C.T., Pentecost, B.T., Magnuson, V.L., Moore, C.M., McGill, J.R. (1988) Chromosomal localization of human lactotransferrin gene (LTF) by *in situ* hybridization. Cytogenet. Cell Genet. 47: 16-17

McKnight, G.S., Hammer, R.E., Kuenzel, E.A., Brinster, R.L. (1983) Expression of the chicken transferrin gene in transgenic mice. Cell 34: 335-341

McKnight, G.S., Lee, D.C., Hemmaplardh, D., Finch, C.A., Palmiter, R.D. (1980) Transferrin gene expression: Effects of nutritional iron deficiency. J. Biol. Chem. 255: 144-147

McKnight, G.S., Palmiter, R.D. (1979) Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. 254: 9050-9058

Mead, P.E., Tweedie, J.W. (1990) cDNA and protein sequence of bovine lactoferrin. Nucleic Acids Res. 18: 7167

Mendelzon, D., Boissier, F., Zakin, M.M. (1990) The binding site for the liver-specific transcription factor Tf-LF1 and the TATA box are the only elements necessary to direct liver-specific transcription *in vitro*. Nucleic Acids Res. 18: 5717-5721

Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J., Jolles, P. (1984) Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins. Eur. J. Biochem. 145: 659-676

Molloy, A.L., Winterbourn, C.C. (1990) Release of iron from phagocytosed *Escherichia coli* and uptake by neutrophil lactoferrin. Blood 75: 984-989

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

Moskaitis, J.E., Pastori, R.L., Schoenberg, D.R. (1990) The nucleotide sequence of *Xenopus laevis* transferrin mRNA. Nucleic Acids Res. 18: 6135

Neilands, J.B. (1981) Microbial iron compounds. Annu. Rev. Biochem. 50: 715-731

Omary, M.B., Trowbridge, I.S. (1981) Covalent binding of fatty acid to the transferrin receptor in cultured human cells. J. Biol. Chem. 256: 4715-4718

Orita, M., Suzuki, Y., Sekiya, T., Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879

Osterhoff, D.R. (1966) Hemoglobin, transferrin and albumin types in Equidae (horses, mules, donkeys and zebras). Proceedings of the 10th European Conference of Blood Groups and Biochemical Polymorphisms, Paris, 1966. pp. 345-351

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

Paabo, S., Higuchi, R.G., Wilson, A.C. (1989). Ancient DNA and the polymerase chain reaction. J. Biol. Chem. 264: 9709-9712.

Palmiter, R.D. (1972) Regulation of protein synthesis in chick oviduct. J. Biol. Chem. 247: 6450-6461

Palmour, R.M., Sutton, H.E. (1971) Vertebrate transferrins. Molecular weights, chemical compositions, and iron-binding studies. Biochemistry 10: 4026-4032

Park, I., Schaeffer, E., Sidoli, A., Baralle, F.E., Cohen, G.N., Zakin, M.M. (1985) Organization of the human transferrin gene: Direct evidence that it originated by gene duplication. Proc. Natl. Acad. Sci. U.S.A. 82: 3149-3153

Penhallow, R.C., Mason, A.B., Woodworth, R.C. (1991) Electrophoretic characterization of human, equine and bovine transferrins. Comp. Biochem. Physiol. 98B: 41-45

Pentecost, B.T., Teng, C.T. (1987) Lactotransferrin is the major estrogen inducible protein of mouse uterine secretions. J. Biol. Chem. 262: 10134-10139

Perrin, F., Cochet, M., Gerlinger, P., Cami, B., LePennec, J.P., Chambon, P. (1979) The chicken conalbumin gene: studies of the organization of cloned DNAs. Nucleic Acids Res. 6: 2731-2748

Pierce, A., Colavizza, D., Benaissa, M., Maes, P., Tartar, A., Montreuil, J., Spik, G. (1991) Molecular cloning and sequence analysis of bovine lactotransferrin. Eur. J. Biochem. 196: 177-184

Powell, M.J., Ogden, J.E. (1990) Nucleotide sequence of human lactoferrin cDNA. Nucleic Acids Res. 18: 4013

Promega Biological Research Products Catalogue (1990/91). Promega Corporation, Madison, WI.

Promega Protocols and Applications Guide (1st edition). Promega Corporation, Madison, WI.

Putt, W., Whitehouse, D.B. (1983) Genetics of four plasma protein loci in *Equus przewalski*: new alleles at prealbumin, postalbumin and transferrin loci. Animal Genetics 14: 7-16

Rado, T.A., Wei, X., Benz, E.J. Jr. (1987) Isolation of lactoferrin cDNA from a human myeloid library and expression of mRNA during normal and leukemic myelopoiesis. Blood 70: 989-993

Rasmusen, B.A., Tucker, E.M. (1973) Transferrin types and reproduction in sheep. Anim. Blood Groups Biochem. Genet. 4: 207-220

Reed, K.C., Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. 13: 7207-7221.

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

Reiter, B. (1978) Review of the progress of dairy science: antimicrobial systems in milk. J. Dairy Res. 45: 131-147

Rey, M.W., Woloshuk, S.L., deBoer, H.A., Pieper, F.R. (1990) Complete nucleotide sequence of human mammary gland lactoferrin. Nucleic Acids Res. 18: 5288

Richardson, D.R., Baker, E. (1990) The uptake of iron and transferrin by the human malignant melanoma cell. Biochim. Biophys. Acta 1053: 1-12

Richardson, D.R., Baker, E. (1991) The release of iron and transferrin from the human melanoma cell. Biochim. Biophys. Acta 1091: 294-302

Rose, T.M., Plowman, G.D., Teplow, D.B., Dreyer, W.J., Hellstrom, K.E., Brown, J.P. (1986) Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence. Proc. Natl. Acad. Sci. U.S.A. 83: 1261-1265

Roskams, A.J., Connor, J.R. (1990) Aluminium access to the brain: a role for transferrin and its receptor. Proc. Natl. Acad. sci. U.S.A. 87: 9024-9027

Ryder, O.A., Sparkes, R.S., Sparkes, M.C., Clegg, J.B. (1979) Hemoglobin polymorphism in *Equus przewalski* and *E. caballus* analyzed by isoelectric focussing. Comp. Biochem. Physiol. 62B: 305-308

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning A Laboratory Manual, 2nd edition. Cold Spring Harbor Laboratory Press. New York

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 2.3 Å resolution. Acta Crystallogr. B46: 763-771

Schade, A.L., Caroline, L. (1944) Raw hen egg white and the role of iron in growth inhibition of *Shigella dysenteriae, Staphylococcus aureus, Escherichia coli* and *Saccharomyces cerevisiae.* Science 100: 14-15

Schaeffer, E., Huchenq, A., Park, I., Lucero, M.A., Cohen, G.N., Zakin, M.M., Constans, J. (1985) Gene organization and DNA polymorphism of the human transferrin gene. Protides Biol. Fluids Proc. Colloq. 33: 131-134

Schaeffer, E., Lucero, M.A., Jeltsch, J.-M., Py, M.-C., Levin, M.J., Chambon, P., Cohen, G.N., Zakin, M.M. (1987) Complete structure of the human transferrin gene. Comparison with analogous chicken gene and human pseudogene. Gene 56: 109-116

Schmid, 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., Phillips, J., Edwards, K., Maggs, J. (1979) The synthesis and secretion of rat transferrin. J. Biol. Chem. 254: 12013-12019

Schryvers, A.B., Gonzalez, G.C. (1990) Receptors for transferrin in pathogenic bacteria are specific for the host's protein. Can. J. Microbiol. 36: 145-147

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

Shapiro, L.E., Wagner, N. (1989) Transferrin is an autocrine growth factor secreted by Reuber H-35 cells in serum-free culture. In Vitro Cell. & Dev. Biol. 25: 650-654

Shimazaki, K.-I., Kawano, N., Yoo, Y.C. (1991) Comparison of bovine, sheep and goat milk lactoferrins in their electrophoretic behavior, conformation, immunochemical properties and lectin reactivity. Comp. Biochem. Physiol. 98B: 417-422

Shirsat, N.V., Bittenbender, S., Kreider, B.L., Rovera, G. (1992) Structure of the murine lactotransferrin gene is similar to the structure of other transferrin-encoding genes and shares a putative regulatory region with the murine myeloperoxidase gene. Gene 110: 229-234

Short, R.V., Chandley, A.C., Jones, R.C., Allen, W.R. (1974) Meiosis in interspecific equine hybrids II. The Przewalski horse/domestic horse hybrid (*Equus przewalski x E. caballus*). Cytogenet. Cell genet. 13: 465-478

Singhvi, N.M., Khanna, N.D. (1988) Transferrin polymorphism in Indian horses, mules and donkeys. Indian Vet. J. 65: 39-43

Soltys, H.D., Brody, J.I. (1970) Synthesis of transferrin by human peripheral blood lymphocytes. J. Lab. Clin. Med. 75: 250-257

Spik, G., Coddeville, B., Montreuil, J. (1988) Comparative study of the primary structures of sero-, lacto-, and ovotransferrin glycans from different species. Biochimie 70: 1459-1469

Stallard, B.J., Collard, M.W., Griswold, M.D. (1991) A transferrinlike (hemiferrin) mRNA is expressed in the germ cells of rat testis. Mol. Cell. Biol. 11: 1448-1453

Stratil, A., Glasnak, V. (1981) Partial characterization of horse transferrin heterogeneity with respect to the atypical type, Tf C. Anim. Blood Groups Biochem. Genet. 12: 113-122

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., Bobak, P., Glasnak, V. (1984) Heterogeneity of horse transferrin: the role of carbohydrate moiety. Anim. Blood Groups Biochem. Genet. 15: 89-101

Sutherland, R., Delia, D., Schneider, C., Newman, R., Kemshead, J., Greaves, M. (1981) Ubiquitous cell-surface glycoprotein on tumour cells is proliferation-associated receptor for transferrin. Proc. Natl. Acad. Sci. U.S.A. 78: 4515-4519

Suzuki, T., Yamauchi, K., Kawase, K., Tomita, M., Kiyosawa, I., Okonogi, S. (1989) Collaborative bacteriostatic activity of bovine lactoferrin with lysozyme against *Escherichia coli* 0111. Agric. Biol. Chem. 53: 1705-1706

Talmadge, K., Gilbert, W. (1980) Construction of plasmid vectors with unique *Pst*I cloning sites in a signal sequence coding region. Gene 12: 235-241

Teahan, C.G., McKenzie, H.A. (1990) Iron(III) binding proteins of echidna (*Tachyglossus aculeatus*) and platypus (*Ornithorhynchus anatinus*). Biochem. Int. 22: 321-328

Teng, C.T., Pentecost, B.T., Marshall, A., Solomon, A., Bowman, B.H., Lalley, P.A., Naylor, S.L. (1987) Assignment of the lactotransferrin gene to human chromosome 3 and to mouse chromosome 9. Somatic Cell Mol. Genet. 13: 689-693

Thibodeau, S.N., Lee, D.C., Palmiter, R.D. (1978) Identical precursors for serum transferrin and egg white conalbumin. J. Biol. Chem. 253: 3771-3774

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

Tormey, D.C., Mueller, G.C. (1972) Biological effects of transferrin on human lymphocytes *in vitro*. Exp. Cell Res. 74: 220-226

Trowbridge, I.S., Omary, M.B. (1981) Human cell surface glycoprotein related to cell proliferation is the receptor for transferrin. Proc. Natl. Acad. Sci. U.S.A. 78: 3039-3043

Tu, G.-F., Achen, M.G., Aldred, A.R., Southwell, B.R., Schreiber, G. (1991) The distribution of cerebral expression of the transferrin gene is species specific. J. Biol. Chem. 266: 6201-6208

van Eijk, H., van Noort, W., Kroos, M., Dekker, K., van der Heul, C. (1983) The heterogeneity of human transferrin in biological fluids. Differences in physiological functions. pp.261-264 in Structure and Function of Iron Storage and Transport Proteins. Urushizaki, I., Aisen, P., Listowsky, I., Drysdale, J.W., eds. Elsevier, Amsterdam

van Ginkel, M.F., van der Voet, G.B., van Eijk, H.G., de Wolff, F.A. (1990) Aluminium binding to serum constituents: a role for transferrin and for citrate. J. Clin. Chem. Clin. Biochem. 28: 459-463

Wang, A.-C., Sutton, H.E. (1965) Human transferrin C and D_1: chemical difference in a peptide. Science 149: 435-437

Wang, A.-C., Sutton, H.E., Howard, P.N. (1967) Human transferrins C and D_{Chi}: an amino acid difference. Biochem. Genet. 1: 55-59

Wang, A.-C., Sutton, H.E., Riggs, A. (1966) A chemical difference between human transferrins B_2 and C. Am. J. Hum. Genet. 18: 454-458

Wang, C.-S., Chan, W.-Y., Kloer, H.U. (1984) Comparative studies on the chemical and immunochemical properties of human milk, human pancreatic juice and bovine milk lactoferrin. Comp. Biochem. Physiol. 78B: 575-580

Weatherall, D.J., Clegg, J.B. (1976) Molecular genetics of human hemoglobin. Annu. Rev. Genet. 10: 157-178

Weinberg, E.D. (1978) Iron and infection. Microbiol. Rev. 42: 45-66

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, S. (1990) A comparison of the structure and properties of serum transferrin from 17 animal species. Comp. Biochem. Physiol. B. 97: 417-427

Welch, S., Langmead, L. (1990) A comparison of the structure and properties of normal human transferrin and a genetic variant of human transferrin. Int. J. Biochem. 22: 275-282

Wenn, R.V., Williams, J. (1968) The isoelectric fractionation of hen's-egg ovotransferrin. Biochem. J. 108: 69-74

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

Williams, J. (1968) A comparison of the glycopeptides from the ovotransferrin and serum transferrin of the hen. Biochem. J. 108: 57-67

Williams, J., Elleman, T.C., Kingston, I.B., Wilkins, A.G., Kuhn, K.A. (1982a) The primary structure of hen ovotransferrin. Eur. J. Biochem. 122: 297-303

Williams, J., Grace, S.A., Williams, J.M. (1982b) Evolutionary significance of the renal excretion of transferrin half-molecule fragments. Biochem. J. 201: 417-419

Wong, C.T., Saha, N. (1986) Effects of transferrin genetic phenotypes on total iron-binding capacity. Acta Haematol. 75: 215-218

Wong, C.T., Saha, N. (1991) Lack of influence of maternal and fetal transferrin phenotypes and concentrations on normal fetal growth. Biol. Neonate 59: 156-160

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

Yang, F., Friedrich, W.E., Buchanan, J.M., Herbert, D.C., Weaker, F.J., Brock, J.H., Bowman, B.H. (1990) Tissue specific expression of mouse transferrin during development and aging. Mech. Ageing Dev. 56: 187-197

Yang, F., Lum, J.B., Bowman, B.H. (1985) Molecular genetics of human transferrin. Protides Biol. Fluids Proc. Colloq. 33: 31-34

Yang, F., Lum, J.B., McGill, J.R., Moore, C.M., Naylor, S.L., van Bragt, P.H., Baldwin, W.D., Bowman, B.H. (1984) Human transferrin: cDNA characterization and chromosomal localization. Proc. Natl. Acad. Sci. U.S.A. 81: 2752-2756

Young, S.P., Bomford, A., Williams, R. (1983) Dual pathways for the uptake of rat asialotransferrin by rat hepatocytes. J. Biol. Chem. 258: 4972-4976

Zak, O., Aisen, P. (1986) Nonrandom distribution of iron in circulating human transferrin. Blood 68: 157-161

Zak, O., Aisen, P. (1990) Evidence for functional differences between the two sites of rabbit transferrin: effects of serum and carbon dioxide. Biochim. Biophys. Acta 1052: 24-28

Zaman, Z., Heynen, M.-J., Verwilghen, R.L. (1980) Studies on the mechanism of transferrin iron uptake by rat reticulocytes. Biochim. Biophys. Acta 632: 553-561