Haemochromatosis: genetic and functional studies

Jason Partridge

Royal Free and University College Medical School
University College London
University of London

Submitted for the degree of:

Doctor of Philosophy
Abstract

Genetic haemochromatosis (GH) is an autosomal recessive condition common in Northern Europeans. It causes excess absorption of iron, resulting in tissue damage. Two approaches were used to study GH:

Positional cloning: P1 derived artificial chromosomes (PACs), containing large inserts of human genomic DNA, were isolated from around D6S1260, at that time the observed peak of linkage disequilibrium with GH. These clones were used to screen a human small intestine cDNA library. Four of the cDNAs isolated from this library contained C2H2 zinc finger motifs, comprising a single locus. The cDNA and genomic sequence and expression pattern of this locus were determined. The locus has characteristics similar to that of an expressed processed pseudogene, although it retains an open reading frame of 1.2kb. No evidence for a parent locus at another chromosomal site was detected using sequence database screening, somatic cell hybrid analysis and fluorescent in situ hybridisation. The sequence conservation displayed by this zinc finger pseudogene makes it an excellent tool for the identification of zinc finger genes in model organisms.

Assessment of a cellular phenotype: a cellular phenotype for GH would allow functional cloning of the gene, as well as investigation of GH in vitro. The enzyme ferric reductase was previously shown to have increased activity in duodenal biopsies from GH patients compared to controls. This increase paralleled increased uptake of $^{59}$Fe in GH.

GH patients and controls for ferric reductase studies were characterised by investigation of both haplotype and $HFE$ mutations. No significant difference in lymphocyte, monocyte or macrophage ferric reductase activity was observed when GH and control preparations were compared. However, a significant ($p<0.05$) ten-fold increase in ferric reductase activity accompanied the differentiation of monocytes to macrophages. This increase most likely reflects the co-ordinate upregulation of proteins of iron metabolism during the differentiation of monocyte to macrophage.
Contents

Chapter 1. General introduction: Index 12

1.1. THE HISTORY OF GENETIC HAEMOCHROMATOSIS 13
1.2. CLINICAL FEATURES 14
1.2.1. Diagnosis 15
1.2.1.1. Liver biopsy 16
1.2.1.3. CT and MRI scanning 18
1.2.2. Symptoms 18
1.2.2.1. Cirrhosis 18
1.2.2.2. Skin abnormalities 19
1.2.2.3. Arthropathy 19
1.2.2.4. Diabetes Mellitus 20
1.2.2.5. Other Endocrine Abnormalities 21
1.2.2.6. Cardiac Abnormalities 21
1.2.3. Treatment 22

1.3. GENETICS OF HAEMOCHROMATOSIS 23
1.3.1. Mode of inheritance 23
1.3.2. Human Leukocyte Antigen (HLA) alleles 23
1.3.3. Linkage disequilibrium with telomeric markers 24
1.3.4. Recombinant chromosomes 26
1.3.5. Chromosome rearrangements 26
1.3.6. Candidate Genes 27

1.4. POSITIONAL CLONING OF THE HFE GENE REGION 30

1.5. ISOLATION OF THE HFE GENE 30
1.5.1. The HFE protein and possible function 32
1.5.2. Genetic heterogeneity of haemochromatosis 37

1.6. MOLECULAR ADVANCES IN THE FIELD OF IRON METABOLISM 40
1.6.1. Ferric reductase 40
1.6.2. Divalent Metal Transporter 1 (DMT1 / DCT1 / Nramp2) 40
1.6.4. Hephaestin 44
1.6.5. Hepcidin 44

1.7 PREVIEW OF THESIS 45

Chapter 2. Analysis of GH and Control genotypes: Index. 49

2.1. AIMS OF THE STUDY 50
2.2. INTRODUCTION 51
2.2.1. Polymorphic Markers 51
2.2.1.1. Protein polymorphisms 51
2.2.1.2. Restriction fragment length polymorphisms 51
2.2.1.3. Polymorphism information content 52
2.2.1.4. Minisatellite repeats (VNTRs) 53
2.2.1.5. Microsatellites 54
2.3. METHODS 56
2.3.1. Multiplex fluorescent PCR 56
2.3.1.1. Microsatellite analysis 57
2.3.1.2. Sample preparation 57
2.3.1.3. Setup of the 310 ABI prism autosampler 57
2.3.1.4. Run conditions 58
2.3.1.5. Computer analysis of data 59
2.3.2. HFE mutation analysis 59
2.4. RESULTS 59
2.5. DISCUSSION 63
2.5.1. General problems of microsatellite analysis 63
2.5.2. GH patient microsatellite data versus controls 64

Chapter 3. Isolation and characterisation of genes around D6S1260: Index. 67

3.1. INTRODUCTION 68
3.2. METHODS 69
3.2.1 YAC and PAC clones 69
3.2.2. Plating the cDNA library 71
3.2.2.1. The Clontech human small intestine cDNA library HL1133a 71
3.2.2.2. Titering the cDNA library 72
3.2.2.3. Plating a 1X coverage of the library 73
3.2.2.4. Preparation of duplicate library filters for screening 73
3.2.3. Screening cDNA library 74
3.2.3.1. PAC insert probe preparation 74
3.2.3.2. Pre-hybridisation of filters with placental DNA competition 75
3.2.3.3. Competition of the radiolabelled probe 76
3.2.3.4. Washing of cDNA filters 77
3.2.3.5. Analysis of results from the 715 / 672 PAC insert probe 77
3.2.3.6. Secondary Screen 78
3.2.3.7. Tertiary screen 78
3.2.4. Subcloning the cDNA clones 79
3.2.4.1. PCR amplification of positive plaques 79
3.2.4.2. EcoR I subclones of cDNA PCR products 79
3.2.4.3. Cloning of PCR products using CloneAmp pAMP1 vector 80
3.2.4.4. Amplification of cDNAs for pAMP1 cloning 80
3.2.4.5. Ligation of cDNAs into pAMP1 vector 81
3.2.5. Sequencing of the cDNAs 82
3.2.6. Computer analysis of cDNA sequences 83
3.2.7. Hybridisation analysis of cDNA inserts 84
3.3. RESULTS 84
3.4. DISCUSSION 89
Chapter 5. Ferric reductase functional studies: Index. 149

5.1. INTRODUCTION 150
5.2. PATIENTS AND METHODS 153
5.2.1. Patients 153
5.2.2. Mutation analysis 153
5.2.3. Isolation of peripheral blood monocytes and differentiation to macrophages 154
5.2.4. Ferric reductase assay 158
5.3. RESULTS 159
5.4. DISCUSSION 164

Chapter 6. Conclusions and Perspectives. 166

References. 180

Appendix 1. 220

Appendix 2. 222
Index of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Chromosomal location of known genes of iron metabolism prior to the isolation of HFE</td>
<td>28</td>
</tr>
<tr>
<td>1.2</td>
<td>Genetic Heterogeneity of Haemochromatosis.</td>
<td>39</td>
</tr>
<tr>
<td>2.1</td>
<td>Patient founder haplotype microsatellite and HFE mutation analysis results</td>
<td>61</td>
</tr>
<tr>
<td>3.1</td>
<td>End sequence and hybridisation analysis of eleven cDNA clones isolated by hybridisation of PACs 715 and 672 to a small intestinal cDNA library.</td>
<td>87</td>
</tr>
<tr>
<td>3.2</td>
<td>Low resolution YAC / PAC map of the haemochromatosis gene region showing the position of ZNF204.</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Multiple mutations in the zinc finger and linker motifs of ZNF204.</td>
<td>104</td>
</tr>
<tr>
<td>4.2</td>
<td>Predicted binding sequence of ZNF204 of the 10 putative functional</td>
<td>106</td>
</tr>
<tr>
<td>5.1</td>
<td>Human ferric reductase activity has the characteristics of a membrane bound enzyme.</td>
<td>162</td>
</tr>
<tr>
<td>5.2</td>
<td>Ferric reductase activities.</td>
<td>163</td>
</tr>
</tbody>
</table>
Index of Figures

Figure 1.1. A predicted model of the HFE protein based on sequence homology with classical MHC class I proteins. 34
Figure 1.2. Diagram of iron transport from intestine (enterocyte) to plasma to liver hepatocyte 42
Figure 2.1. GeneScan trace showing relative fluorescent intensity plotted against approximate size (bp). 60
Figure 4.1. Possible structures of the two histidine two cysteine zinc finger. 95
Figure 4.2. Mapping, isolation and characterization of ZNF204. 92
Figure 4.3. ZNF204 sequence, as deposited in GenBank (Accession No. AF033199) 101
Figure 4.4. Northern analysis of the ZNF204 locus expression on a polyA(+) multi-tissue northern blot. 114
Figure 4.5. Diagrammatic overview of 5' RACE procedure. 119
Figure 4.6. Fluorescent in situ hybridisation of PAC 715 on normal metaphase chromosomes, showing hybridisation to chromosome 6p21.3. 132
Figure 4.7. Fluorescent in situ hybridisation of cDNA 3AS2 on normal metaphase chromosomes, showing hybridization to chromosome 6p21.3. 133
Figure 4.8. Fluorescent in situ hybridisation of PAC 832 on normal metaphase chromosomes, showing hybridisation to chromosome 14. 134
Figure 4.9. ZNF204 cross hybridizes with multiple zinc finger loci. 138
Figure 4.10. Results of the TNT analysis of ZNF204. 143
Figure 6.1 A model showing the proposed role of Hepcidin in intestinal iron absorption. 175
Acknowledgments

I would like to thank the following people for their kind help and support:

Firstly I would like give my deepest thanks to my supervisor: Dr. Ann Walker for her unfailing support throughout all aspects of this thesis and wise advice through thick and thin.

I would also like to thank my secondary supervisors; Firstly Dr. James Dooley for his support guidance and copious quantities of blood, and secondly, Dr. J Paul Simons for his kind help and reviews of my progress.

Special thanks to Daniel Wallace for help and support throughout my time in the lab and to all the members of the UK Haemochromatosis Consortium for their wise words of advice.

I would especially like to thanks the members of my family for putting up with me writing this thesis; my mother and father Brenda and John, my sister Gemma and my wonderful wife Melanie who makes it all worthwhile.

This work was supported by the Medical Research Council.
Dedication

I wish to dedicate my thesis to my daughter Emily Rebecca Partridge who decided to be born 5 and a half weeks early on the same day I was due to take my PhD viva (9th of July 2003). Both me and mummy love you very much.
Chapter 1

General Introduction
Chapter 1. General introduction: Index.

1.1. THE HISTORY OF GENETIC HAEMOCROMATOSIS .......... 13
1.2. CLINICAL FEATURES .................................................... 14
  1.2.1. Diagnosis ................................................................. 15
  1.2.1.1. Liver biopsy ............................................................ 16
  1.2.1.3. CT and MRI scanning ............................................ 18
  1.2.2. Symptoms ............................................................... 18
   1.2.2.1. Cirrhosis ............................................................... 18
  1.2.2.2. Skin abnormalities ................................................. 19
  1.2.2.3. Arthropathy ............................................................ 19
  1.2.2.4. Diabetes Mellitus .................................................. 20
  1.2.2.5. Other Endocrine Abnormalities ............................... 21
  1.2.2.6. Cardiac Abnormalities ........................................... 21
  1.2.3. Treatment ............................................................... 22
1.3. GENETICS OF HAEMOCROMATOSIS ............................. 23
  1.3.1. Mode of inheritance .................................................. 23
  1.3.2. Human Leukocyte Antigen (HLA) alleles ...................... 23
  1.3.3. Linkage disequilibrium with telomeric markers .............. 24
  1.3.4. Recombinant chromosomes ........................................ 26
  1.3.5. Chromosome rearrangements ..................................... 26
  1.3.6. Candidate Genes ..................................................... 27
1.4. POSITIONAL CLONING OF THE HFE GENE REGION ............. 30
1.5. ISOLATION OF THE HFE GENE ...................................... 30
  1.5.1. The HFE protein and possible function ......................... 32
  1.5.2. Genetic heterogeneity of haemochromatosis ................... 37
1.6. MOLECULAR ADVANCES IN THE FIELD OF IRON METABOLISM .... 40
  1.6.1. Ferric reductase ...................................................... 40
  1.6.2. Divalent Metal Transporter 1 (DMT1 / DCT1 / Nramp2) ....... 40
  1.6.4. Hephaestin ............................................................. 44
  1.6.5. Hepcidin ............................................................... 44
1.7 PREVIEW OF THESIS .................................................... 45
Chapter 1. General introduction:

1.1. The history of genetic haemochromatosis

Genetic haemochromatosis (GH) was first described by Trousseau in 1865 and later Troisier in 1871 in patients with the classical symptoms of diabetes mellitus, cirrhosis and bronzing of the skin. Von Recklinghausen (1889) was the first to attribute these features to a specific disorder. Sheldon (1935) was the first to propose that GH was caused by an inborn error of metabolism that led to increased iron accumulation. GH was one of the first genetic disorders to be mapped to a specific chromosome via linkage. This seminal work by Simon and co-workers in 1975 showed an association between GH and the human leukocyte antigen (HLA)-A3. Linkage was shown with HLA-A, which was later mapped as part of the major histocompatibility complex (MHC) to chromosome 6p21.3 (Simon et al., 1987; Franke and Pellerino, 1977). The approximate chromosomal position of GH was known for 19 years before the \textit{HFE} gene was finally cloned.
1.2. Clinical features

The clinical presentation of GH is dependent on both genetic and environmental factors. The major genetic component results in increased iron absorption but this will be affected by environmental factors including the dietary iron intake and blood loss. Patients present between the third and fifth decade of life with a variety of symptoms including arthralgia, lethargy, upper abdominal pain, abnormal liver function tests, loss of libido or the onset of diabetes mellitus. Late features are skin pigmentation, liver cirrhosis, endocrine abnormalities and cardiac disease. There is often a delay between the first presenting symptoms and diagnosis of GH of approximately 5 years (Adams et al., 1991).

GH is 10 times more common in male than female patients because the onset is delayed in women due to menstrual blood loss and pregnancy (Finch et al., 1966). Precirrhotic GH patients have a normal life expectancy and cirrhotic patients may only have a normal life expectancy if iron is removed (Adams et al., 1991). Those with cirrhosis and diabetes mellitus have reduced survival, due to the development of complications including hepatocellular carcinoma (Niederau et al., 1985). Since cirrhosis is prevented by the removal of iron by venesection there is a need for early diagnosis.
1.2.1. Diagnosis

The initial diagnosis of GH is usually based on the phenotypic expression of the disease, assessed by serum iron (normal range 11-36 µmol/l), total iron binding capacity (normal range 53-85 µmol/l), serum ferritin concentration and liver biopsy (normal ranges dependent on age/sex). From the serum iron concentration and total iron binding capacity (TIBC) the transferrin saturation value can be calculated (serum iron/TIBC). The normal serum iron concentration is between 11 and 36 µmol/l and rises as body iron stores increase.

Patients with iron overload and increased serum iron concentrations have a reduced serum transferrin level, due to a feedback mechanism reducing transferrin synthesis.

Serum ferritin levels generally reflect body iron stores and are increased in GH. Ferritin is an iron storage protein, which is present in all tissues. The normal serum ferritin concentration ranges for the Royal Free Hospital laboratory are 10-150 µg/l in females and 20-200 µg/l in males. Levels are increased in patients with GH; with advanced disease levels can be increased to approximately 10 times the normal level.
1.2.1.1. Liver biopsy

Liver biopsy is important in the definite diagnosis of GH because it allows both histological examination and estimation of liver iron. As iron accumulation occurs in GH it is deposited as both ferritin and haemosiderin in periportal hepatocytes and in the pericanalicular cells within lysosomes. Later in the disease deposition is found in all areas of the hepatic acinus as well as the bile duct epithelium and connective tissue. Untreated this iron causes perilobular fibrosis and in advanced disease broad fibrous septa and cirrhosis. Areas of preserved parenchyma surrounded by fibrotic septa lead to a “holly leaf” pattern (Powell et al., 1981).

The pattern of iron distribution in GH is different to that found in secondary iron overload, for example following transfusion for haemolytic anemia. Kupffer cells (differentiated liver macrophages) do not become iron loaded until late in classical GH. This is in contrast to the iron loaded Kupffer cells observed in transfusion related siderosis (Yam et al., 1968).

The liver tissue obtained from a needle biopsy is split into two pieces:

(a) One section is fixed in formalin for histological examination to determine the degree of fibrosis and cirrhosis. Haemosiderin iron is
estimated histologically using Perls (Prussian) stain or Tirmann-Schmelzer stain.

(b) The remainder is washed to remove any contaminating blood and then placed in a drying oven to give the dry weight (approximately 30% of the hepatic wet weight). The iron content is then determined by either spectrophotometric methods or atomic absorption spectrophotometry.

The iron content is used to determine the hepatic iron index:

\[
\frac{\text{Liver iron concentration of iron (µmol/g dry wt)}}{\text{Age of patient in years}} = \text{Hepatic iron index}
\]

A hepatic iron index of above 2.0 is indicative of GH. The index has been valuable in differentiating between GH homozygotes (generally > 2.0) and heterozygotes and patients with abnormal iron indices caused by other factors such as alcoholic liver damage and hepatitis C (generally < 2.0) (Basset et al., 1986, Summers et al., 1990).
1.2.1.3. CT and MRI scanning

The computed axial tomography (CT) is a non-invasive method, which may show hepatic iron overload. However its main drawback, apart from cost, is low sensitivity for mild and moderate iron deposition. Magnetic resonance imaging (MRI) can detect moderate iron deposition but is not sensitive enough for detection of early iron deposition. Liver iron Measurements by Biosusceptometry (SQID) is a far more sensitive method however, the only two available SQID machines are in North America. For this reason, liver biopsy with iron measurement may be used to assess liver iron in the diagnosis of GH.

1.2.2. Symptoms

1.2.2.1. Cirrhosis

Approximately 10 % of GH patients present with cirrhosis (Stremmel et al., 1995, Adams et al., 1991). Hepatocellular carcinoma is a major complication and cause of mortality with prevalence more than 200 times that of the general population (Powell et al., 1994). In cirrhotic patients hepatocellular carcinoma is still a risk despite the removal of the iron. Since the advent of genetic testing, the goal is early identification and treatment of haemochromatosis before establishment of cirrhosis.
1.2.2.2. Skin abnormalities

Untreated GH patients with late disease can have excessive skin pigmentation. This is caused by increased melanin deposition in the skin.

GH has been called "bronze diabetes" but "bronze" is misleading as most GH patients display a slate grey discolouration of the skin. The abnormal melanin deposits are increased within scars, genitalia, flexation creases and on rare occasions the oral mucosa (Ferrand et al., 1977). Another cutaneous feature of late GH is lack of facial and pubic hair. This may be secondary to endocrine abnormalities.

1.2.2.3. Arthropathy

Arthropathy was first described in patients with GH in 1964 by Schumacher and co-workers. It is now recognized that 25 to 50% of all GH patients suffer from joint pain with or without signs of arthritis (Niederau et al., 1985). The prevalence of arthropathy is not related to the degree or duration of the iron overload. It can occur despite venesection and can even occur after iron stores have returned to a normal level. Arthropathy is often present early in the disease process preceding diabetes mellitus and cirrhosis (Adams 1991). It is not specific for GH and is observed also in secondary iron overload (Dymock et al., 1972).
The most common form of arthropathy observed normally progresses slowly, although degenerative arthritis is observed. The arthropathy involves the joints of the hands, proximal interphalangeal joints and metacarpophalangeal joints (Schumacher 1988). Early radiological changes consist of subchondrial cyst formation and narrowing of joint spaces. Knees and hips may also be affected. Chondrocalcinosis occurs in these joints and is related to the deposition of calcium pyrophosphate, which is thought to be a direct result of iron deposition. Iron inhibits pyrophosphatase resulting in the accumulation of pyrophosphates.

1.2.2.4. Diabetes Mellitus

Diabetes mellitus develops in 30 to 60 % of patients with GH (Niederau et al., 1985). An investigation of 163 GH patients showed 55 % had diabetes mellitus with another 10 % having glucose intolerance (Niederau et al., 1985). Insulin dependent diabetes mellitus is associated with late disease. Complications of diabetes mellitus such as retinopathy, peripheral neuropathy and nephropathy may also occur. Diabetes is thought to be due to damage to the islet cells (beta cells) of the pancreas by excess deposition of iron (Stremmel et al., 1988).
1.2.2.5. Other Endocrine Abnormalities

Loss of libido and testicular atrophy are frequently observed in advanced GH in men, and may be the first sign of the disorder. The cause of hypogonadism is pituitary dysfunction due to the selective deposition of iron within the gonadotrophic cells of the pituitary gland. There is reduced secretion of lutenizing hormone and follicle stimulating hormone (Stremmel et al., 1988), with a subsequent reduction of plasma testosterone levels (Kley et al., 1985). In young females amenorrhea may result (Adams et al., 1990).

1.2.2.6. Cardiac Abnormalities

GH patients may present rarely with cardiomyopathy and congestive cardiac failure. Cardiomyopathy is normally observed in long-standing GH patients diagnosed late in the disease but may be the presenting feature in younger adults (Dabestani et al., 1988). The risk of death due to cardiomyopathy is approximately 300 times higher for GH patients compared with the general population. Although 20 - 35 % of GH patients have arrhythmias, such as supraventricular ectopic beats, tachyarrhythmias and atrial flutter and fibrillation, these are normally not life threatening (Adams et al., 1991).
1.2.3. Treatment

GH is treated by venesection of 1 unit of blood (approximately 450ml) weekly until the transferrin saturation and ferritin concentration are normal (Crawford et al., 1991, Bourel and Lenoir 1965). Maintenance venesection is then continued at 3 to 6 monthly intervals. Each unit of blood removed represents approximately 250mg of iron. The range of iron overload in GH patients is 5 to 40g (Bomford and Williams 1976) equivalent to 20 to 160 venesections. For comparison the normal body iron content is approximately 5g. Venesection increases the rate of erythropoiesis thus mobilizing stored iron from parenchymal tissues.

In the rare patient with cardiomyopathy the removal of iron by venesection alone may be too slow, and the iron chelator desferrioxamine is used in combination with venesection. Desferrioxamine however causes side effects, so treatment must be kept to a minimum and used only in extreme cases.
1.3. Genetics of haemochromatosis

1.3.1. Mode of inheritance

In 1974 Saddi and co-workers proposed an autosomal recessive mode of inheritance for GH, which is now accepted. Before this it was thought that the inheritance of GH was autosomal dominant. The autosomal dominant mode of inheritance was proposed due the high proportion of children with an affected parent having GH. However, this was an effect of the high gene frequency allowing for a high proportion of homozygous / heterozygous marriages.

1.3.2. Human Leukocyte Antigen (HLA) alleles

In 1975 Simon and coworkers reported an association of GH with certain alleles of HLA antigens, in particular the HLA-A3 and B-14 antigens. These preliminary results were published in French literature in early 1975 (Simon et al., 1975). Full details were later published in English in 1976 (Simon et al., 1976). HLA-A3 and HLA-B14 were found in significantly greater frequencies in GH patients than healthy blood donor controls. HLA-A3 was observed in 78% of GH patients versus 27% of controls, and HLA-B14 in 26% of GH patients compared with only 3% of controls (Simon et al., 1976). HLA-B7 was also shown to be significantly
increased in GH patients in a later study (Simon et al., 1981). Family studies also demonstrated linkage of GH to HLA-A (Simon, 1987). The findings of Simon were a landmark in the history of haemochromatosis research, and were the first evidence to confirm Sheldon's hypothesis, made in 1935, that GH was dependent on genetic factors.

Research in haemochromatosis prior to this date hung on the question whether haemochromatosis was acquired or inherited, with some authors arguing strongly for GH being an acquired disorder.

Simon's observations did not rule out environmental factors influencing the extent of iron overload, but did show that the primary defective locus was linked to the major histocompatibility complex (MHC), HLA antigen A, which was subsequently mapped to chromosome 6p21.3 (Franke and Pellegrino et al., 1977)

1.3.3. Linkage disequilibrium with telomeric markers

The GH candidate region initially centered on the HLA region but was extended greatly with the investigation of new microsatellite markers. One crucial CA repeat marker D6S105 was isolated and mapped approximately 2cM telomeric of HLA-A. (Stone et al., 1994; Raha-Chowdury et al., 1995). Allele 8 of this marker has a stronger association with GH than the HLA-A3 allele (Jaswinska et al., 1993): D6S105 allele 8
was observed in 93% of GH patients compared to 21% of controls. Additional CA repeat markers around D6S105 and towards the telomeric marker D6S299 were also investigated. This study revealed that 70% of GH chromosomes contain the haplotype: D6S306 allele 5, D6S105 allele 8, D6S464 allele 9 and D6S1260 allele 4. This common GH haplotype was only found in 6% of non-GH control chromosomes (Raha-Chowdhury et al., 1995). The linkage of GH to HLA-A and the presence of a common haplotype in the majority of patients suggested that GH has spread by multiplication of a founder mutation, and that the majority of GH chromosomes are derived from this founder individual.

The gene position in the founder haplotype could therefore be mapped by linkage disequilibrium. Linkage disequilibrium is defined as the nonrandom association of alleles at linked loci.

The lambda likelihood method of linkage disequilibrium analysis is powerful as it analyses all recombinations since the original mutation and not just recombination occurring in a single family (Hasbacka et al., 1994). Using this method of linkage disequilibrium analysis in the GH critical region Raha-Chowdhury et al. (1995) showed that the marker D6S1260 was at the peak of linkage disequilibrium with GH in their data set. The peak indicated the most probable region for the gene to be located at this time. The GH gene was thought most likely to lie in the region between D6S105 and D6S1260 (equates to approximately 900
kb). This mapping was reproduced by an independent American study (Seese et al., 1996) using a similar set of markers.

1.3.4. Recombinant chromosomes

Several recombinations involving GH and HLA-A and HLA-B have been proposed but subsequently withdrawn (Edwards et al., 1986, David et al., 1986, Panajotopoulos et al., 1989; Powell et al., 1990). However an informative recombination in a GH patient was described by Calandro and co-workers (1995). This positioned the HFE gene telomeric to HLA-F. This recombinant was therefore consistent with the linkage disequilibrium data (Calandro et al., 1995).

1.3.5. Chromosome rearrangements

An initial pulsed field gel electrophoresis (PFGE) study by Lord and co-workers (1990) showed no GH associated changes. In this study however the most telomeric marker used was HLA-A. More recently, a pericentromeric inversion (6p21.1-6p23) was demonstrated to segregate with disease in a family. However, the breakpoints appeared to flank the predicted site of the HFE gene (Venditti et al., 1994). Recent studies using PFGE in the region around D6S1260 however showed no rearrangement (Wallace et al., 1998). The lack of any observable PFGE rearrangements in GH suggested that the common founder mutation
responsible for GH would not involve large chromosomal rearrangement. This implied that the GH defect was either a point mutation or a micro deletion or rearrangement, beyond the resolving power of techniques such as PFGE.

1.3.6. Candidate Genes

Known proteins of iron metabolism were excluded as candidates on the basis of their chromosomal location (see Table 1.1.). None of the functional genes of iron metabolism mapped to 6p21.3.
Table 1.1. Chromosomal location of known genes of iron metabolism prior to the isolation of HFE.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gene Symbol</th>
<th>Map Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferritin (H)</td>
<td>FTH1</td>
<td>11q3</td>
</tr>
<tr>
<td></td>
<td>FTHL 1-4</td>
<td>1p31-p22; 1q32.2-q42; 2q32-q33; 3q21-q23</td>
</tr>
<tr>
<td></td>
<td>FTHL7,8, 10-16</td>
<td>13q12; Xq26-q28</td>
</tr>
<tr>
<td></td>
<td>FTHP1</td>
<td>6p21.3-p12</td>
</tr>
<tr>
<td>Ferritin (L)</td>
<td>FTL</td>
<td>19q13.3-13.4</td>
</tr>
<tr>
<td></td>
<td>FTLL1</td>
<td>20q12-pter</td>
</tr>
<tr>
<td></td>
<td>FTLL2</td>
<td>Xp22.3-p21.2</td>
</tr>
<tr>
<td>Transferrin</td>
<td>TF</td>
<td>3q21</td>
</tr>
<tr>
<td></td>
<td>TFP</td>
<td>3</td>
</tr>
<tr>
<td>Transferrin receptor</td>
<td>TFRC</td>
<td>3q26.2-pter</td>
</tr>
<tr>
<td>Lactoferrin</td>
<td>LTF</td>
<td>3q21</td>
</tr>
<tr>
<td>Iron regulatory binding protein</td>
<td>IRP1</td>
<td>9</td>
</tr>
<tr>
<td>Aconitase (soluble)</td>
<td>ACO1</td>
<td>9q22-q32</td>
</tr>
<tr>
<td>Haptoglobin</td>
<td>HP</td>
<td>16q22.1</td>
</tr>
<tr>
<td>Haemopexin</td>
<td>HPX</td>
<td>11q15.5-15.4</td>
</tr>
<tr>
<td>Haem Oxygenase –1</td>
<td>HMOX1</td>
<td>22q12</td>
</tr>
<tr>
<td>Haem Oxygenase –2</td>
<td>HMOX2</td>
<td>16p13.3</td>
</tr>
<tr>
<td>Caeruloplasmin</td>
<td>CP</td>
<td>3q21-24</td>
</tr>
<tr>
<td>Iron Response Protein 2</td>
<td>IRP2</td>
<td>15</td>
</tr>
</tbody>
</table>

* Adapted from Worwood 1999
Pappas and co-workers (1995) described 10 expressed sequence tagged sites (ESTs) mapping around the HFE gene region (Pappas et al., 1995). ESTs are sequences defined by a pair of primers derived from cDNA. Pappas and co-workers mapped 10 ESTs to chromosome 6p21-6p23 using a panel of somatic cell hybrids containing different portions of chromosome 6. The fine mapping of these ESTs was not described. The DNA library used to generate the ESTs however was fetal brain, which was not the most likely tissue for the expression of HFE, unless HFE is ubiquitously expressed.

A zinc finger protein described by Beutler and co-workers was proposed as a candidate for the HFE gene, lying 500kb centromeric of D6S105 (Beutler et al., 1996). The gene was shown to be polymorphic in 3 out of 55 GH patients and 1 out of 44 control patients, however it was most highly expressed in ovary, testis and prostate, as well as the small intestine. Although zinc finger proteins, as transcriptional regulators, were considered potential candidates for the HFE gene, the expression pattern of the gene and its position made it very unlikely to be the gene causing GH.
1.4. Positional cloning of the \textit{HFE} gene region

Positional cloning is the isolation of a gene by chromosomal map position alone, without the use of antibodies to the protein or prior information on function, structure or sequence. For positional cloning, the candidate region must be refined down to a size that allows efficient use of gene isolation techniques. In GH no recombinations or chromosome rearrangements became apparent that would allow the gene position to be narrowly defined (see section 1.3.4.).

1.5. Isolation of the \textit{HFE} gene

Mercator Genetics Incorporated used a combined positional cloning approach to isolate the \textit{HFE} candidate gene (Feder et al., 1996). They originally and incorrectly named the gene \textit{HLA-H} (there was already an HLA-H locus, OMIM 142925). They first generated a Yeast Artificial Chromosome (YAC) contig stretching from HLA-A to D6S276 and used this to isolate smaller Bacterial Artificial Chromosome (BAC), P1 Phage Artificial Chromosome (PAC) and cosmid clones. Using these clones as a resource they generated 26 short repeat markers across the region, which they used for linkage disequilibrium studies (Feder et al., 1996) The highest $P_{excess}$ obtained (D6S2241) was 0.81, suggesting that at least 81% of disease-bearing chromosomes carried a common mutation. They used this $P_{excess}$ data, with the deviation from Hardy-Weinberg equilibrium
along with ancestral haplotype analysis to narrow the critical region containing the gene to a region of 600kb. Where family data was unobtainable a cell line was created containing each individual chromosome 6, to enable haplotype analysis. The precise boundaries for the gene were identified by haplotype analysis of GH chromosomes not bearing the full extended GH ancestral haplotype, to identify the positions of historic recombination events. This analysis defined the minimal $HFE$ gene region of 250kb between the markers D6S2241 and D6S2238.

Exon trapping and cDNA selection were then used on clones spanning the critical region to identify candidate sequences. The entire genomic region of 250,000bp was also sequenced to ensure complete identification of all possible coding regions. The 250kb region, like the neighboring MHC region, was gene rich, yielding 15 genes: 12 histone genes, a Ro/SSA ribonucleotide protein (an auto antigen that is frequently recognised in patients with Systemic Lupus Erythematosus and Sjogren Syndrome), a sodium phosphate transporter and a HLA-A2 like sequence. To identify the most probable candidate, each gene was individually sequenced, including intron-exon boundaries, in two haemochromatosis patients homozygous for the founder haplotype and 2 control individuals. A single base mutation of a G to A at nucleotide 845 of the open reading frame was found to be homozygous in the HLA-A2 like gene in the patients but not present in the controls. This mutation predicted the substitution of a cysteine at amino acid 282 by tyrosine.
(Cys282Tyr). Sequence homology with other HLA molecules predicted that this mutation would disrupt an intrastrand disulphide bond. This mutation was detected in 85% of GH chromosomes and only 3% of control chromosomes. A second variant H63D (His63Asp) was also detected (Feder et al., 1996)

1.5.1. The HFE protein and possible function

Based upon the sequence similarities, the product of the HFE gene was predicted to have similarities with atypical class I MHC proteins (Figure 1.1). HFE is a MHC class I-like molecule that interacts with β2-microglobulin. HFE was predicted to have with three extracellular domains (α1, α2 and α3 with the binding site for β2-microglobin on α3, a transmembrane helix and a short intracellular domain. HFE retains the structural features of MHC class I molecules such as the four cysteine residues that form disulphide bridges within the α2 and α3 domains, but unlike classical antigen presenting HLA molecules only two of the normally four critically spaced tyrosine residues that are important in peptide binding are present (Madden et al., 1992). This two tyrosine structure will destroy the peptide binding groove as is observed with nonclassical members of the MHC Class I family, such as the Fc receptor that do not bind peptides (Ravetch and Marguiles, 1994; Story et al., 1994). In the Fc receptor the peptide binding groove is effectively closed by a proline in the α2 domain; in HFE there is an equivalent proline at
codon 188 (Feder et al., 1996) It seemed unlikely that HFE is involved in presentation of peptides, but how it could be involved in regulation of iron metabolism was unclear.
Figure 1.1. A predicted model of the HFE protein based on sequence homology with classical MHC class I proteins.

The model shows the three extracellular domains $\alpha_1$, $\alpha_2$, $\alpha_3$, the binding site for $\beta_2$-microglobulin on $\alpha_3$, the transmembrane helix and intracellular domain, as well as the sites of the C282Y and H63D mutations (adapted from Feder et al., 1996; Bacon et al., 1999).
The Cys282Tyr mutation disrupts a critical disulphide bridge in the α3 loop of the HFE protein (Figure 1.0.) This affects the protein’s ability to interact with β2-microglobulin, retarding its presentation on the surface of the cell and binding to the transferrin receptor protein (TfR). The His63Asp mutation does not affect the association of HFE with β2-microglobulin, and *in vitro* this form is presented on the cell surface (Waheed *et al.*, 1997; Feder *et al.*, 1997).

Northern blot analysis, showed a single major transcript present in most tissues of about 4kb. Highest levels were observed in small intestine, the site of iron absorption and the liver, the major site of iron deposition in haemochromatosis (Feder *et al.*, 1996).

Immunohistochemical studies showed staining for the HFE protein in some epithelial cells in every segment of the alimentary canal. The most intense staining was seen in the deep crypts of the small intestine (Parkkila *et al.*, 1997). The crypt cells do not absorb iron; this occurs as they migrate up the villus before they are sloughed off. This was an unexpected finding as it was assumed the staining would be more intense on the mature villus enterocytes, the site of iron absorption.
A major breakthrough was the demonstration of binding of HFE to the transferrin receptor (Waheed et al., 1999, Parkkila et al., 1997b and Feder et al., 1999). Although the mechanism of action of HFE is not fully understood, it appears to modulate the classical transferrin receptor - mediated endocytosis pathway of iron entry into cells.

In the small intestine, the transferrin receptor is expressed on the basolateral surface of the enterocyte. It remains controversial, but it may be that the HFE - transferrin receptor complex in the basolateral crypts may "sense" body iron stores, and determine the iron absorption of the mature enterocyte.
1.5.2. Genetic heterogeneity of haemochromatosis

Since the discovery of the \( HFE \) gene, several other causes of haemochromatosis have been found (Table 1.2.).

Juvenile haemochromatosis is a rare autosomal recessive disorder causing accelerated iron overload presenting in the second and third decades of life and affects men and women equally. The gene has been mapped by linkage analysis to chromosome 1q, but is as yet undiscovered (Roetto et al. 1999)

Transferrin receptor 2-related haemochromatosis is an autosomal recessive disorder caused by mutations in the gene \( TFR2 \), which encodes a transferrin receptor isoform (Camaschella et al., 2000, Roetto et al., 2002)

Ferroportin 1 related iron overload is an autosomal dominant disorder caused by mutation of the gene for the ferroportin iron transporter molecule (also known as iron-regulated transcript 1, IREG1; metal transporter protein 1, MTP1, SLC11A3). It causes iron overload, though markedly in cells of the reticulo endothelial system (e.g. Kupffer cells). It is difficult to manage clinically using venesection as this may cause anaemia (Njajou et al., 2001; Montosi et al., 2001; Devalia et al., 2002; Wallace et al., 2002)
H-Ferritin related iron overload is an autosomal dominant iron overload that has (so far) only been observed in a single Japanese family. It is caused by a mutation in 5' iron responsive element (IRE) in the gene that encodes the H-subunit of the iron storage molecule ferritin (Kato et al., 2001).
Table 1.2. Genetic Heterogeneity of Haemochromatosis.

<table>
<thead>
<tr>
<th>Disease (onset)</th>
<th>OMIM*</th>
<th>Gene (year identified)</th>
<th>Locus</th>
<th>Mode of inheritance</th>
<th>Mutation(s)</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFE-related haemochromatosis (adult)</td>
<td>235300</td>
<td>HFE (1996)</td>
<td>6p31.3</td>
<td>Autosomal recessive</td>
<td>C282Y, H63D, S65C, IVS3+1G→T</td>
<td>Venesection</td>
</tr>
<tr>
<td>Juvenile haemochromatosis (juvenile / early adult)</td>
<td>602390</td>
<td>Not yet cloned</td>
<td>1q21</td>
<td>Autosomal recessive</td>
<td>-</td>
<td>Chelation e.g. desferrioxamine (Venesection – care)</td>
</tr>
<tr>
<td>Transferrin receptor 2-related haemochromatosis (adult)</td>
<td>604720, 604250</td>
<td>TFR2** (2000)</td>
<td>7q22</td>
<td>Autosomal recessive</td>
<td>Y250X, M172K, 84-88insC</td>
<td>Venesection (disease appears to be similar to HFE-related haemochromatosis)</td>
</tr>
<tr>
<td>Ferroportin 1 – related Iron overload (Adult)</td>
<td>604653, 606069</td>
<td>FPN1 (also called IREG1 or SLC11A3 or MTP 1) (2000)</td>
<td>2q32</td>
<td>Autosomal dominant</td>
<td>N144H, A77D, V162del</td>
<td>Not yet defined. Some patients become anaemic if venesected.</td>
</tr>
<tr>
<td>H-ferritin IRE*** – related Iron overload (Adult)</td>
<td>134770</td>
<td>H-ferritin (mutation identified 2001)</td>
<td>11q12-q13</td>
<td>Autosomal dominant</td>
<td>A49U</td>
<td>Not yet defined</td>
</tr>
</tbody>
</table>

1.6. Molecular advances in the field of iron metabolism

1.6.1. Ferric reductase

The ferric reductase gene was cloned by a subtractive cloning strategy designed to identify intestinal genes involved in iron absorption. The cDNA sequence predicted a plasma membrane di-haem protein, named duodenal cytochrome B (DcytB). Immunohistochemical staining for the protein localised to the duodenal brush border membrane (McKie et al., 2000). Ferric reductase reduces FelII to Fell in preparation for its uptake by Divalent Metal Transporter 1 (DMT1) (Figure 1.2.).

1.6.2. Divalent Metal Transporter 1 (DMT1 / DCT1 / Nramp2)

DMT1 was identified by cDNA expression cloning in Xenopus oocytes (Gunshin et al., 1997). It is a plasma membrane glycoprotein with 12 membrane spanning domains. It has been demonstrated to be a metal ion transporter with a broad substrate range that includes Fell, ZnII, Coll, CdII, Cull, Nill and PbII. DMT1 is found on luminal side of the upper gastrointestinal tract and on the plasma membrane of most cells. It is also present in endosomal vesicles of macrophages and other cells, where it functions to transport iron from within the endosome into the cytoplasm (Flemming et al., 1998; Gruenheid et al., 1999). On the small intestinal brush border, it is the site of iron uptake by the cell after
its reduction by ferric reductase (Figure 1.2.). Expression of DMT1 (possibly ferroportin 1 also) has been reported to be increased (or inappropriately high in relation to body iron stores) in haemochromatosis patients, but as yet the mode of this regulation is not yet fully understood (Zoller et al., 2001)
Figure 1.2. Diagram of iron transport from intestine (enterocyte) to plasma to liver hepatocyte (Fleming and Sly, 2001).
1.6.3. Ferroportin (Ireg1)

Ferroportin was identified using three separate methodologies: subtractive cloning (McKie et al., 2000); positional cloning of the gene responsible for hypochromic anemia in zebrafish (Donovan et al., 2000) and by using an affinity column to bind messenger RNAs containing Iron Response Elements (IREs) (Abboud et al., 2000). It is an iron export protein localised to the enterocyte basolateral membrane (connecting the enterocyte to the portal vein circulation; see Figure 1.2.). Ferroportin (Fp1) is also found on Kupffer cells and on placental syncytiotrophoblast cells (Donovan et al., 2000). An IRE has been reported in the 5' untranslated region (UTR), but its function is not yet understood. Expression of ferroportin is increased in conditions that increase iron absorption, for instance iron deficiency. This is in contrast to the effect of the classical 5'-IRE in H-ferritin, where under low iron conditions, ferritin expression is decreased (Klausner et al., 1993). Ferroportin has been shown to be mutated in some forms of autosomal dominant haemochromatosis (Montosi et al., 2001; Njajou et al., 2001; Devalia et al., 2002). Basolateral transport has been proposed as the rate limiting step of iron absorption and this step is probably mediated by ferroportin 1 (Fp1) (McLaren et al., 1991)
1.6.4. Hephaestin

The Hephaestin gene (HEPH) was cloned as a result of looking for the gene defect in the sex linked anemia (sla) mouse (Vulpe et al., 1999). It was shown to be a caeruloplasmin homologue with predominantly perinuclear localization within intestinal enterocytes. This multicopper protein seems to act as the ferroxidase, necessary for the exit of the iron from the intestinal enterocyte into the systemic circulation (see Figure 1.2. above). However the exact function of hephaestin has yet to be determined.

1.6.5. Hepcidin

Knock out mice for the USF2 gene (upstream of the hepcidin gene) that also prevented expression of the hepcidin gene showed increased circulatory iron and hepatic iron (Nicolas et al., 2001). The iron loading pattern was similar to that displayed by the HFE knockout mouse. Hepcidin has sequence homology with anti microbial peptides and is synthesized in the liver. It is suggested to be part of the signaling between the body iron stores in the liver and iron absorption in the intestine (Fleming and Sly 2001).
1.7 Preview of thesis

Chapter 2 describes the genetic characterisation of the Royal Free cohort of haemochromatosis patients and a control group. The genotype for four microsatellite markers flanking the haemochromatosis gene region was determined. Following the identification of the HFE gene, the patients were also investigated for the Cys282Tyr and His63Asp mutations. This was done to identify a genetically homogeneous group of Cys282Tyr homozygous patients for later ferric reductase studies. It also confirmed absence of disease related genotypes in the control group, and highlighted possible non HFE-related haemochromatosis patients.

Chapter 3 describes the isolation and characterisation of genes from around the microsatellite marker D6S1260. At the time of study this was the apparent peak of linkage disequilibrium within haemochromatosis chromosomes. The study was done in an attempt to identify candidate genes for haemochromatosis. A small intestinal cDNA library was screened with two PAC clones known to contain D6S1260. Eleven cDNA clones were identified via hybridisation, four of which were identified to have single copy inserts that showed sequence homology to zinc finger loci. It was decided to further investigate these four clones.
Chapter 4 describes the complete sequencing and characterisation of the four cDNA clones, which were all derived from a novel zinc finger locus on chromosome 6p21.3, designated ZNF204 (deposited as GenBank accession number AF033199). Genomic and RNA analyses suggested that ZNF204 was an expressed processed pseudogene and is unrelated to haemochromatosis. However, no parent locus could be identified using somatic cell hybrid panels or fluorescent in situ hybridisation (FISH). No product could be identified following coupled transcription and translation of ZNF204. This locus is probably an expressed processed pseudogene. The sequence is however highly conserved across species and may prove valuable in identifying other zinc finger genes in model organisms.

Chapter 5 describes the first analysis of ferric reductase activity in white blood cells from haemochromatosis patients and control individuals. Previous studies had shown increased activities in haemochromatosis compared with control duodenal biopsy ferric reductase activity. Macrophages, like duodenal enterocytes, remain paradoxically iron poor in haemochromatosis and were therefore considered likely to express the same abnormality in ferric reductase activity as duodenum. This study was conducted in an attempt to identify an in vitro index of the disease and to investigate the possibility of functional complementation by candidate haemochromatosis genes. No significant difference was observed in either monocytes or macrophages from
haemochromatotic patients versus controls, ruling out a functional complementation approach. However, a ten fold increase in ferric reductase activity was observed as monocytes matured into macrophages, possibly reflecting the putative upregulation of gene of iron metabolism in differentiation.
Chapter 2

Analysis of GH and control genotypes
Chapter 2. Analysis of GH and Control genotypes: Index.

2.1. AIMS OF THE STUDY ................................................................. 50
2.2. INTRODUCTION ........................................................................ 51
2.2.1. Polymorphic Markers .......................................................... 51
2.2.1.1. Protein polymorphisms ..................................................... 51
2.2.1.2. Restriction fragment length polymorphisms ...................... 51
2.2.1.3. Polymorphism information content ................................... 52
2.2.1.4. Minisatellite repeats (VNTRs) ........................................... 53
2.2.1.5. Microsatellites ................................................................. 54
2.3. METHODS ............................................................................... 56
2.3.1. Multiplex fluorescent PCR .................................................. 56
2.3.1.1. Microsatellite analysis ..................................................... 57
2.3.1.2. Sample preparation ......................................................... 57
2.3.1.3. Setup of the 310 ABI prism autosampler ......................... 57
2.3.1.4. Run conditions ............................................................... 58
2.3.1.5. Computer analysis of data ............................................... 59
2.3.2. HFE mutation analysis ......................................................... 59
2.4. RESULTS ............................................................................... 59
2.5. DISCUSSION ........................................................................... 63
2.5.1. General problems of microsatellite analysis ......................... 63
2.5.2. GH patient microsatellite data versus controls ..................... 64
Chapter 2. Analysis of GH and control genotypes.

2.1. Aims of the study

Four polymorphic (CA)$_n$ repeat markers from the *HFE* gene region were chosen for analysis of GH patient and control genotypes. D6S265, D6S105 and D6S1260 have been shown to form part of the founder haplotype in GH using linkage disequilibrium studies (Jazwinska *et al.*, 1993, Raha-Chowdury *et al.*, 1995, Seese *et al.*, 1996). A marker towards the telomeric extent of the GH gene region, D6S1621, was also tested. Towards the end of this study, the *HFE* gene was identified (Feder *et al.*, 1996), and so the C282Y and H63D mutations of the *HFE* gene were also determined. These studies were performed:

(1) To determine microsatellite markers of the *HFE* gene region, and the C282Y and H63D mutations of *HFE*, in order to identify patients with a secure diagnosis of *HFE*-related haemochromatosis, and to compare GH and control patients to be investigated by ferric reductase assay (Chapter 5).

(2) To screen the Royal Free Hospital cohort of GH patients, and control patients, to identify appropriate individuals for a pulsed field gel electrophoresis approach to mapping the GH gene (Wallace *et al.*, 1996).
2.2. Introduction

2.2.1. Polymorphic Markers

2.2.1.1. Protein polymorphisms

The first genetic linkage studies in humans relied on protein polymorphisms of blood to define linkage (Clark et al., 1956). Linkage is defined as the occurrence of two loci sufficiently close together on a chromosome so that their assortment is non-independent. The first linkage to be shown in GH was with HLA-A (Simon et al., 1977).

2.2.1.2. Restriction fragment length polymorphisms

Due to the limitations of protein analysis investigations, research moved towards recombinant DNA technology to investigate linkage in GH. Restriction fragment length polymorphism (RFLP) occurs because sites recognized by restriction endonucleases are polymorphic. Single base-pair changes are frequent in the human genome (estimated to be as high as 1 in 1000bp. The nucleotide sequence difference can create or destroy enzyme recognition sites. Individual chromosomes can be traced down generations using these polymorphisms. There are however two main drawbacks to using RFLPs for haplotype analysis: (1) the level of polymorphism is normally low and limited by the number of alleles, and
(2) RFLP analysis requires restriction enzyme digestion, electrophoresis, blotting and hybridization. This is time consuming and not well suited to automation.

### 2.2.1.3. Polymorphism information content

The degree of polymorphism of a marker is expressed as the polymorphism information content (PIC). PIC values are calculated using the formula:

\[
PIC = 1 - \left( \sum p_i^2 \right) - \sum \sum 2p_i^2 p_j^2
\]

where \( p_i \) and \( p_j \) are the frequency of the \( i \)th and \( j \)th alleles.

High PIC values indicate a greater likelihood that a locus will be informative for linkage analysis. RFLP analysis on the human genome, due to its large size and low rates of polymorphism, is limited in application. Therefore more informative polymorphic markers were sought.
2.2.1.4. Minisatellite repeats (VNTRs)

Minisatellites are highly polymorphic regions of the genome. They are composed of tandem repeats of 8-90bp with variation deriving from the number of tandem repeats. The length of the total array is variable from 1-30 kb giving rise to a large number of alleles, each of which may be rare in the population. Over 70% of minisatellites are highly polymorphic (Litt et al., 1989). The overall size of the minisatellite is related to its degree of polymorphism.

Small minisatellites (2 kb or smaller) are frequently associated with monomorphic or only minimally variable loci. Large minisatellites (bigger than 4kb) are nearly always derived from hyper-variable loci.

Minisatellites are found throughout the genome, although there is clustering in human centromeric and telomeric regions. Uninterrupted minisatellite repeats are termed perfect repeats; interrupted minisatellites are termed imperfect repeats, and if fragmented into several blocks termed compound repeats (Weber et al., 1989). They are inherited in a Mendelian fashion with a low mutation rate. Their uses include genetic fingerprinting and linkage analysis although this is hampered by the clustering.
2.2.1.5. Microsatellites

Microsatellites consist of around 10-50 copies of a 1-6 bp motif; therefore they can be highly polymorphic. They were originally analyzed and detected by radiolabelled PCR and denaturing polyacrylamide gel electrophoresis (Weber et al., 1989). Microsatellites therefore became a frequently used marker in genetic analysis. The di-nucleotide repeat CA is most commonly used for genetic analysis. The abundance of the CA repeat has been investigated (Tautz et al 1989, Litt et al., 1989, Weber et al., 1989) showing that there are approximately $0.5 \times 10^5$ to $1 \times 10^5$ copies in the human genome with an average separation of between 30-65 kb.

This high frequency throughout the genome and high degree of polymorphism are advantageous for genetic analysis. Examples of the usefulness of microsatellites include the Genethon human genetic linkage map (Gypay et al., 1994) and the Genome Directory (Venter et al., 1995).

The Genethon (1994) map is a second generation genetic linkage map of the human genome containing over 2000 CA repeats, about 56% of these at a distance of less than 1cM from another of its markers. The Genome Directory described the use of CA repeat markers to create a first order YAC contig map across the entire human genome, and to
obtain high resolution second generation maps of chromosomes 3, 12, 6 and 22. These maps are particularly useful in mapping both monogenic diseases and multifactorial traits. The recent publication of the draft human genome sequence built upon these successes, and marked a culmination of the efforts of the Human Genome Project (International Genome Sequencing Consortium, 2001).

The mechanism for the creation of variation in the length of microsatellite repeats is thought to be that of replication slippage (polymerase stutter) randomly occurring in internally repetitive stretches. Another possibility causing variation in repeat length is unequal mitotic exchange. This variation in length however occurs at a very slow pace estimated to be $6.2 \times 10^{-4}$ mutations per locus per gamete for the longest of repeats (Weissenbach et al., 1996). This slow mutation rate enables alleles to be traced successfully through families. CA repeats are highly conserved as a repetitive element in mammalian genomes. CA repeats are effectively single locus probes, and exploited in genomic mapping. Once mapped, their highly polymorphic nature and Mendelian inheritance allows them to be used effectively in linkage and linkage disequilibrium analysis to map disease genes. In this Chapter, microsatellite analysis of the founder haplotype and $HFE$ mutation analysis were used to identify patients with a secure diagnosis of $HFE$-related haemochromatosis.
2.3. Methods

2.3.1. Multiplex fluorescent PCR

The following work was performed at the kind invitation of Professor Mark Worwood within the Department of Haematology, University of Wales College of Medicine, Cardiff. Multiplex fluorescent PCR amplification of CA repeats was performed on genomic DNA from both GH and control patients using a Hybaid Omnigene PCR machine. Four primer pairs were used in the multiplex PCR, and the forward primer of each pair was fluorescently labeled as indicated: D6S105 (HEX), D6S265 (FAM), D6S1260 (TET) and D6S1621 (TET). It was possible to use the fluorescent marker TET for both D6S1260 and D6S1621 primers as these microsatellites give PCR products differing in size by approximately 200bp, which can be readily differentiated from each other. The reaction profile was: 94.2°C for 40 seconds; 55°C for 40 seconds; 72°C for 40 seconds (30 cycles) followed by a final extension at 72°C for 5mins then a 26°C soak. PCR reactions contained; 1X Bioline PCR buffer; 1mM MgCl₂; 100μM each of dGTP, dCTP, dATP, dTTP; 1μM each of forward and reverse primers; 1U of Bioline Taq polymerase, in a total volume of 15μl.
2.3.1.1. Microsatellite analysis

The allele sizes of amplified CA repeats were analyzed on a Perkin Elmer ABI Prism 310 GeneScanner using the GeneScan application kit. The 310 was run in high resolution mode containing a 6.6M urea denaturing gel. A 3 percent polymer was used for this resolution giving a 50-250bp read range. The capillary tube containing the polymer was set to 40cm gel length, with sample injection set at 10 seconds to optimize resolution. Internal size standards were included with each sample to permit accurate sizing.

2.3.1.2. Sample preparation

PCR reactions (1μl) were mixed with 15μl deionised formamide and 0.5μl GS 500 size standards. Samples were denatured 2mins at 95°C, snap cooled on ice, and loaded into the 310.

2.3.1.3. Setup of the 310 ABI prism autosampler

Buffers and polymer for the electrophoresis were made according to the Perkin Elmer operational manual. For each sample the 310 will automatically form a new matrix within the capillary tube. After sample analysis the capillary is cleaned and neutralized by 0.3M NaOH and 1M HCl washes respectively and washed with deionised water. The next
aliquot of polymer is then drawn up into the capillary and analysis of the
next sample starts. Tubes containing buffers and polymers were placed
into the 310 in the following positions: 1, GeneScan polymer solution; 2,
dH₂O; 3, dH₂O; 4, 0.3M NaOH; 5, 1M HCl; 6, left empty for emergency
back flush. The anode buffer consisted of 1g genetic analyzer buffer
(10x) and 2.8g GeneScan polymer (7% w/w) made up to 10ml. The
cathode buffer consisted of 0.5g genetic analyzer buffer 10X, 2.2g
Genescan polymer (7% w/w), 2.0g urea plus 800ul of deionized water,
giving a 3% Gene scan polymer solution containing 6.6M urea. After
dissolving the urea the solution was passed through a 0.45μm filter
before use.

2.3.1.4. Run conditions

The ABI 310 was set to the following run and injection parameters:
Analyzer range 3250 - 5500; size call range 75 - 500bp; split peak
detection off; data smooth on; peak detection 70
RED/BLUE/GREEN/YELLOW; fit to line "global Southern"; run module
denatured; GS XT long denatured; injection 10 sec; Kv 15.00; run 15.00
Kv; temperature of block 42°C; runtime for each sample 20 mins; matrix
file dye set C and autoanalysis default on.
2.3.1.5. *Computer analysis of data*

Raw data were analyzed using Perkin Elmer Gene Scan Analysis software version 1 alpha on an Apple Macintosh™. An example of the data obtained is shown in Figure. 2.1.

2.3.2. *HFE mutation analysis*

DNA was prepared from peripheral blood from GH patients and controls using the QIAamp Blood kit (Qiagen). Cys282Tyr and His63Asp mutations were analyzed by PCR and restriction enzyme digestion (Worwood *et al.*, 1997).

2.4. *Results*

An example result of the output from the ABI 310 GeneScanner is shown in Figure. 2.1., below. The results from the analysis of 14 GH patients and 7 controls are summarised in Table. 2.1.
Figure 2.1. GeneScan trace showing relative fluorescent intensity plotted against approximate size (bp).

This Genescan trace above shows homozygosity for one marker (D6S1260-4/4, Green), also showing heterozygosity for two markers (D6S105-8/6, Black, D6S265-1/3, Blue), peaks for size standards are shown in red. The conversion between allele size and allele number is shown in Appendix 1.
Table 2.1. Patient founder haplotype microsatellite and $HFE$ mutation analysis results

<table>
<thead>
<tr>
<th>Individual</th>
<th>Age</th>
<th>Sex</th>
<th>Genotype</th>
<th>$HFE$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>D6S1621</td>
<td>D6S1260</td>
</tr>
<tr>
<td>GH1$^a$</td>
<td>42</td>
<td>m</td>
<td>5-5</td>
<td>4-4</td>
</tr>
<tr>
<td>GH2$^a$</td>
<td>34</td>
<td>m</td>
<td>5-6</td>
<td>3-4</td>
</tr>
<tr>
<td>GH3</td>
<td>52</td>
<td>m</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH4$^a$</td>
<td>74</td>
<td>m</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH5</td>
<td>68</td>
<td>m</td>
<td>5-7</td>
<td>3-3</td>
</tr>
<tr>
<td>GH6</td>
<td>34</td>
<td>m</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH7</td>
<td>36</td>
<td>m</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH8</td>
<td>46</td>
<td>f</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH9$^a$</td>
<td>33</td>
<td>m</td>
<td>5-7</td>
<td>4-4</td>
</tr>
<tr>
<td>GH10</td>
<td>66</td>
<td>f</td>
<td>5-5</td>
<td>3-4</td>
</tr>
<tr>
<td>GH11$^b$</td>
<td>39</td>
<td>m</td>
<td>7-10</td>
<td>4-4</td>
</tr>
<tr>
<td>GH1$^c$</td>
<td>54</td>
<td>m</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>GH13$^a$</td>
<td>50</td>
<td>m</td>
<td>5-7</td>
<td>4-5</td>
</tr>
<tr>
<td>GH102$^a$</td>
<td>48</td>
<td>M</td>
<td>5-10</td>
<td>4-4</td>
</tr>
<tr>
<td>C1$^a$</td>
<td>48</td>
<td>m</td>
<td>10-10</td>
<td>4-4</td>
</tr>
<tr>
<td>C2$^a$</td>
<td>57</td>
<td>m</td>
<td>8-8</td>
<td>3-4</td>
</tr>
<tr>
<td>C3</td>
<td>77</td>
<td>f</td>
<td>8-8</td>
<td>4-4</td>
</tr>
<tr>
<td>C4</td>
<td>53</td>
<td>m</td>
<td>10-10</td>
<td>4-4</td>
</tr>
<tr>
<td>C5$^a$</td>
<td>60</td>
<td>m</td>
<td>4-11</td>
<td>4-5</td>
</tr>
<tr>
<td>C6</td>
<td>54</td>
<td>m</td>
<td>6-10</td>
<td>4-5</td>
</tr>
<tr>
<td>C7$^a$</td>
<td>71</td>
<td>m</td>
<td>5-10</td>
<td>4-4</td>
</tr>
</tbody>
</table>

Legend on next page.
Results are shown for GH and control (C) patients. Homozygosity for alleles of the founder haplotype (D6S1621-5; D6S1260-4; D6S105-8; D6S265-1) is dark shaded. Regions of possible heterozygosity for the founder haplotype are lightly shaded.

NT, not tested; C282Y, Cys282Tyr; H63D, His63Asp; + and -, presence and absence of mutation; a patients investigated in ferric reductase assays (Chapter 5); b juvenile haemochromatosis; c homozygous for HLA-A3 and HLA-B7. The position of HFE is between the markers D6S1621 and D6S1260.
2.5. Discussion

2.5.1. General problems of microsatellite analysis

Microsatellites were originally observed on standard DNA sequencing gels. The microsatellites were frequently amplified using Taq polymerase and \( \alpha^{32}\mathrm{P} \) dCTP, to allow the polymorphic bands to be detected by autoradiography. This manual method is prone to the occurrence of three main artifacts:

(1) The differing mobility of the two strands of DNA may give rise to multiple bands. This is because the CA and GT strands of the amplified DNA migrate with differing mobilities under denaturing conditions. It may be remedied by using a single radiolabelled primer which will label only one strand, rather than incorporating a radiolabelled deoxynucleotide, which will label both strands.

(2) Polymerase slippage (by mispairing during the PCR) may give rise to extra bands differing by 2n in size. Changing the PCR reaction conditions however does not lead to any significant improvement of resolution.

(3) Non complementary base addition at the 3' end of PCR products (Clark 1988). This is treatable by treating the PCR end products with 3'-5' exonuclease.
The ABI 310 is not prone to mobility shift problems as only one of the primers is fluorescently labeled. Polymerase stutter may occur, but by using the GeneScan software it is possible to reduce any ladder to background, leaving clear peaks. Non-complementary base addition may be remedied by 3'-exonuclease treatment as for conventional microsatellite analysis. Although a drawback to using the 310 is its initial cost, the data produced are of high quality.

2.5.2. GH patient microsatellite data versus controls

The numbers of patients analysed are too small for independent linkage disequilibrium analysis. Even in the limited data set, however, there is evidence for the conservation of the founder haplotype (D6S265-1; D6S105-8; D6S1260-4; D6S1621-5; Raha-Chowdhury et al., 1995) in GH patients from the Royal Free Hospital.

Two patients (GH2 and GH11) had early onset of iron overload with cardiac involvement, and were diagnosed as having juvenile haemochromatosis. Of note was the observation that these patients did not have the founder haplotype, or either the C282Y or H63D mutation of HFE. Linkage of juvenile haemochromatosis to HFE was later excluded, and this disease has subsequently been recognized as a distinct
disorder, and linked to another locus on chromosome 1q (Roetto et al., 1999).

The genotype analysis also allowed selection of appropriate controls and GH patients for physical mapping of patient chromosomes by pulsed-field gel electrophoresis (Wallace et al., 1998).

These studies allowed the comparison of genotypes between the GH and control patients studied for ferric reductase analysis (Chapter 5). Following mutation analysis it was also possible to select patients with a secure diagnosis of HFE-related haemochromatosis who were homozygous for C282Y, and controls who were negative for C282Y, for ferric reductase analysis.
Chapter 3

Isolation and characterisation of genes around D6S1260
Chapter 3. Isolation and characterisation of genes around D6S1260:

Index.

3.1. INTRODUCTION .................................................................................. 68
3.2. METHODS ........................................................................................... 69
  3.2.1 YAC and PAC clones ....................................................................... 69
  3.2.2. Plating the cDNA library .............................................................. 71
    3.2.2.1. The Clontech human small intestine cDNA library HL1133a .... 71
    3.2.2.2. Titration of the cDNA library .................................................. 72
    3.2.2.3. Plating a 1X coverage of the library ......................................... 73
    3.2.2.4. Preparation of duplicate library filters for screening ............... 73
  3.2.3. Screening cDNA library .................................................................. 74
    3.2.3.1. PAC insert probe preparation .................................................. 74
    3.2.3.2. Pre-hybridisation of filters with placental DNA competition ...... 75
    3.2.3.3. Competition of the radiolabelled probe ..................................... 76
    3.2.3.4. Washing of cDNA filters ......................................................... 77
    3.2.3.5. Analysis of results from the 715 / 672 PAC insert probe ............ 77
    3.2.3.6. Secondary Screen ................................................................... 78
    3.2.3.7. Tertiary screen ....................................................................... 78
  3.2.4. Subcloning the cDNA clones ........................................................... 79
    3.2.4.1. PCR amplification of positive plaques ...................................... 79
    3.2.4.2. EcoR I subclones of cDNA PCR products ............................... 79
    3.2.4.3. Cloning of PCR products using CloneAmp pAMP1 vector ...... 80
    3.2.4.4. Amplification of cDNAs for pAMP1 cloning ............................. 80
    3.2.4.5. Ligation of cDNAs into pAMP1 vector ................................... 81
  3.2.5. Sequencing of the cDNAs ............................................................... 82
  3.2.6. Computer analysis of cDNA sequences ......................................... 83
  3.2.7. Hybridisation analysis of cDNA inserts ......................................... 84
  3.3. RESULTS............................................................................................. 84
  3.4. DISCUSSION....................................................................................... 91
3.1. Introduction

This work was undertaken in spring 1996, before the cloning of the haemochromatosis candidate gene (*HFE*) by Feder and co-workers (1996). As described in Section 1.3 (Genetics of haemochromatosis), at this time the apparent peak of linkage disequilibrium was the chromosome 6 polymorphic CA repeat marker D6S1260 (Raha-Chowdury et al., 1995). We therefore concentrated on this region to isolate and characterise new genes hoping to identify potential candidate genes for haemochromatosis as well as other novel gene sequences.

There are several methods, which can be used to isolate genes from cloned regions of genomic DNA. The method used was the direct hybridisation of genomic clones to a cDNA library. This approach requires the use of cDNA from a defined tissue. This is potentially useful in reducing the number of transcripts to be screened, but involves the presumption that the gene or genes are expressed in the particular tissue. The major advantage of this method, however, is that it is simple and rapid.
Two overlapping PACs 11g1pCYPAC2N (715) and 302f12pCYPAC2N (672), spanning 148kb of genomic DNA, and containing D6S1260, were used to screen a cDNA library of human small intestine. The choice of cDNA reflected the probability that this tissue expresses the gene for haemochromatosis, since the small intestine is the site of iron absorption from the diet.

3.2. Methods

3.2.1 YAC and PAC clones

Yeast artificial chromosome (YAC) clones from the HFE gene candidate region were obtained from the CEPH library (Albertsen et al., 1990; Chumakov et al., 1995).

P1-derived artificial chromosome (PAC) clones from the Roswell Park Park Cancer Institute (RPCI-1) PAC library were obtained from the Human Genome Mapping Project Resource Centre (HGMP-RC). PACs from the HFE gene region were identified by several methods: direct request of previously published PACs; hybridization of whole YACs onto gridded PAC library filters (Baxendale et al., 1991); PCR amplification from primary, secondary and tertiary PAC library pools; and by PAC ‘walking’ using T7 and SP6 end specific riboprobes on gridded PAC library filters (Stratagene SuperCos protocol).
YAC clones were grown at 30°C for two days in selective media deficient in uracil and tryptophan. YAC DNA was prepared in agarose blocks according to the lithium dodecyl sulphate (LiDS) method of the Wellcome Trust Summer School – Human genome analysis: from genome to function (http://www.wellcome.ac.uk/summerschools). Cells were immobilized in agarose blocks, and DNA isolated via treating with zymolase and yeast lysis solution containing lithium dodecyl sulphate.

PAC clones were grown at 37°C overnight in Luria-Bertani (LB) media containing the selective antibiotic kanamycin. PAC DNA was prepared using standard alkaline lysis mini-preparation methods (Sambrook et al., 1989) or by using QIAfilter plasmid midi kits (Qiagen). The alkaline lysis procedure is based on the following principle: Cells are first pelleted and lysed by treating them with sodium dodecyl sulphate and NaOH. The SDS solubilizes the cell membranes causing the cells to lyse thus releasing their contents. NaOH causes the DNA and proteins to denature. Subsequent addition of acetate buffer neutralizes the lysate and allows the DNA to renature. The smaller size of the plasmid DNA and the proximity of the two interlocked strands means that the plasmid DNA renatures quickly and remains in solution. The chromosomal DNA however remains single stranded, and together with the denatured proteins and other cellular debris becomes trapped in the insoluble salt and detergent complexes. The Qiagen plasmid purification procedure uses an anion exchange column, which binds double stranded DNA.
Contaminating RNA, proteins and other impurities can then be removed in a low salt wash and DNA eluted in a high salt wash. The purified plasmid can then be precipitated using either ethanol or isopropanol and plasmid DNA resuspended in TE buffer (10mM Tris-HCl, 1mM EDTA, pH8).

Pulsed-field gel electrophoresis analysis, Southern blotting and hybridization with 32P-radio-labeled total human DNA determined the size of YAC clones. The size of the PAC inserts was determined by Not I digestion and PFGE analysis. YAC and PAC clones were digested with Hind III, fragments separated by agarose gel electrophoresis and Southern blotted to Hybond N+ membranes. Probes and polymorphic markers across the \textit{HFE} region were localized on the YACs and PACs by hybridization to Southern blots and by PCR analysis.

3.2.2. Plating the cDNA library

3.2.2.1. \textit{The Clontech human small intestine cDNA library HL1133a}

A cDNA or \textit{complementary} DNA is synthesized as a single-stranded copy of an RNA template, using an RNA-dependent DNA polymerase (reverse transcriptase). The first strand cDNA is copied using a DNA-dependent DNA polymerase, giving double stranded cDNA, which is cloned into a suitable vector. A cDNA library is a collection of cDNAs that ideally represents all the RNA molecules expressed by a specific tissue. The
cDNA library chosen for screening was a Clontech human small intestine 5'-Stretch cDNA library (HL1133a). This is a commercial library prepared from the duodenal tissue of a 15 year old Caucasian female. The mRNA was primed with both oligo (dT) and random primers for reverse transcription, and then EcoR I cloned into the vector λGT10.

3.2.2.2. Titering the cDNA library

The following methodology is modified from Sambrook et al., (1994). The *Escherichia coli* host strain Y1090r- was revived from glycerol stock held at -70°C and streaked onto an LB agar 1.5% (w/v) Petri dish containing 50μg/ml ampicillin and grown at 37°C overnight. A single colony was used to inoculate 10ml of LB broth containing 10mM MgSO₄ and 0.2% (w/v) maltose without added antibiotic. The broth was incubated at 37°C in a shaker at 200rpm overnight until the OD₆₀₀ reached approximately 2.0. The culture was centrifuged at 650g for 10mins and the pellet re-suspended in the same volume of 10mM MgSO₄ ("plating cells") and stored at 4°C. Titering dilutions of 1/500, 1/5,000 and 1/50,000 were prepared from the library using 1X lambda dilution buffer (0.1M NaCl, 10mM MgSO₄, 7H₂O, 33mM Tris-HCl (pH7.5) and 0.01% (w/v) gelatine).

The titering dilutions (100μl) and control (100μl 1X dilution buffer) were added to 200μl of the Y1090r- plating cells, and incubated at 37°C for 15mins. Melted (45°C) LB top agarose 0.7% (w/v) containing 10mM MgSO₄ was mixed with each dilution thoroughly and poured onto pre-
warmed (37°C) LB plus 10mM MgSO₄ agar plates, ensuring even spreading across the plate. The plates were allowed to solidify and then incubated at 37°C inverted for approximately 7 hours, until plaques were clearly visible. Plaques were counted and plaque forming units (pfu) per ml of stock calculated.

3.2.2.3. Plating a 1X coverage of the library

Using the titering information from above, dilutions were prepared from stock to give 250,000 pfu on each of eight 22 X 22cm LB plus 10mM MgSO₄ agar plates. The Clontech HL1133a small intestine library is estimated to contain 1.6 X 10⁶ independent cDNA clones. Inoculating 250,000 pfu on eight 22 x 22cm agar plates thus gave a single coverage of all the cDNAs present. The plates were poured using the same method as for titering, except that each 22 x 22cm plate contained 200ml of LB plus 10mM MgSO₄ agar and 50ml LB top agarose plus 10mM MgSO₄. The plates were inverted and incubated at 37°C for approximately 7 hours until the plaques were clearly visible, and not touching each other.

3.2.2.4. Preparation of duplicate library filters for screening

When plaques reached the optimum size, the plates were removed from the 37°C incubator and chilled at 4°C to harden the top agarose. Sterile Hybond™ N+ (Amersham) 22X22cm nylon membrane filters (2 filters per plate) were marked with the corresponding plate numbers. These were
placed using sterile forceps carefully on top of the top agarose for 60 seconds. Care was taken not to smudge the plate, introduce bubbles or disturb the pattern of plaques.

The orientation of the nylon filters was marked by piercing each one with an 18 gauge needle to give an asymmetric pattern. The first filter was removed, and the duplicate filter placed on the same plate for 60 seconds, and marked with the same needle holes. The filters were incubated twice plaque-side up on Whatman 3MM filter paper saturated with DNA denaturing solution (1.5M NaCl, 0.5M NaOH) for 60 seconds. The filters were then placed on Whatman 3MM saturated with neutralizing solution (1.5M NaCl, 0.5M Tris-HCl (pH8.0) twice for 60 seconds. The filters were briefly washed in 4 X SSC, placed on damp Whatman 3MM and UV cross linked at 1200J using a Stratagene StrataLinker, and allowed to dry prior to hybridization.

3.2.3. Screening cDNA library

The human genomic inserts of PACs 715 and 672 were used to screen the small intestine cDNA library.

3.2.3.1. PAC insert probe preparation

PACs 715 and 672 were digested using the restriction enzyme Not I to release the human insert. The digest was separated by pulsed field gel
electrophoresis (PFGE) using a BioRad CHEF DRIII system (Ho & Monaco 1992). A 0.8% (w/v) agarose gel was pulsed in three blocks of 24 hours with switch times of 12sec (96°, 2V/cm), 15sec (100°, 2V/cm), 18 sec (96°, 2V/cm).

The DNA was visualised by ethidium bromide staining. The human DNA insert bands (PAC 715, 10kb and 115kb; PAC 672, 33kb and 49kb) were excised from the gel under long wave UV light. The excised human DNA inserts from each PAC were pooled, β-agarase-digested and phenol chloroform extracted. The DNA was resuspended in 30µl TE pH8.0. The DNA concentration of the inserts was estimated by absorbance at A_{260}. 25ng of the human insert DNA was radiolabelled by random hexanucleotide priming (Feinberg & Vogelstein, 1983) using α^{32}P dCTP (Amersham).

3.2.3.2. Pre-hybridisation of filters with placental DNA competition

The duplicate cDNA library filters were pre-wet using 4X SSC, sandwiched between hybridisation meshes and placed in an AppliGene hybridisation bottle. 20ml of formamide pre-hybridisation solution containing 100µg/ml denatured, sheared human placental DNA was added in order to quench repetitive DNA sequences. The bottle was placed in a rotating oven at 42°C overnight.
3.2.3.3. **Competition of the radiolabelled probe**

The radiolabelled probe was pre-annealed with a mixture of sheared human placental DNA, Alu / Line repeats and ribosomal probes to quench repeat sequences contained within human genomic DNA.

Approximately 500,000 to 1,000,000 copies of the Alu repeat element and 100,000 - 800,000 copies of the Line repeat have been estimated to occur in the human genome (Boeke, 1997). To compete the probe, the following reagents were added to give a final concentration as indicated: human placental DNA (2.5mg/ml); Bluescript DNA (50µg/ml); Alu repeat DNA (5µg/ml); LINE repeat (5µg/ml); ribosomal DNA (R7.3 and R5.8 probes; 5µg/ml each). The probe mixture was denatured at 100°C for 10min and snap-cooled on ice. The probe was briefly centrifuged to collect evaporation, cold 20X SSC was added to a final concentration of 5X, and the mixture was placed at 65°C for 90mins to allow the repeat DNA to anneal to the probe.

The prehybridisation solution was then removed from the filters and replaced with 20ml fresh hybridisation solution. The competed probe was added to give 1 X 10⁶cpm/ml and the filters were hybridised overnight at 42°C.
3.2.3.4. Washing of cDNA filters

After hybridisation the hybridisation solution was discarded and the filters were washed twice in 4X SSC at room temperature. The filters were washed twice in 0.2X SSC, 0.1% SDS at 60°C for 15mins, rinsed in 4X SSC and placed between plastic wrap. The filters were placed together on X-OMAT XAR-5 x-ray film for 2 days (Kodak) at -70°C using an intensifying screen to enhance the signal.

3.2.3.5. Analysis of results from the 715 / 672 PAC insert probe

After developing, the x-ray films were marked at the 3 asymmetrical needle holes. The duplicate filter films were aligned and a total of 24 potentially positive plaques were found to be reproduced on both filters. Due to the high density of the plaques on the 22 X 22cm agar plates it was not possible to pick out individual positive plaques for PCR amplification at this stage. A 5mm diameter agar plug was pulled from the plates containing the positive plaques and placed in 1ml of sterile lambda dilution buffer and left to elute overnight. A second round of screening was then undertaken.
3.2.3.6. Secondary Screen

A 1 μl aliquot from each of the 24 eluted phage was added to 100 μl of the Y1090r- plating cells. The mixture was incubated for 10 mins at 37°C then added to melted (45°C) LB top agarose 0.7% (w/v) containing 10mM MgSO₄, and poured onto pre-warmed (37°C) LB plus 10mM MgSO₄ 15cm diameter agar plates. The plates were incubated inverted for 12-14 hours until individual plaques could be seen. Sterile Hybond™ N+ (Amersham) 15cm circular nylon membrane filters were used to copy the plates as described previously. Again the orientation of the nylon filters was marked by piercing asymmetrically 5 times with an 18 gauge needle. The filters were probed by hybridisation as described in screening cDNA library.

3.2.3.7. Tertiary screen

From this second round of screening 11 positive plaques were identified. Due to the high density of plaques on the second round plates there was the possibility of cross contamination of clones picked for PCR. As a precaution a third round of screening was undertaken to reduce the density of plaques and possibility of cross contamination while picking. The 11 positives were pulled as plugs and re-screened using the above method. This third round of screening yielded single positive plaques that were distinct enough to be picked for PCR.
3.2.4. Subcloning the cDNA clones

3.2.4.1. PCR amplification of positive plaques

The 11 positive plaques identified by hybridisation were picked using sterile pipette tips and inoculated into the following PCR reaction mixture to act as DNA template: 1X GibcoBRL PCR buffer; 1.5mM MgCl$_2$; 200μM each of dGTP, dCTP, dATP, dTTP; 500ng λGT10 forward and reverse primers in a total volume of 100μl. A no DNA negative control was also included. The reaction profile was that of a hot start: 94°C for 2mins then hold and add 2.5U (0.5μl) GibcoBRL Taq polymerase, then 94°C for 45 seconds; 72°C for 45 seconds; 55°C for 2mins increasing by 1 second per cycle for 35 cycles followed by a 10°C soak. The PCR products were separated by electrophoresis on a 1% (w/v) agarose gel and visualised by ethidium bromide staining. From the 100μl total PCR reaction, 60μl was purified using QIAquickspin (Qiagen) purification columns and eluted with 50μl 10mM Tris HCl (pH8.0).

3.2.4.2. EcoR I subclones of cDNA PCR products

10μl of purified DNA was digested with EcoR I, examined on an agarose gel and subcloned into the vector pBluescript (Stratagene). Before ligation, EcoR I was heat inactivated for 20mins at 68°C. The ligation
reaction contained: 100-200ng EcoR I-digested PCR product (3.0µl); 200ng EcoR I-digested and calf intestinal alkaline phosphatase-treated Bluescript; 6.25µl dH₂O; 0.5µl 10X T4 ligation buffer (NEB); 0.25µl T4 DNA ligase (2 X10⁶ U/ml (NEB) at 11°C overnight. XL-1 Blue (Stratagene) competent cells were transformed using 5µl of this ligation mixture. Transformants were picked for plasmid DNA preparation by alkaline lysis "mini-preps" (Sambrook et al., 1994), and for preparation of glycerol stocks.

3.2.4.3. Cloning of PCR products using CloneAmp pAMP1 vector

CloneAmp is a system developed by GibcoBRL to facilitate the high efficiency cloning of PCR amplification products. Sequence specific PCR primers are designed so that they incorporate dUMP residues in their 5’ ends (forward 5’ CAD CAU CAD CAD 3’, reverse 5’ CUA CUA CUA CUA 3’). After PCR, the amplified product is treated with uracil DNA glycosylase (UDG), resulting in the dUMP residues becoming abasic and unable to base pair. This exposes a 12 base pair 3’ overhang which can be annealed to the complementary sequence in the linearised vector pAMP 1.

3.2.4.4. Amplification of cDNAs for pAMP1 cloning

PCR was performed on the QIAquickspin-purified primary amplification products of the cDNAs. The primers used were adaptations of λGT10
primers, \( \lambda \text{GT10F-CAU} \) and \( \lambda \text{GT10R-CUA} \). The PCR reaction contained: 
0.3\( \mu \)l of QiaQuick (Qiagen) cDNA amplification product, or sterile \( \text{dH}_2\text{O} \) 
negative control; 1X GibcoBRL PCR buffer; 1.5mM MgCl\(_2\); 200\( \mu \)M each of 
dGTP, dCTP, dATP, dTTP; 500ng \( \lambda \text{GT10F-CAU}/\lambda \text{GT10R-CUA} \) 
forward and reverse primers in a total volume of 100\( \mu \)l. The reaction 
profile was that of a hot start: 94°C for 2mins, then hold and add 2.5U 
(0.5\( \mu \)l) GibcoBRL Taq polymerase; then 94°C for 45 seconds, 55°C for 
45 seconds, 72°C for 3mins, increasing by 1 second per cycle for 20 
cycles; followed by a 10°C soak. The PCR products were separated on a 
1\% (w/v) agarose gel and then visualised by ethidium bromide staining.

3.2.4.5. Ligation of cDNAs into pAMP 1 vector

The dUMP-containing PCR product (1\( \mu \)l; 50-100ng) was added to a 
reaction mixture containing: 1X annealing buffer (20mM Tris-HCl (pH8.4), 
50mM KCl, 1.5mM MgCl\(_2\)) (Gibco BRL); 0.5\( \mu \)l pAMP 1 vector DNA 
(12.5ng); 0.25\( \mu \)l (10U/\( \mu \)l) uracil DNA glycosylase in a total volume of 5\( \mu \)l. 
The mixture was incubated at 37°C for 30mins and placed on ice. 1\( \mu \)l of 
the reaction mixture was used to transform competent XL-1 blue \( \text{E. coli} \) 
(Stratagene). Transformants were picked for plasmid DNA preparation by 
alkaline lysis "mini-preps", and for preparation of glycerol stocks.
3.2.5. Sequencing of the cDNAs

A single colony from each of the transformed cDNAs was used to inoculate 100ml of LB broth containing 50μg/ml ampicillin. The cultures were grown overnight at 37°C in a shaking incubator at 300rpm. When the culture reached approximately 1 optical density unit (ODU) at A600, the culture was harvested and centrifuged at 3000rpm to obtain a pellet. This pellet was then used for alkaline lysis maxi-preparation of plasmid DNA using QIAfilters (Qiagen).

The maxi-preparations of plasmid DNA were sequenced by the Sanger dideoxy-chain termination method using the Sequenase Version 2.0 sequencing kit (United States Biochemical). The 5' and 3' ends of the cDNAs were sequenced with the λGT10 forward and reverse PCR primers. Sequencing products were subjected to electrophoresis under denaturing conditions for 5 hours at 60V in a BioRad Sequi-Gen sequencing tank containing 6% acrylamide : bisacrylamide (>200bp range) SequaGEL (National Diagnostics). The sequencing gel was vacuum dried at 80°C for 2 hours and exposed to BioMAX MR film (Kodak).
3.2.6. **Computer analysis of cDNA sequences**

The DNA sequences obtained from the ends of the cDNAs were screened for regions of homology to known sequences using the BLAST and FASTA (Altschul *et al.*, 1994; Pearson & Lipman 1988) citations. These programs search for regions of similarity between the test sequence and sequences held within specified databases.

Two types of BLAST search were performed: BLASTN (DNA test sequence searching for homology to DNA sequences held on the database) and BLASTX (DNA test sequence with six frame translation searching for homology to predicted protein sequence held on the database). The nucleic acid databases searched with BLASTN were: G, GenBank Library; E, EMBL Library; R, Alu-like sequences; C, cDNA sequence library; P, Eukaryotic Promoter Data; J, Fugu Project sequences; 2, Sanger CpG Island library; 3, EMBL minus EST/STS/GSSs.

The following are sections of EMBL that were used for more focused searches: x, ESTs; h, Human; l, Invertebrates; o, Organelles; a, Other Mammals; v, Other Vertebrates; t, Patent Sequences; d, Rodents; s, STSs; u, Unclassified; q, New EMBL sequences. Protein databases searched with BLASTX were: S, Swissprot Protein Library; M, TREMBL (translation of EMBL); N, NBRF (PIR) Protein Library. The databases were also searched using the command nr for non-redundant database.
FASTA searches were made using the same combination of databases as BLASTN and BLASTX.

Results from the homology search (BLASTN, BLASTX) are returned with a homology score and a probability score. The homology score shows local alignment as a percentage identity (nucleotide against nucleotide and predicted protein against protein) and as a percentage similarity (predicted protein against protein only). The probability gives an indication of the strength of the similarity. The data obtained must however be taken as whole when considering possible homologies.

3.2.7. Hybridisation analysis of cDNA inserts

cDNA insert probes were prepared by EcoR I digestion of PCR products and gel purification, radiolabelled by the random hexamer method (Feinberg and Vogelstein 1983) and hybridised to Southern blots of the 11 cDNA inserts; Hind III-digested PAC clones from the region; and Hind III-digested human genomic DNA.

3.3. Results

The results from screening the cDNA library and analysis of the positive clones are summarised in Table 3.1. Of the 11 cDNAs isolated, only 4 clearly mapped back to one of the parent PACs originally used to isolate the cDNAs. The cDNAs 1, 2, 3 and 4 all hybridised to PAC 715
(11g1pCYPAC2N) as well as to YACs 81 and 119, (CEPHy900E03878 and CEPHy904727B2). These four cDNAs all cross-hybridised to each other. All four had BLAST and FASTA sequence homologies to zinc finger transcription factors. cDNA3 was shown to contain the expressed sequence tag STS-9945. Of these 4 cDNAs, two gave single copy bands when hybridised to Hind III-digested human genomic DNA, the remaining two zinc finger cDNAs, however, gave a smear on hybridisation.

The ends of these two zinc finger clones when sequenced were found to be chimeric. The cDNA 3 (2CS) contained 110bp of the β-amyloid gene, from amino acid 172 to 226 (100% nucleotide identity) adjoining the pBluescript cloning vectors SK primer site. The β-amyloid gene maps to chromosome 21 with no pseudogenes reported on chromosome 6p21.3. The cDNA 4 (4CS) contains approximately 3kb of the epidermal growth factor sequence (100% nucleotide identity), adjoining the λGT10 forward primer in the cloning vector pBluescript. The gene for epidermal growth factor has been mapped to chromosome 19 with no reports of any pseudogenes in chromosome 6p21.3. Chimeric clones such as these are frequently observed in cDNA libraries, and are a consequence of the method by which the libraries are created.

Of the other 7 cDNAs, sequence from cDNA 5 (4BS) did not have significant homology to any sequences held in the BLAST database. The sequence of cDNA 6 (1AL) had 92% nucleotide identity to various
cytochrome oxidase subunits. The cDNA 7 (2AL) sequence showed a 100% amino acid identity to bacterial rRNA sequence.

The cDNAs 8 (2BS), 9 (4AS), 10 (5AS) and 11 (5AS2) when sequenced all showed high homology to repeat sequences held in the BLAST databases: cDNAs 8 and 9 contained Alu sequence; cDNA 10 had high homology to the Alu class of retrotransposons and cDNA 11 had high homology to the LINE class of repeat elements, and also contained a CA repeat of unknown origin. On hybridisation cDNAs 5 to 11 did not give a signal on any of the PACs originally used to screen the cDNA library, and gave a smear on Hind III digested genomic DNA.
Table. 3.1. End sequence and hybridisation analysis of eleven cDNA clones isolated by hybridisation of PACs 715 and 672 to a small intestinal cDNA library.

YACs 81 and 119 correspond to clones CEPHy900_e_03878 and CEPHy904727_b_2, respectively; PACs 715 and 672 correspond to clones 11g1pCYPAC2N and 302f12pCYPAC2N, respectively.

<table>
<thead>
<tr>
<th>cDNA No.</th>
<th>Name</th>
<th>Size (kb)</th>
<th>Homology by BLAST analysis</th>
<th>Hybridisation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>cDNA</td>
<td>PAC</td>
<td>YAC</td>
</tr>
<tr>
<td>1</td>
<td>5FL</td>
<td>0.9</td>
<td>zinc finger protein 1.2.3.4</td>
<td>1,2,3,4</td>
<td>715</td>
</tr>
<tr>
<td>2</td>
<td>3AS2</td>
<td>2.7</td>
<td>zinc finger protein 1.2.3.4</td>
<td>1,2,3,4</td>
<td>715</td>
</tr>
<tr>
<td>3</td>
<td>2CS</td>
<td>1.6</td>
<td>zinc finger protein 1.2.3.4</td>
<td>1,2,3,4</td>
<td>715</td>
</tr>
<tr>
<td>4</td>
<td>4CS</td>
<td>5.5</td>
<td>zinc finger protein 1.2.3.4</td>
<td>1,2,3,4</td>
<td>715</td>
</tr>
<tr>
<td>5</td>
<td>4BS</td>
<td>5.5</td>
<td>None</td>
<td>5</td>
<td>-ve</td>
</tr>
<tr>
<td>6</td>
<td>1AL</td>
<td>1.7</td>
<td>Cytochrome oxidase</td>
<td>6</td>
<td>-ve</td>
</tr>
<tr>
<td>7</td>
<td>2AL</td>
<td>1.2</td>
<td>bacterial rRNA</td>
<td>7</td>
<td>-ve</td>
</tr>
<tr>
<td>8</td>
<td>2BS</td>
<td>0.9</td>
<td>Alu</td>
<td>8,9</td>
<td>-ve</td>
</tr>
<tr>
<td>9</td>
<td>4AS</td>
<td>2.9</td>
<td>Alu</td>
<td>8,9</td>
<td>-ve</td>
</tr>
<tr>
<td>10</td>
<td>5AS</td>
<td>1.2</td>
<td>retro-transposon</td>
<td>10,11</td>
<td>-ve</td>
</tr>
<tr>
<td>11</td>
<td>5DS2</td>
<td>1.5</td>
<td>LINE and CA repeat</td>
<td>11</td>
<td>-ve</td>
</tr>
</tbody>
</table>
Table 3.2. Low resolution YAC / PAC map of the haemochromatosis gene region showing the position of ZNF204.

<table>
<thead>
<tr>
<th>YAC</th>
<th>Size [kb]</th>
<th>SSADH</th>
<th>D6S1261</th>
<th>P44S</th>
<th>P44T</th>
<th>BTN</th>
<th>PY138</th>
<th>PY147</th>
<th>D6S1260</th>
<th>700</th>
<th>ZNF204</th>
<th>PY117</th>
<th>820</th>
<th>825</th>
</tr>
</thead>
<tbody>
<tr>
<td>878_e_3</td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>950_h_11</td>
<td>1760</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>727_b_2</td>
<td>1200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>899_g_1</td>
<td>850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>779_d_7</td>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>947_f_6</td>
<td>750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>792_g_12</td>
<td>610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>790_g_7</td>
<td>780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>755_g_10</td>
<td>1510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The position of ZNF204 is shown in red; blue indicates the presence of markers in YACs; green indicates presence of markers in PACs; light blue indicates YAC deletions; NT indicates not tested. Yeast artificial chromosome (YAC) clones were obtained from the CEPH library (Albertsen et al., 1990; Chumakov et al., 1995). The position of HFE was subsequently shown to be between probe p44T and butyrophilin (BTN) (Feder et al., 1996).
3.4. Discussion

At the time of isolation before the cloning of the haemochromatosis candidate gene (Feder et al., 1996), genes isolated from around D6S1260 (at that time the apparent peak of linkage disequilibrium) were potential candidates for the haemochromatosis gene. Haemochromatosis is a disease where the normal regulation of iron absorption is lost leading to increasing iron overload. Zinc finger proteins regulate the expression of other genes and loss of function due to mutation may therefore lead to loss of regulation.

The 4 cDNAs containing the zinc finger homology were chosen to be fully sequenced, because they were considered potential haemochromatosis candidates on the basis of the available mapping data and their possible functional role, and because they represented a potentially novel zinc finger locus which had been isolated by positional cloning from the 6p21.3 region.

The other 7 cDNAs were excluded from further analysis at this time because of their cross-homology to known repeat sequences, repetitive hybridisation pattern on genomic Hind III digested DNA and the fact that they did not map back to either of the PACs used to original isolate the cDNAs (PAC 715 or 672). The identification of these cDNAs was most likely due to cross-hybridisation of repeat sequences contained within the PACs, despite the use of competing DNA in the hybridisation.
Repeat sequences are found throughout the genome, although there is evidence of clustering in centromeric and telomeric regions in humans. There are two classes of dispersed repetitive DNA: SINEs (short interspersed elements) and LINEs (long interspersed elements). SINEs are shorter than 500bp and occur $10^5$-$10^6$ times in the human genome. An example of which is the Alu family of repeats, classically around 300bp in length. LINEs are longer (average size 5kb) and occur approximately $10^4$ times in the human genome. Retrotransposons comprise part of this family of repeats.

This identification of false positives due to cross hybridisation with SINEs and LINEs is one drawback to screening a cDNA library by hybridisation with genomic clones. Other methods such as exon trapping (Buckler et al., 1991) and genomic sequencing are unaffected by the problem of cross-hybridisation, although these methods are more time consuming and / or large scale procedures.

The characterisation of the novel zinc finger locus encoded by cDNAs 1-4 is described in Chapter 4: Characterisation of ZNF204.
Chapter 4

Characterisation of ZNF204
Chapter 4. Characterisation of ZNF204: Index.

4.1. INTRODUCTION ........................................................................... 93
4.2. COMPLETE CDNA SEQUENCE .................................................. 98
  4.2.1. Methods ............................................................................... 98
  4.2.2. Results ................................................................................ 98
  4.2.3. Conclusion from sequence investigation of ZNF204 ............. 108
4.3. ORGANISATION OF THE GENOMIC LOCUS ......................... 109
  4.3.1. Introduction ....................................................................... 109
  4.3.2. Methods ............................................................................ 109
  4.3.3. Results and Discussion ...................................................... 110
4.4. NORTHERN BLOT ANALYSIS ................................................. 112
  4.4.1. Introduction ....................................................................... 112
  4.4.2. Methods ............................................................................ 112
  4.4.3. Results and Discussion ...................................................... 116
4.5. 5' RAPID AMPLIFICATION OF CDNA ENDS (5' RACE) .......... 118
  4.5.1. Introduction ....................................................................... 118
  4.5.2. Methods ............................................................................ 118
  4.5.3 Results ............................................................................... 118
4.6. SEARCHING FOR ZNF204 HOMOLOGUES ............................. 122
  4.6.1. Somatic cell hybrid panel analysis ....................................... 122
  4.6.2 Methods ............................................................................. 122
  4.6.3. Results and Discussion ...................................................... 123
4.7. HYBRIDISATION SCREENING OF A GENOMIC PAC LIBRARY 125
4.8. FLUORESCENT IN SITU HYBRIDISATION (FISH) ..................... 126
  4.8.1. Introduction ....................................................................... 126
  4.8.2. Methods ............................................................................ 127
  4.8.2.1. cDNA probe .................................................................. 127
  4.8.2.2. PAC probes used for FISH .............................................. 127
  4.8.2.3. Method used for FISH ................................................... 127
  4.8.3. Results and Discussion ...................................................... 130
4.9. ZOO BLOT ANALYSIS OF ZNF204 ......................................... 135
  4.9.1. Methods ............................................................................ 135
  4.9.2. Results and Discussion ...................................................... 135
4.10. IN VITRO TRANSCRIPTION AND TRANSLATION OF ZNF204 139
  4.10.1. Methods .......................................................................... 140
  4.10.1.1. TNT reaction conditions ............................................... 140
  4.10.1.2. Denaturing gel analysis of translation products .............. 141
  4.10.2. Results and Discussion ...................................................... 141
4.11. ZNF 204 LOCUS DISCUSSION ............................................. 144
Chapter 4. Characterisation of ZNF204.

4.1. Introduction

Zinc finger (ZNF) proteins are transcription factors, which regulate the transcription of other genes. One of the first gene regulators to be characterized was TFIIIA, a zinc finger protein from Xenopus that regulates the transcription of ribosomal 5S RNA (Miller et al., 1985). The zinc finger transcription factors take their name from their structure, where four key residues are arranged in a tetrahedral formation, and together co-ordinate a zinc atom. The zinc finger domain is composed of 25 to 30 amino acid residues, with two cysteine and two histidine residues arranged at the extremities of the domain (the so called C2H2 domain). It is these residues that co-ordinately bind the zinc atom. Multiple zinc finger domains may be found tandemly arranged in transcription factors. The protein sequence of the full ~30 amino acid zinc finger domain is highly conserved:

Tyr/Phe-X-Cys-X-Cys-X_{2,4}-Cys-X_5-Phe-X_5-Leu-X_2-His-X_{2,5}-His-X_5
where X is any amino acid. The 30 amino acid repeating unit is referred to as the zinc finger domain on the basis of its tertiary structure forming a loop of 12 amino acids. This loop contains the conserved leucine and phenylalanine residues as well as being rich in basic residues. It is these basic residues such as arginine and lysine that interact with the acidic DNA (see Figure 4.1).
Figure 4.1. Possible structures of the two histidine two cysteine zinc finger.

![Diagram](image)

(a) (b)

Adapted from Latchman1995.

Showing the loop of 12 amino acids and (b) showing the proposed structure of an antiparallel beta-pleated sheet and an alpha-helix around the co-ordination sites of the zinc ion. This binding of the zinc atom has been directly confirmed by Extended X-ray absorption fine structure Spectroscopy (EXAFS) (Diakun, et al., 1986).
The zinc fingers interact with the major groove of the DNA helix. The finger interacts with five bases of the DNA sequence (or half a helical turn). Successive fingers bind on opposite sides of the helix (Klug and Rhodes 1993). Although first shown in TFIIIA, an RNA polymerase III-associated transcription factor, C2H2 fingers have subsequently been identified mainly as transcription factors of RNA polymerase II (Struhl 1989). The association between the presence of a zinc finger motif and transcription factors is so strong that it is possible to probe with the zinc finger domain to isolate new transcription factors.

An example of this was the isolation of a zinc finger ZNF141, the gene responsible for Wolf-Hirschhorn syndrome on chromosome 4p. The gene was isolated using a degenerate oligonucleotide of the C2H2 consensus sequence (Tommerup et al., 1993). It is estimated that the human genome contains several hundred zinc finger genes containing the highly conserved C2H2 domain (Bray et al., 1991). Screening of cosmid and cDNA libraries with zinc finger sequences has led to the identification of over 130 novel zinc finger sequences (Hoovers et al., 1992). Another example of a disease caused by mutation of a C2H2 zinc finger gene is Greig cephalopolysydactyly syndrome Engelkamp and Heyningen (1996). Wilm's tumor is also associated with the presence of a defective C2H2 zinc finger on chromosome 11p13. This defective zinc finger causes both Wilm's tumor (Gessler et al., 1990), and congenital malformations of the urogenital tract (Pritchard-Jones et al., 1991).
Zinc finger proteins therefore have been shown to play a key role in regulation of gene transcription. Several human diseases result from mutation of a zinc finger protein. The complete characterization of the novel zinc finger locus, which was isolated from chromosome 6p21.3, is described in this Chapter.

To characterise the novel ZNF cDNA the complete sequence was first determined and analysed. The sequence obtained was screened for sequence homologies to known gene sequences, to determine motifs and potential open reading frames. The genomic locus was also sequenced, to search for introns.

Expression was investigated using northern blot analysis to determine sites of expression. 5’ rapid amplification of cDNA ends was used to search for additional 5’ transcribed sequences. The possibility of translation from the ZNF was investigated by coupled transcription translation analysis.

Four approaches were used to search for loci homologous to the ZNF: Computer based database homology search; hybridisation screening of a genomic PAC library; single chromosomal hybrid panel analysis and fluorescent in situ hybridisation.
4.2. Complete cDNA sequence

4.2.1. Methods

The complete cDNA sequences of the 4 zinc finger-containing cDNAs, 2CS, 3AS2, 5FL and 4CS were determined by primer walking. Each primer walk was sequenced by the Sanger dideoxy-chain termination method, using the Sequenase Version 2.0 sequencing kit (United States Biochemical). Individual sequences were aligned by eye and using the BESTFIT and PILEUP programs of the GCG suite available via telnet at menu.hgmp.mrc.ac.uk.

4.2.2. Results

The sequence analysis revealed that the four cDNAs were all from the same locus (Genome nomenclature approved name ZNF204, GenBank accession number AF033199). Two of the four cDNAs were found to be chimeric. Figure 4.2. summarises the mapping and positional cloning of the cDNAs encoding the ZNF204 locus. The ZNF204 sequence, as deposited in GenBank, is given in Figure 4.3.
Figure 4.2. Mapping, isolation and characterization of ZNF204. (See following page for legend).
(a) Fine resolution genomic map around D6S1260 indicating marker positions, with ZNF204 shown in bold. Below, PAC clones used to screen the small intestinal cDNA library; dots indicate the presence of markers (Raha-Chowdhury et al., 1995; Pappas et al., 1995; Volz et al., 1997; Wallace et al., 1998) as determined by PCR or hybridization. (b) Overlapping cDNA clones showing the position of EST449 and STSG-9945; a double slash indicates the junction of ZNF204 cDNA clones with chimeric sequence. (c) Schematic representation of ZNF204 showing the 1.1 kb ORF consensus sequence and the positions of its start ATG (arrow) and polyadenylation consensus sequence (pA). Blocks indicate the positions of individual zinc finger motifs with * representing termination codons and lightning indicating the location of the frame shift mutation. The position of the PCR hybridization probe is indicated above the consensus. Below the consensus, the 3 bars represent the 3 overlapping PCR products amplified from genomic DNA using cDNA primers; cDNA and genomic DNA products were of identical size.
Figure 4.3. ZNF204 sequence, as deposited in GenBank (Accession No. AF033199)

LOCUS AF033199 2630 bp mRNA PRI 25-JUN-1998
DEFINITION Homo sapiens C2H2 zinc finger protein pseudogene, mRNA sequence.
ACCESSION AF033199
NID g3252864
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2630)
TITLE Cloning and molecular characterization of a cross-homologous zinc finger locus ZNF204
MEDLINE 98292504
REFERENCE 2 (bases 1 to 2630)
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1997) Academic Medicine, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK
FEATURES Location/Qualifiers
  source 1..2630
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="6p21.3; between D6S1260 and D6S105"
    /tissue_type="small intestine"
  gene 1..2630
    /note="expressed processed pseudogene"
    /gene="C2H2 zinc finger protein"
    /pseudo
Chapter 4  Characterisation of ZNF204  Page 102

ORIGIN

1  gggggcatgc tggctcctctt cactctccac catgatgta aagttctctga ggcttcccac
gttgcctgc tggagaattg accaaagagct cgggaagagg tggctatggg
21  acatcaaccc ggagaaattg aagcaaaagt gcccttgctca aaagactgca taaagaaaca
181  taactggaga aataatagaa ccgtgttggat ctgtatggga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
301  aagaatctgcc aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
301  aagaatctgca aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
301  aagaatctgct aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
301  aagaatctgcc aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
301  aagaatctgctc aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
301  aagaatctgcc aagagaagaa ttccttcctg aagacccagt aagttggagt tggctatgct
601  aaaaaaccttt gtaataataag tgaagtgagg aagacccagtcc tactcttcag tggagagaga
601  aaaaaaccttt gtaataataag tgaagtgagg aagacccagtcc tactcttcag tggagagaga
721  tttaatcaga gttcccccct aagaaaaatt cagaaaaaggg cagaaaaaggg
781  tcacacagtc gtagctggtg ggagcctttg actgttctttct tgaataataag tgaagtgagg
841  gaattccacc tgggagagaaaa cctttgttcttg gagttctgag aatgtaataag tgaagtgagg
901  tcgtaataac ctatctctag atacaaagac tatcataatgc tcttcattttt tgaataataag tgaagtgagg
961  cttaaatggt gcaggcttctt cagacactaat ctatctctag atacaaagac tatcataatgc tcttcattttt tgaataataag tgaagtgagg
1021 actgagagag aacccattttg atgcgagaaacc atgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1081 cttaataatacc taattctggtg gtacactataa gtagcagactt atcataaatgc tcttcattttt tgaataataag tgaagtgagg
1141 aagacccactt atctctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1201 ccctttgttcttg gagttctgag aatgtaataag tgaagtgagg
1261 aagacccactt atctctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1321 ttcttggtcg gcacatatgtt tggagagagaaaa cctttgttcttg gagttctgag aatgtaataaa tgaagtgagg
1381 caattgggtat ggagccactttg cagcaacacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1441 ggagagaagccc atcataatgg ttagaagagc cagacactaat ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1501 acatcaaccc ggagaaattg aagcaaaagt gcccttgctca aaagactgca taaagaaaca
1561 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1621 gtctttttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1681 catcaacagct aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
1741 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1801 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1861 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1921 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
1981 taatactttggg aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2041 gcttgcctcc tctttctttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
2101 taataactttggg aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2161 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2221 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2281 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2341 ttctttttttt gtaacccacta ctgcaattct ttcgagagtt gatgtaataaa tgaagtgagg
2401 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2461 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2521 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
2581 aagacccagtcc tactcttcag tggagagaga aagttgaggag tgggcccaag
gacacgagc tagagacgca accagccacc actctcagcc tattattattt tattagaaaa
The ZNF204 locus contained the primer sequences for two expressed sequence
tags (ESTs), STSG-9945 (Whitehead Institute) and EST449 (Pappas et al.,
1995), (Figure. 4.2.). As expected, STSG-9945 was located within the 3'
untranslated region, originating from oligo dT priming of the polyA tail. However
EST449 was derived from an internal polyA region, and was located close to the
5' end of the transcript. Neither of the EST sequences contained a zinc finger
motif, thus it is their placement within the transcript which allowed them to be
classified as part of a zinc finger locus.

The ZNF204 transcript contained a total of 18 two cysteine-two histidine (C2H2)
zinc finger motifs, the complete sequence of the ZNF204 cDNAs revealed the
presence of a premature stop codon upstream of zinc finger motifs 14-18 and a
frame shift mutation within zinc finger motif 17 (Figure. 4.2.). Sequence analysis
of both genomic and cDNA clones revealed that only 10 of the 18 zinc finger
motifs conformed fully with the core consensus \( C-X_{2.4}-C-X_3-F-X_5-L-X_2-H-X_{2.5}H; \)
Klug and Rhodes, 1987). The remaining motifs contain mutations which predict
amino acid substitutions at conserved cysteine and histidine residues (see Table.
4.1.). In Drosophila and yeast, mutations of these conserved residues have been
shown to eliminate the function of the Kruppel and ARD1 zinc finger proteins,
respectively (Blumberg et al., 1987; Redemann et al., 1988).
Table 4.1. Multiple mutations in the zinc finger and linker motifs of ZNF204.

The consensus motif for the zinc finger loop (Choo and Klug 1994) is shown along with the 13 ZNF204 motifs before the first stop codon. The most highly conserved residues of the motif are highlighted in bold red; mutations are shown in blue (ZNFs 3, 12, 13).

<table>
<thead>
<tr>
<th>Zinc finger No.</th>
<th>Consensus Sequence;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CX_{2-4} CXXFXXXXLXXXH_{2-5}</td>
</tr>
<tr>
<td>1</td>
<td>CKYSECGRTRGHITLVQHQITH</td>
</tr>
<tr>
<td>2</td>
<td>CSECGRVSVCLVLNQHQRIH</td>
</tr>
<tr>
<td>3</td>
<td>CLEQRLFSQTHLTMQRIH</td>
</tr>
<tr>
<td>4</td>
<td>CNECEKAHDHSALIQHHV</td>
</tr>
<tr>
<td>5</td>
<td>CNECGNASDCSSLQHQRTH</td>
</tr>
<tr>
<td>6</td>
<td>CKQCGKAFSTLTQHQRSI</td>
</tr>
<tr>
<td>7</td>
<td>CNDGIPFSCALYQHKRIH</td>
</tr>
<tr>
<td>8</td>
<td>CNDCKPFSRCALIHERI</td>
</tr>
<tr>
<td>9</td>
<td>CTDGKAFCDLASLQHQITQH</td>
</tr>
<tr>
<td>10</td>
<td>CNECGKASQTYLVKIH</td>
</tr>
<tr>
<td>11</td>
<td>CTECGRAFSWSTDKNHKTH</td>
</tr>
<tr>
<td>12</td>
<td>CNECRKAFSGLYQCVI</td>
</tr>
<tr>
<td>13</td>
<td>YGKCGKAFQRTDLKKHKMH</td>
</tr>
</tbody>
</table>
A possible DNA binding sequence for the 10 putative functional zinc finger motifs is postulated in Table. 4.2. It is possible to postulate a DNA binding sequence as the zinc finger binds to the DNA double helix by inserting the $\alpha$-helical part of the finger into the major groove. The positioning of the three amino acids at positions $-1$, 3 and 6 of the ZNF consensus control the binding specificity (Choo and Klug 1994; and Nardelli, et al., 1991).

It was hoped that the predicted DNA binding sequence of the functional zinc finger motifs could be used to determine the gene this zinc finger regulated, however, the predicted sequence., nCnnT/cnnGA/TnGA/TnTCnTcCnnGnnnnTnnnnnTcnnn is too non specific to enable further electronic or PCR investigations (Table 4.2.).
Table 4.2. Predicted binding sequence of ZNF204 of the 10 putative functional zinc finger motifs (legend on next page).

<table>
<thead>
<tr>
<th>Zinc Finger No.</th>
<th>Zinc finger motif amino acid sequence</th>
<th>Predicted DNA binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CKYSECGRTRGHITLVQHITH</td>
<td>n C n</td>
</tr>
<tr>
<td>2</td>
<td>CSECGKAFSVCLVNLHQRIH</td>
<td>n Tc n</td>
</tr>
<tr>
<td>4</td>
<td>CNECEKAFHDSALIQHIVH</td>
<td>n G A/T</td>
</tr>
<tr>
<td>5</td>
<td>CNECGNAFSDKSLIQHQRTH</td>
<td>n G A/T</td>
</tr>
<tr>
<td>6</td>
<td>CKQCGKAFSTRYLTQHQRSH</td>
<td>n T C</td>
</tr>
<tr>
<td>7</td>
<td>CNDGIPFSCASLYQHKRIH</td>
<td>n Tc C</td>
</tr>
<tr>
<td>8</td>
<td>CNDCKPFSRCSALIQHIRH</td>
<td>n n G</td>
</tr>
<tr>
<td>9</td>
<td>CTDCGKAFLVQHQTQH</td>
<td>n n n</td>
</tr>
<tr>
<td>10</td>
<td>CNECGKAFSQSTYLVQHQTIKH</td>
<td>n T n</td>
</tr>
<tr>
<td>11</td>
<td>CTECGRASWSTDKKNHQKTH</td>
<td>n Tc n</td>
</tr>
</tbody>
</table>
Only the ten zinc fingers upstream of the first stop which conform to the consensus are shown. The amino acids highlighted in blue determine the interaction of that finger with DNA, and allow some prediction of the bound DNA sequence, as shown (Choo and Klug 1994 and Nardelli et al., 1991). Where both capital and lower case letters are used to show the predicted nucleotides bound, capitals indicate the statistically more likely nucleotide.
4.2.3. Conclusion from sequence investigation of ZNF204.

The detection of multiple mutations of conserved zinc finger residues lead to the conclusion that although ZNF204 was discovered via a cDNA library, thus indicating ZNF204 RNA was produced, it may well be a pseudogene. To investigate this possibility the genomic locus was investigated to look for introns.
4.3. Organisation of the genomic locus

4.3.1. Introduction

PCR primer pairs were designed to span the ZNF204 cDNA consensus sequence. These were used to investigate genomic DNA to define intronic boundaries. Seven overlapping sets of primers were used to amplify cDNA and genomic DNA templates. Intron containing PCR products would give rise to larger PCR products detectable by agarose gel electrophoresis. Three key overlapping PCR reactions were used to screen ZNF204 (Figure 4.2.).

4.3.2. Methods

The PCR reaction was as follows: 1X GibcoBRL PCR buffer; 1.5mM MgCl₂; 100μM each of dGTP, dCTP, dATP, dTTP; 500ng of each primer set and template DNA or sterile dH₂O negative control in a total volume of 25μl. The reaction profile was that of a hot start: 94°C for 2mins then hold and add 2.5U (0.5μl) GibcoBRL Taq polymerase then 94°C for 45 seconds; 55°C for 45 seconds; 72°C for 2mins for 35 cycles followed by a 10°C soak. The PCR products were separated by electrophoresis on a 1% (w/v) agarose gel and visualised by ethidium bromide staining.
The genomic sequence was determined by analysis of Hind III subclones of PAC pCYPAC2N11g1 in pBluescript phagemid. Sequence homology analysis was performed using the BLAST program (Altschul et al., 1994).

4.3.3. Results and Discussion

Amplification of the genomic ZNF204 locus in 7 overlapping segments using cDNA primers revealed that the 2.7kb ZNF204 is colinear with the genomic locus at the level of resolution of electrophoresis of PCR products. Lack of introns is suggestive evidence for ZNF204 being a pseudogene. However, functional genes can exist without introns an example of which is the histone family.

The possibility that a small exon was not detected by amplification was investigated by sequence analysis of the genomic locus. This analysis was performed on Hind III subclones of PAC pCYPAC2N11g1 in pBluescript phagemid. The genomic sequence of the 2.7 kb ZNF204 locus was identical with the cDNA sequence, confirming the lack of introns within this region. The lack of introns, multiple mutations in functional zinc finger motifs, and apparent truncation of the 5' region together indicate that the locus is may be an expressed processed pseudogene.
To search for a "parent" locus and investigate the similarity of ZNF204 to published sequences, homology analysis was performed using the BLAST program (Altschul et al., 1994) on a non-redundant database. BLASTN revealed the highest homology to ZNF135: 66% nucleic acid identity over 117 nucleotides. However reanalysis of the highest 10 homologies using the BESTFIT (GCG) program showed that ZNF184 (second highest homology with BLASTN) had the closest homology to ZNF204 with 68% nucleic acid identity over 1963 nucleotides. ZNF184 is also from the 6p21.3 region and within 500kb of ZNF204 (Goldwurm et al., 1997; Wallace et al., 1998). When only the putative 5' UTR of ZNF204 was searched for homology, ZNF184 was again detected. This homology was however to the predicted coding "spacer" region of ZNF184 suggesting that the 5' region of ZNF204 may have been derived from coding sequence. It seems likely that the putative "parent" locus is in 6p21.3, and possibly related to ZNF184.
4.4. Northern blot analysis

4.4.1. Introduction

To investigate the size of the mRNA transcript(s) homologous with the ZNF204 locus, 2 cDNAs 1 and 2 (spanning the ZNF204 locus) and a short PCR product (STSG-9945) were used as hybridisation probes on a multi tissue northern blot. By analysing hybridisation on a northern blot it may be possible to ascertain if the overlapping DNA sequence for cDNAs 1, 2, 3 and 4 represent the entire cDNA sequence of the ZNF204 locus, or only part of it. Northern analysis also enables the simultaneous analysis of mRNA expression and transcript size in different tissues.

The multi-tissue northern blot analysed contained human polyA(+) mRNA from spleen, thymus, prostate, testis, ovary, small intestine, colon (non-mucosa) and peripheral blood leukocyte (Clontech Human II #7759-1).

4.4.2. Methods

cDNAs 1,2 and 3 (25ng) were radiolabelled by random hexanucleotide priming (Feinberg and Vogelstein, 1983) using $^{32}$P dCTP (Amersham). PCR products from STSG-9945 (cDNA1 as a template) and from ZNF184 (Goldwurm et al., 1997), genomic DNA as a template were also radiolabelled. The northern blot
and suitable positive and negative control Southern blots were pre-hybridised. The prehybridisation solution was removed from the filters and replaced with 20ml fresh hybridisation solution. The probe was competed as previously described (Chapter 3) and added to give $1 \times 10^6$cpm/ml. The filters were hybridised overnight at 42°C.

The blots were briefly washed using 4X SSC then washed twice in 0.5X SSC 0.1% SDS at 55°C for 15mins. The filters were rinsed in 4X SSC and placed on BioMAX MS x-ray film (Kodak) at -70°C using an intensifying screen to enhance the signal.
Figure 4.4. Northern analysis of the ZNF204 locus expression on a polyA(+) multi-tissue northern blot.

Legend on next page.
a) The hybridisation illustrated is for STSG-9945, it is however representative of all the ZNF204 hybridisations performed (cDNA 1 and 2 and STSG-9945 PCR products). A 3.2kb transcript was visible in all tissues tested: spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocyte. An additional 8.5kb band was also seen in prostate, ovary and testis. The lane loading observed using β-actin would indicate however, the 8.5kb band could be observable in other tissues if loading was more comparable.

b) Hybridisation of the probe β-actin on the multi-tissue northern blot to check lane loading. The hybridisation pattern suggests near equal lane loading.
4.4.4. Results and Discussion

ZNF204 hybridization was detected in spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocyte (Figure 4.4.). The polyA(+) mRNA transcript size for ZNF204 is 3.2kb. The total overlapping cDNA sequence obtained for ZNF204 is however 2.7kb. Therefore approximately 500bp of sequence is missing from the overlapping cDNAs sequences. The observation that both ZNF204 and ZNF184 hybridise to a 3.2kb transcript on northern analysis is consistent with the BLAST analysis, suggestive of a common ancestry. The BLAST analysis also suggests that the 5' region of ZNF204 may be truncated, consistent with the northern analysis. In an attempt to clone this missing 5' sequence the 5' Rapid Amplification of cDNA Ends approach was used.

On three tissues, prostate, ovary and testis an additional 8.5kb transcript was observed. There are several possible explanations for this larger transcript. An alternative downstream polyadenylation sequence could be used, giving rise to a larger 3' untranslated region. An alternative splicing event could give rise to a larger transcript; however, no evidence has been found for introns in the genomic locus of ZNF204. The 8.5kb band could correspond to mRNA from a homologous locus. Finally, the “polyA(+) mRNA” used for the northern blot has previously been shown to contain nuclear RNA precursors (Clontech personal
communication of unpublished data). This therefore seems a likely explanation for the additional 8.5kb bands.
4.5. 5' Rapid Amplification of cDNA Ends (5' RACE)

4.5.1. Introduction

The creation of a cDNA library is often based on the priming of reverse transcriptase with oligo[dt]. This method ensures the production of 3' sequence, however, depending on length and secondary structure there is often loss of the 5' terminal sequence.

ZNF204 has a 3.2kb mRNA transcript on northern analysis. The overlapping cDNA sequence for ZNF204 however only gives a total of 2.7kb. There is therefore potentially 0.5kb of missing 5' sequence to ZNF204. In an attempt to clone this missing 5' end of ZNF204, the 5' rapid amplification of cDNA ends method was used (5' RACE).

4.5.2. Methods

The method of 5' RACE used was a modification of that of Frohman (1993). 5' RACE or “anchored” PCR allows the isolation of 5' terminal sequence from mRNA via the utilisation of an anchored PCR primer.
Figure 4.5. Diagrammatic overview of 5' RACE procedure.

1. **mRNA**
   - Anneal first strand primer, GSP1, to mRNA

2. **Copy mRNA into cDNA with SuperScript™ II RT**

3. **Degrade RNA with RNase Mix**

4. **Purify cDNA with GelMax Spin Cartridge**

5. **Tail purified cDNA with dCTP and TdT**

6. **PCR amplify dC-tailed cDNA using the Abridged Anchor Primer and nested GSP2**

7. **Copy PCR product using AUAP, or UAP, and nested GSP"
Poly(A)+ lymphocyte mRNA was used for the first round cDNA synthesis. Total RNA can be used for this procedure but, with the low level of ZNF204 mRNA expected from the northern blot analysis, poly(A)+ was used to enrich for rare cDNA sequences (Figure 4.7). GSP1 was designed 730bp downstream of the 5' end of the ZNF204 cDNA sequence: (a) to allow hybridisation screening using ZNF204 5' probes on the 5' RACE products and (b) to give a PCR product over 500bp, for better recovery using the GlassMAX spin columns.

The cDNA product was purified to remove un-incorporated dNTPs and GSP1. The purified product was then incubated with the recombinant enzyme terminal deoxynucleotidyl transferase (rTdT) and dCTP. This enzyme produces a homopolymeric tail on the 3' termini of the cDNAs. A PCR reaction was then used to amplify the cDNA of interest using an “abridged-anchored” primer to the 3' homopolymeric tail and a nested PCR primer 3' to the reverse transcription primer GSP1, GSP2 (GSP2 was designed to have an annealing temperature close to that of the “abridged-anchor” primer, 660bp from the ZNF204 cDNAs 5' termini). A second round of PCR was used to increase the concentration of product produced for cloning using a 3' nested primer to GSP2, GSP3. The primers used for this final stage of amplification were designed to allow the use of the CloneAmp system (GibcoBRL) of PCR cloning by the incorporation of a CAU-tail to the upstream “anchored” primer and a CUA-tail to GSP3 (630bp from ZNF204 cDNA sequence 5' terminus).
4.5.3 Results

The PCR amplification of cDNA using GSP2 gave a smear of bands on the ethidium bromide stained agarose gel (all negative controls were negative). PCR amplification using the nested primer GSP3 however failed to give any bands (all negative controls were also negative).

No bands were observed after hybridization of these Southern blots, indicating no specific 5' RACE product had been amplified. The GSP2 and GSP3 primer positioning 660bp and 630bp from the current ZNF204 cDNA 5' terminus means products of 1.66Kb and 1.63Kb respectively should be amplified from ZNF204 RNA if the transcript was 3.2kb and not 2.7kb as indicated by the northern analysis. The lack of any products therefore probably reflects the technical limitation of this procedure.
4.6. Searching for ZNF204 homologues

4.6.1. Somatic cell hybrid panel analysis

A somatic cell hybrid panel is made by the fusion of a mouse or hamster cell line with a human cell line. As the fused cells divide, the human chromosomes are systematically lost. The process of human chromosome retention is random, creating chromosomally distinct cell lines some eventually containing a single human chromosome. These "human mono-chromosomal" cells are then isolated from the cell population and cultured into a single human chromosome cell line (Warburton et al., 1990).

4.6.2 Methods

As described previously the 2.7kb ZNF204 cDNA is colinear with genomic DNA. Three pairs of PCR primers spanning the ZNF204 locus (Figure 4.2.) were used to screen the human DNA contained within a monochromosomal somatic cell hybrid panel of Povey and co-workers (UK HGMP Resource Centre), in order to search for a putative "parent" locus. Each of the three PCR reactions contained 50ng of each chromosome somatic cell hybrid DNA, a sterile dH2O control and parental rodent cell line controls (50ng human, mouse or hamster genomic DNA).
Mouse and hamster genomic DNA templates were included with each primer set because the somatic cell hybrid DNA also contains the genomic DNA from the host cell line. This "host" DNA could potentially also be amplified with the ZNF204 primers.

4.6.3. Results and Discussion

All three primer sets only gave the expected size bands in the genomic positive control PCR reaction. The lack of PCR amplification from the chromosome 6 monochromosomal hybrid cell line appears to be due to the cell line used for chromosome 6, MCP6BRA. This cell line (MCP6BRA) has a breakpoint in the short arm of chromosome 6 and is missing the tip of this arm including the region 6p21.3. Hybridisation of PCR reactions with the radiolabeled cDNA 3AS2 was consistent with this interpretation, only showing a positive signal with the genomic control.

The lack of an apparent parent locus detected by the monochromosomal somatic cell hybrid panel PCR analysis may be due to several reasons. The primers chosen to screen the panel may have flanked large introns in the "parent" locus, the PCR conditions being optimised for the smaller ZNF204 intron-less copy of the gene. The "parent" locus may be contained in a "gap" in the library panel's chromosomal coverage. The parent may be close to ZNF204, hence missing in the chromosome 6 cell line MCP6BRA. In an attempt to resolve
these possibilities, homologues of ZNF204 were sought by genomic library screening and fluorescent in situ hybridisation.
4.7. Hybridisation screening of a genomic PAC library

The PCR product of STSG-9945 was used as a hybridisation probe to screen human PAC library filter grids (UK HGMP Resource Centre) in order to search for homologues of ZNF204. The sequence of STSG-9945 is contained within the 3' UTR of ZNF204 and as a consequence contained within PAC 715 (pCYPAC2N11g1). Three PACs were found to be positive with STSG-9945: PAC 715 (as expected) as well as 832 (pCYPAC2N145g14), and 833 (pCYPAC2N102h8).

Further PCR analysis of these PACs showed that PAC 833 contained PY117, a marker approximately 110kb centromeric to D6S1260. When 833 was analysed by PCR using primers spanning ZNF204, it gave the same size PCR products as 715, indicating it was most likely also from chromosome 6p21.3, and probably overlaps the same locus as 715. PAC 832 however did not contain PY117, and failed to give PCR products from the set of primers spanning ZNF204. PAC 832 therefore might contain a possible homologue and/or “parental” locus of ZNF204, and as such was included as a probe in fluorescent in situ hybridisation.
4.8. Fluorescent *in situ* hybridisation (FISH)

4.8.1. Introduction

Fluorescent *in situ* hybridisation (FISH) is a technique that allows the detection of specific nucleic acid sequences on morphologically preserved chromosomes by the use of a fluorescent probe. The position of the fluorescent probe on the metaphase chromosomes gives mapping data on the probe used. *In situ* hybridisation was originally developed by Pardue and Gall (1969). This was before the advent of molecular cloning and utilised biochemically isolated DNA (e.g. satellite DNA, ribosomal RNA) that was radiolabeled as the probe. There are now non-radioactive methods for labeling DNA for *in situ* hybridisation such as fluorescence.

FISH may use either direct or indirect reporting. In direct reporting, the detectable molecule (reporter) is directly bound to the nucleic acid probe. In the indirect method, the probe contains a reporter molecule, but detection of the probe requires additional incubation with an affinity partner to produce the signal.

The presence of the premature stop and frame shift mutations and lack of introns in ZNF204 makes it most likely an expressed processed pseudogene. By using FISH it was possible to screen human metaphase spread chromosomes for a “parent” locus or identify the position of other copies of this locus.
4.8.2. Methods

4.8.2.1. cDNA probe

The cDNA used as a probe for a “parent” locus was cDNA2 as it was not only the largest of the four cDNAs (2.7kb) it did not contain any chimeric sequence. The use of a 2.7kb cDNA sequence in FISH is close to the limit of resolution (approximately 1.5kb of probe sequence). For this reason an indirect method of labelling was adopted, as it allows the enhancement of signal strength.

4.8.2.2. PAC probes used for FISH

PAC 832 was tested as a possible parental locus of ZNF204. PAC 715, which contains cDNA2, was included as a positive control probe for FISH.

4.8.2.3. Method used for FISH

FISH was performed using a modification of the method described by Pinkel and co-workers (1986) (Gillet et al., 1993, Banfi et al., 1996). DNA from PAC and cDNA clones was labeled with biotin-14-dATP using a GibcoBRL (Life Technologies) Bionick Kit. The labeled probe was mixed with two competitor DNAs: Cot-1 DNA (GibcoBRL) and polydeoxyadenylic acid-deoxythymidylic
acid (poly[dA]-[dT] (Sigma). Herring sperm DNA was also added as a blocking agent. The DNA was precipitated using sodium acetate and ethanol and re-suspended in hybridisation buffer containing 50% formamide (v/v), 10% dextran sulphate (w/v) and 2X SSPE at pH 7.0.

Slides of human metaphase chromosomes were prepared from normal male lymphocyte culture. The normal male lymphocyte culture was synchronised by the addition of thymidine to block DNA synthesis. The block was then removed and 5 bromo-deoxyuridine was allowed to incorporate prior to harvest for use in background chromatin staining. Standard cytogenetic methods were used in the colcemid arrest and hypotonic treatment of the cultured cells prior to their fixing using methanol/acetic acid onto glass slides (Banfi et al., 1996). The slides of metaphase chromosomes can be stored at -70°C for up to 6 months. Metaphase spreads were kindly provided by Dr M. Fox, Galton Laboratory, UCL.

The metaphase slides used were prepared for FISH by pre-treating with RNAs (10µg/ml) and proteinase K (0.035µg/ml) and prefixed with 1% (v/v) formaldehyde (in PBS containing 5% MgCl₂ w/v). The slides were dehydrated in an ethanol series (70% / 90% / 100%) before denaturing in 70% formamide, 2 X SSC for 5min at 75°C. The slides were dehydrated again in a series of 4°C ethanol washes (70% / 90% / 100%). The probe was then denatured at 75°C for 5mins, added to the slides and allowed to anneal overnight at 37°C.
The annealed slides were washed by immersion in 50% (v/v) formamide in 2X SSC at 42°C (three times for 5mins), and in 0.1 X SSC, (two times for 2.5min at 42°C). Slides were blocked using 5% (w/v) Marvel non fat milk in 4 X SSC for 20min, and washed in 0.05% (v/v) Tween 20 / 4 X SSC. The fluorescent signal was produced by incubation in a 5µg/ml solution of fluorescein isothiocyanate-avidin conjugate (FITC-avidin; Vector Laboratories) in 5% (w/v) Marvel, 4X SSC, for 20mins. This signal was amplified by adding biotinylated rabbit anti-avidin (5µg/ml in Marvel, 4X SSC; Vector laboratories) for 20mins, followed by a second round of FITC-avidin incubation again for 20mins.

Chromosomes were identified by banding made visible by using the fluorescent counter stain diamidino phylindole (DAPI) and propidium iodide which were included in the cover slip mounting medium Vectashield (Vector laboratories).

The slides were investigated using a Nikon Optiphot fluorescence microscope connected to a confocal laser microscopy (MRC600) image capturing system. Digital images were obtained using a cooled CCD (Charge Coupled Device) camera (Photometrics) attached to a Zeiss Axiophot fluorescent microscope. The digital image obtained was analysed and stored on an Apple Macintosh using the SmartCapture software from Digital Scientific.
4.8.3. Results and Discussion

Numerous individual chromosomal metaphase spreads were examined using fluorescent microscopy and digital imaging enhancement software.

PAC 715 gave a strong signal only on chromosome 6p21.3 indicating that it was not chimeric and confirming the original PCR mapping data in (Figure 4.6.).

The ZNF204 probe, cDNA2, had high background fluorescence due to the small size of the probe and the length of exposure time required to observe a signal (Figure 4.7.) cDNA2 did however give small specific signals only on chromosome 6p21.3. To confirm this signal 11 cells were imaged in detail. Of these 64% of the chromatids contained a clear signal on chromosome 6p21.3. A small number of other chromosomes contained a signal. These were however not reproduced in either sister chromatids or other cells examined, making them most likely to be non-specific. However, the resolution of standard FISH is approximately 1Mb. A second “parent” locus within about 1Mb of ZNF204 on chromosome 6p21.3 therefore cannot be ruled out.

PAC 832 gave a strong clear band on chromosome 14q in the immunoglobulin region (Figure 4.8.). Further analysis of the insert of PAC 832 using STSG-9945, 3AS2, 2CS, and 5FL as hybridisation probes however failed to give any positive a signals. The non-reproducibility of the original positive identification of PAC
832, and the lack of FISH or human monochromosomal somatic cell hybrid panel PCR evidence for a second locus for ZNF204 on chromosome 14, leads to the conclusion that PAC 832 may have been identified originally by cross-hybridisation and not by sequence specificity.
Figure 4.6. Fluorescent in situ hybridisation of PAC 715 on normal metaphase chromosomes, showing hybridisation to chromosome 6p21.3.

Chromosomes were counter stained with 2,6'-diamidino-phenylindole (DAPI) and the signal detected with FITC. Images were captured by cooled CCD (Photometrics) from a Zeiss Axioskop fluorescence microscope using Smartcapture software (Digital Scienific).
Figure 4.7. Fluorescent in situ hybridisation of cDNA 3AS2 on normal metaphase chromosomes, showing hybridisation to chromosome 6p21.3.

Chromosomes were counter stained with 2,6'-diamidino-phenylindole (DAPI) and the signal detected with FITC. Images were captured by cooled CCD (Photometrics) from a Zeiss Axioskop fluorescence microscope using Smartcapture software (Digital Scienific).
Figure 4.8. Fluorescent in situ hybridisation of PAC 832 on normal metaphase chromosomes, showing hybridisation to chromosome 14.

Chromosomes were counter stained with 2,6'-diamidino-phenylindole (DAPI) and the signal detected with FITC. Images were captured by cooled CCD (Photometrics) from a Zeiss Axioskop fluorescence microscope using Smartcaputre software (Digital Scienific).
4.9. Zoo blot analysis of ZNF204

To analyse the evolutionary conservation of ZNF204 zinc finger locus, zoo Southern blots of Hind III-digested and Taq I-digested DNA were screened with a PCR fragment from ZNF204. The zoo Southern blots contained restriction enzyme digested DNA from human, pig, cow, dog and hamster.

4.9.1. Methods

A 0.5kb region of the cDNA2 was amplified and radiolabelled. The PCR probe contained the cDNA sequence of the zinc finger motifs 1, 2, 3, 4 and 5. The hybridisation of the probe was performed at 42°C. The blots were washed at low stringency: once with 4X SSC at room temperature, twice for 20mins in 0.5X SSC, 0.1% SDS at 42°C and once for 20mins in 0.5X SSC, 0.1% SDS at 65°C.

4.9.2. Results and Discussion

The ZNF204 probe was shown to be conserved across species. The Hind III digested zoo blot displayed a ladder of hybridising bands increasing uniformly in size by approximately 80bp in all species tested (human, cow, pig, and hamster; Figure 4.9.). This ladder of bands was absent from the Taq I digested zoo blot where only a smear was observed (data not shown).
The step-wise increase in the size of the Hind III fragments corresponds to the distance between the phenylalanine residues of adjoining zinc finger domains of 84bp. The DNA-encoding sequence Lys-Ala-Phe occurs commonly in zinc fingers (where phenylalanine is part of the conserved zinc finger consensus), and gives rise in some cases to a Hind III site. A computer analysis of the 10 zinc finger genes with highest homology to ZNF204 also revealed that 15 out of the 20 Hind III sites present were positioned at the phenylalanine residue within the conserved sequence. Two zinc finger genes contained 5 Hind III sites, all at the phenylalanine residue of the consensus, giving rise to a ladder of Hind III fragments as detected by hybridisation in Figure 4.9.

The conservation of such zinc finger probes has been previously reported, with one or several cross hybridizing bands detected on Southern blots of human DNA, and on zoo blots (Shannon et al., 1996; Villa et al., 1996; Chowdhury et al., 1987). This experiment demonstrated a remarkable regular ZNF204 hybridization pattern conserved across all species tested, which may be explained in terms of the periodicity of Hind III sites within the consensus sequence of multiple zinc finger proteins. At least 35 individual bands can be resolved in human and bovine DNA, suggesting that the zinc finger genes, and therefore proteins, of these species can in theory contain at least 35 motifs. Indeed Xfin protein of Xenopus, contains 37 C2H2 zinc finger motifs, consistent with these results (Ruiz-i-Altaba et al., 1987). The cross-hybridization of ZNF204
to multiple zinc finger loci across species may provide a tool for the analysis of C2H2 zinc fingers in different species throughout evolution.
Figure 4.9. ZNF204 cross hybridizes with multiple zinc finger loci.

Southern blot of a Hind III digested DNA from different species probed with a 0.5kb fragment of ZNF204, with size in kb shown on right. A ladder of bands increasing uniformly by approximately 80bp was seen in all lanes reflecting cross-hybridization of ZNF204 to multiple zinc finger loci across species.
4.10. *In vitro* transcription and translation of ZNF204

As described earlier the sequence investigation of cDNA ZNF204 showed ATG and inframe termination codons upstream of the main ZNF open reading frame, as well as a "premature" stop codon upstream of zinc finger motifs 14 –18, and a frame shift mutation within zinc finger motif 14. To investigate whether the ZNF open reading frame may be translated, and whether these mutations may result in a truncated protein, cDNA2 was analysed using a cDNA-directed protein synthesis reaction, utilising a coupled eukaryotic *in vitro* transcription-translation system. The 2.7kb of cDNA2 contains an ATG start codon that conforms to the Kozak consensus (Kozak 1997), followed by all 18 zinc finger motifs which comprise the 1.1kb open reading frame, the 3' untranslated region, a polyadenylation sequence and a polyA tail, making it the most suitable of all the cDNA to use in this analysis.

The coupled rabbit reticulocyte lysate system; used for the *in vitro* transcription was an adaptation of the plasmid cDNA-directed protein synthesis coupled transcription-translation system described by Jackson and co-workers (1992) namely the TNT (Transcription 'n' Translation) coupled reticulocyte lysate system (Promega). This is a single tube system that enables the transcription of gene sequences cloned downstream of SP6, T3 or T7 RNA polymerase promoters, and translation by using rabbit reticulocyte lysate.
4.10.1. Methods

Firstly PCR of cDNA2 plasmid was used to determine which of the vector (pBluescript) RNA polymerase promoters (T3, SP6) was suitable for in vitro transcription. The PCR results showed that the promoter T3 was upstream of the cDNA2 ATG start codon.

The RNA polymerase T3 was therefore used to transcribe cDNA2. Two reactions were performed: (a) using the standard conditions and, (b) with the addition of 500μM ZnSO₄. Zinc is structurally important in the zinc finger protein, and its addition has been shown to be beneficial in the in vitro translation of zinc finger proteins (Choo and Klug, 1994). A T3 luciferase positive control plasmid (Promega) and a dH₂O negative control were also tested. A second group of locally available controls was also included (kindly provided by Dr. P. DeZoysa). These plasmids use the RNA polymerase T7: pKAL and pGAL (deZoysa et al., 1998).

4.10.1.1. TNT reaction conditions

Stock TNT rabbit lysate stored at -70°C was quickly thawed on ice and immediately aliquoted (5μl each) into reaction tubes containing: 1X TNT reaction buffer, 1,000 Ci/mmol ³⁵S-methionine, 1mM amino acid mixture (minus methionine), 1U RNA polymerase (T3 or T7) and 1U RNase Guard (GibcoBRL)
in a total volume of 4μl. 1μl (approximately 200ng) QiaQuick column-purified plasmid DNA was added to the TNT rabbit lysate mixture to start the reaction. The tube was incubated for 2 hours at 30°C then placed on ice.

4.10.1.2. Denaturing gel analysis of translation products

2μl aliquots of the TNT reactions and 2μl protein molecular weight markers ([methyl-14C] methylated NEC-811 (NEN)) were separated on two 12% SDS-polyacrylamide gels containing a 4% acrylamide stacking gel at 60V for 5 hours. The two identically loaded polyacrylamide gels were washed and vacuum dried at 80°C for 2 hours, and exposed to Kodak BioMAX XR x-ray film at -70°C overnight, and developed.

4.10.2. Results and Discussion

The results of the TNT analysis are presented in Figure. 4.10. The T3 luciferase positive control gave a strong signal at the correct position 61kDa. The two negative dH2O controls gave no products. However no protein bands were observed in either the standard or zinc addition lanes of cDNA2. The predicted protein pi and size for cDNA2 using the e-mail SwissProt program was 8.83 pi / 42 kDa.
The lack of observable translation could be due to a weak signal from the protein product, as the predicted protein sequence for cDNA2 contains only 4 methionine residues, which could be labelled by $^{35}$S. The T7 pKAL control however was detected using the same protocol in later experiments, the protein product of this plasmid also contains 4 methionine residues. Another possible reason for the lack of signal is the need for optimisation in the addition of ZnSO$_4$.

The concentration of ZnSO$_4$ was previously optimized by Choo and Klug 1994 in *in vitro* transcription. A weaker signal was observed from the non ZnSO$_4$ supplemented *in vitro* translations. However, the presence of start and inframe stop codons up stream of the putative start ATG of the 1.1kb open reading frame may be the major factor preventing the in vitro translation of ZNF204 (Kozak, 1996).
Figure 4.10. Results of the TNT analysis of ZNF204.

Lane 1, cDNA2 / T3; Lane 2, cDNA2 / T3 plus 500µM ZnSO₄; Lane 3 ¹⁴C methylated molecular weight marker NEC811; Lane 4, Positive control luciferase / T3; Lane 5, Empty; Lane 6, Negative control dH₂O / T3. Although translation of the positive control was demonstrated, no signal was detected for cDNA2 with or without ZnSO₄ supplementation.
4.11. ZNF 204 locus discussion

The identification of 4 different ZNF204 cDNAs from a small intestinal cDNA library confirms that the ZNF204 locus is transcriptionally active. Analysis of the complete cDNA sequence for ZNF204 shows homology to the C2H2 class of zinc finger transcription factors. The amino acid structural motif of zinc finger transcription factors is C-X$_{2,4}$-C-X$_3$-F-X$_5$-L-X$_2$-H-X$_{2,5}$-H; (Klug and Rhodes, 1987) and is highly conserved as is this intervening “linker” sequence TGEKPY. The complete sequence of the ZNF204 consensus cDNA reveals the presence of a premature stop codon upstream of zinc finger motifs 14-18 and a frameshift mutation within zinc finger motif 17 (Figure. 4.2).

Sequence analysis of both genomic and cDNA clones reveals that 10 of the 18 zinc finger motifs conform fully to the highly conserved zinc finger consensus (Klug and Rhodes, 1987). The remaining 8 motifs however contain mutations, which predict amino acid substitutions at the conserved cysteine and histidine residues that bind the zinc atom (Table 4.1). In drosophila and yeast, mutations at these conserved residues have been shown to eliminate the function of the Kruppel and ARD1 zinc finger proteins, respectively (Blumberg et al., 1987; Redemann et al., 1988). The amplification of the genomic ZNF204 locus in 3 overlapping segments using cDNA primers revealed that the 2.7kb of the ZNF204 locus contains no intronic sequence.
Together, these data suggest that this locus is most likely an expressed (known to make RNA as isolated from a cDNA library), processed (no introns detected), pseudogene. Such loci are thought to arise by integration into chromosomes of a natural complementary DNA sequence produced by reverse transcriptase from an RNA transcript. The processed appearance of this pseudogene is consistent with a reverse transcribed mRNA intermediate rather than a gene duplication event.

Consistent with this cDNA being non functional is the presence of potential ATG start codons and in frame termination codons within the putative 5’ UTR, often seen in non functional loci (Kozak, 1996). Processed pseudogenes, by the nature of their retrotransposition, typically have the remnants of their polyA tails represented in genomic DNA at their 3’ termini. The first polyadenylation signal in the 3’ genomic region of ZNF204 has no detectable remnant of a polyA tail. However, there is a second putative polyadenylation signal 181bp down stream in genomic DNA that appears to have the remains of a polyA tail. The second polyadenylation signal is not the classical AATAAA but ATAAAA. This non classical polyadenylation signal sequence has been shown to retain a reduced functionality in vitro (Sheets et al., 1990). The parent mRNA of ZNF204 may have therefore been alternatively polyadenylated and 176bp longer than the longest ZNF204 cDNA detected. The wide band observed on the northern blot at 3.2kb could be consistent with two transcripts 176bp apart.
Fluorescent in situ hybridisation and monochromosomal somatic cell hybrid panel analysis only detected ZNF204 on chromosome 6p21.3. A possible explanation for this apparent lack of a “parent” locus could be the technical limitations of these methods in the detection of an intron-containing homologue. However, several other explanations are possible. The original locus may no longer be present in the human genome. The resolution of these methods (approximately 1Mb for FISH) is such that the “parent” locus may also be in 6p21.3. This chromosomal band does indeed appear to contain a cluster of zinc finger genes, six having been isolated from the region 6p21.3: SRE-BP, ZNF165, ZNF193, ZNF192, ZNF184, (Goldwurm et al., 1997; Lee et al., 1997; Tirosvoutis et al., 1995).

The results of the homology analysis performed using the BLASTN program (Altschul et al., 1994) revealed the highest homology to ZNF135: 66% nucleic acid identity over 117 nucleotides. However reanalysis of the highest 10 homologies using the BESTFIT (GCG) program showed that ZNF184 (second highest homology with BLASTN) had the closest homology to ZNF204 with 68% nucleic acid identity over 1963 nucleotides. The zinc finger protein ZNF184 is also from the 6p21.3 region, and has been mapped to within 500kb of ZNF204 (Goldwurm et al., 1997; Wallace et al 1998). When only the putative 5’ UTR of ZNF204 was searched for homology, ZNF184 was again detected. This homology was however to the predicted coding “spacer” domain of ZNF184 suggesting that the 5’ region of ZNF204 may have been derived from coding
sequence. Taking the mapping and homology analysis together, it seems likely that the putative "parent" locus is in 6p21.3, and possibly related to ZNF184. Indeed, ZNF204 and ZNF184 both hybridise to a 3.2kb transcript on northern analysis, which is consistent if they shared their ancestry.

In conclusion, the ZNF204 locus on chromosome 6p21.3 is most likely an expressed processed pseudogene that arose by a local retrotransposition of a gene within the 6p21.3 ZNF184 cluster. Its conservation across species suggests that it could be used to isolate novel ZNF genes from other species.
Chapter 5

Ferric reductase functional studies
Chapter 5. Ferric reductase functional studies: Index.

5.1. INTRODUCTION ................................................................................150
5.2. PATIENTS AND METHODS .............................................................153
  5.2.1. Patients ...........................................................................................153
  5.2.2. Mutation analysis ............................................................................153
  5.2.3. Isolation of peripheral blood monocytes and differentiation to
         macrophages ............................................................................................154
  5.2.4. Ferric reductase assay ...................................................................158
5.3. RESULTS ...........................................................................................159
5.4. DISCUSSION .....................................................................................164
Chapter 5. Ferric Reductase Functional Studies

5.1. Introduction

Iron is of vital importance in all living organisms, being an essential co-factor of enzymes involved in respiration, cell replication and electron shuttling. The homeostasis of iron is maintained primarily by its absorption in the duodenum and the proximal region of the jejunum (McCance and Widdowson, 1937). Normally the daily loss of iron (1-2mg) is matched by absorption from the diet. This fine balance between iron excretion and absorption is lost in GH. This autosomal recessive disease leads to increased absorption of iron from the diet despite increasing body iron stores. Iron is deposited in the parenchymal tissue of the liver, pancreas, heart and other tissues; if untreated, the iron overload eventually leads to tissue damage. The prevalence of GH in Europeans is approximately 1 in 300, which is higher than the combined prevalences of cystic fibrosis, phenylketonuria and Duchenne muscular dystrophy (Bothwell et al., 1995; Dooley et al., 1997; Witte et al., 1996).

Abnormalities of iron metabolism in the intestine, liver and macrophages continue to be implicated in the pathogenesis of GH (Powell et al., 1990; Cox 1990; Olynyk 1995). In the haemochromatotic macrophage, despite the elevation of serum iron and transferrin saturation, concentrations of iron are minimal until late in the disease (Yam et al., 1968). This is in
direct contrast to transfusional siderosis where there is progressive accumulation of iron by the macrophage. This difference exists despite similar increases in serum ferritin levels (Saab et al., 1986). Cultured GH monocytes also have elevated levels of ferritin release, and abnormal kinetics of iron release, compared to control monocytes (Flanagan et al., 1989; Fillet et al., 1989; Moura et al., 1998).

The uptake of “free” iron provided in the form of ferric citrate by mononuclear cells has been established (Parmley et al., 1981; Bonkowsky et al., 1979). Both “free” iron uptake and ferric reductase activity have been shown to be increased in duodenal biopsies from haemochromatotics as compared to controls (Cox and Peters, 1978; Raja et al., 1996). More recent work has confirmed these findings in Hfe knockout mice (Simpson et al., 2003(a); Simpson et al., 2003(b)).

Ferric reductase activity has been described in both prokaryotic and eukaryotic cells, including E.coli, S.cerevisiae, mouse, rat, rabbit and human duodenum, rat liver endosomes, rabbit reticulocytes and various cultured cell lines. Its conservation across species indicates a functional role. The S. cerevisiae gene for ferric reductase (FRE1) has been cloned and sequenced (Dancis et al., 1994; Dancis et al., 1992). The ferric iron uptake mechanism in the yeast, S. cerevisiae, utilises the ferric reductase enzyme to reduce Fe(III) before absorption. The ferrous Fe(II) iron is
transported through the membrane, and is thought to be re-oxidised back to Fe(III) as it passes through the membrane.

The human homologue has now been cloned. Partial purification of human ferric reductase from both human Hutu 80 duodenal adenocarcinoma cells and human duodenal microvillus membranes has shown that the activity is NADH-dependent and the protein is membrane bound (Riedel et al., 1995).

In this chapter, ferric reductase activity is demonstrated in the K562 human erythroleukaemic cell line and primary cultures of human leukocytes, monocytes and macrophages. Ferric reductase activity is investigated in cells from GH patients compared to controls, under conditions of high “free” iron and during monocyte-macrophage differentiation.
5.2. PATIENTS AND METHODS

5.2.1. Patients

Ten male patients were investigated: five GH and five control patients with no family history of iron disorders. The diagnosis of GH was made from clinical and biochemical data and confirmed by a hepatic iron index >1.9 and/or removal of >5g iron during initial phlebotomies (Dooley et al., 1997; Worwood et al., 1997; Summers et al., 1990). The groups were age matched: GH, 49±17.4 y, control, 55±5.3 y, (mean±S.D.).

5.2.2. Mutation analysis

The Cys282Tyr and His63Asp mutations of the haemochromatosis gene HFE were determined by PCR and restriction enzyme digestion (Worwood et al., 1997). All haemochromatosis patients studied were homozygous for the Cys282Tyr mutation and negative for His63Asp. Two controls carried a single copy of His63Asp; all controls were negative for Cys282Tyr.
5.2.3. Isolation of peripheral blood monocytes and differentiation to macrophages

Peripheral blood was collected into 4X 10ml heparinized tubes. White blood cells were isolated from peripheral blood as mononuclear cells by density gradient centrifugation using Lymphoprep™ (Nycomed) medium according to the manufacturers protocol. Monocytes were then separated from white blood cell mass using their adherence to plastic surfaces. This was achieved by incubating the cells in plastic tissue culture dish for 1 hour. After washing the dish, adherent monocytes were released by scraping with 4°C PBS. Monocytes were then counted, and dispensed into wells of a 96-well microtitre plate. Each well contained $0.2 \times 10^8$ cells in a volume of 200µl. The cells were incubated at 37°C, 5% CO₂ for 14 days to fully differentiate into macrophages. Growth medium containing 20% human AB negative serum (Sigma), 100 units/ml penicillin and 100µg/ml streptomycin, 2mM L-glutamine in RPMI 1640 (ICN) was replaced on day 7 (Tormey et al., 1997)

Cell numbers were determined both using a haemocytometer and by fluorescence activated cell sorter (FACs) analysis (Coulter EpicXL). Cell viability was confirmed by trypan blue exclusion (>95%). The purity of the primary cell cultures was assessed by both morphological criteria (Leishmanns stain) and CD14 FACs analysis; an example can be seen in Figure 5.1. It shows before and after separation via plastic adherence.
Typical purities were 80% for monocytes and 95% for macrophages the other 5-20% being contamination with neutrophils, eosinophils, basophils and lymphocytes. The clonal erythroleukaemic cell line K562 was included in all experiments as an internal control. Post experimental macrophage numbers were calculated from DNA concentration using the assumption that each cell contains 6pg of DNA. DNA concentrations were measured spectrophotometrically at 600nm using the Burton assay methodology (Burton 1956). A standard curve of calf thymus DNA (Sigma) was run in parallel to all macrophage DNA. The cell number was then used to calculate ferric reductase activity as nmol Fe/min/10⁶ cells.

Cells were cultured at 37°C either in conditions of “normalised iron” (24μM iron, transferrin saturation 30%), or in normalised medium supplemented with “100μM ferric citrate”, to reflect the elevated in vivo concentrations of serum iron and “free” iron of a heavily iron loaded haemochromatotic (Batey et al., 1980; Gutteridge et al., 1985). The “100μM ferric citrate” solution was freshly prepared as a mixture of 100μM ferric chloride to 150μM sodium citrate, as previously described (Neumannova et al., 1995). Cell cultures supplemented with ferric citrate were protected from light; fresh medium was substituted after 7 days. In addition to the protection from light and chelation by citrate, stability of iron (III) was also conferred by the RPMI 1640 culture medium itself, which contains many other chelating agents at high concentration. These include amino acids, sugars, vitamins, bicarbonate and phosphate, which
also prevent the hydrolytic polymerization of iron (Richardson and Baker, 1991). The amino acids and glucose alone provide an additional 150-fold molar excess of chelating agent to iron (Richardson and Baker, 1991).
Figure 5.1. CD14 FACS analysis showing macrophage purity before and after plastic adherence purification.

FS (forward scatter) log is proportional to the size and SS (side scatter) log is proportional to granularity. a) White blood cells before plastic adherence purification. The monocytes (circled) are contaminated with vesicles, debris and other cell types, which are largely removed after plastic adherence, b).
5.2.4. Ferric reductase assay

Cells were harvested, rinsed, then incubated at a concentration of 2 $\times 10^6$ cells/ml in 200$\mu$l oxygenated physiological buffer (125mM NaCl, 3.5mM KCl, 1mM CaCl$_2$, 10mM MgSO$_4$, 10mM D-glucose in 16mM Hepes/NaOH, pH 7.4) in 96-well tissue culture plates at 37°C. The reaction was started by the addition of 100$\mu$M Fe(III), as a ferric chelate of nitrilotriacetate (NTA) in a 1:2 ratio (Fe:NTA), and ferrozine (1mM). The reaction was performed at 37°C. The rate of reaction was measured spectrophotometrically at 562nm to detect the reduction of the yellow Fe(III) to the purple coloured stable Fe(II)-ferrozine complex (Raja et al., 1996). All controls including analysis of the K562 cell line were performed in parallel with every primary cell culture monocyte and macrophage assay to confirm uniformity of results.
5.3. RESULTS

Preliminary experiments investigated the characteristics of ferric reductase activity of the clonal control K562 erythroleukaemic cell line. In each experiment the rate of ferric reductase activity was determined by linear regression analysis of at least three time courses; each individual time point was performed in duplicate. All time courses were linear for at least 60 minutes (Figures 5.2. and 5.3.). After mild trypsinisation, ferric reductase activity was reduced to 38.7% of control (Table 5.1.). Heating the cells to 65°C for 5min prior to assay reduced activity to 26.7% of control. When the assay was performed at 10°C, 17% of control activity was observed. These observations are consistent with assay of a membrane bound protein. When K562 and macrophage cell-free supernatant was taken for reductase activity at the end of the incubation period, only background levels of reductase activity were detected (Table 5.1.). This indicated that the activity was not due to a soluble, released factor. These observations indicate that ferric reductase is a membrane bound enzyme.

Investigation of primary cultures of lymphocytes, monocytes and macrophages demonstrated ferric reductase activity in these blood-
derived cells (Table 5.2.). No significant difference (p>0.05) was observed between GH (Cys282Tyr homozygous) and control (Cys282Tyr negative) in either lymphocyte, monocyte or macrophage preparations.

Figure 5.2. Monocyte ferric reductase activity: GH versus controls.

Linear regression analysis of ferric reductase data (● = GH, ■ = control), The error bar = 1 S.D. There was no significant difference between GH and control groups (n=5; p>0.05 unpaired student t-test).
Figure 5.3. Macrophage ferric reductase activity: GH versus controls.

Linear regression analysis of ferric reductase data (● = GH, ■ = control), The error bar = 1 S.D. There was no significant difference GH and control groups (n=5; p>0.05 unpaired student t-test).
Table 5.1. Human ferric reductase activity has the characteristics of a membrane bound enzyme.

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Treatment</th>
<th>Ferric reductase activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K562</td>
<td>none</td>
<td>100</td>
</tr>
<tr>
<td>K562</td>
<td>0.1mg/ml trypsin, 37°C, 60min</td>
<td>38.7 ± 1.5**</td>
</tr>
<tr>
<td>K562</td>
<td>65°C, 5min</td>
<td>26.7 ± 1.5**</td>
</tr>
<tr>
<td>K562</td>
<td>assay performed at 10°C</td>
<td>17.0 ± 1.7**</td>
</tr>
<tr>
<td>K562</td>
<td>supernatant activityχ</td>
<td>0.33 ± 0.58**</td>
</tr>
<tr>
<td>macrophage</td>
<td>supernatant activityχ</td>
<td>0.02 ± 0.02**</td>
</tr>
</tbody>
</table>

The cells were subjected to various treatments and then ferric reductase activity was measured and expressed as a percentage of K562 control activity at 37°C (mean ± standard deviation; n = 3). For the control K562 cells, the ferric reductase activity was 6.93nmol Fe/min/10^9 cells, where 1μM ferric iron gives an absorbance of ~0.03 at 562nm. The reaction was performed in the absence of ferrozine, cells were removed by centrifugation, and supernatant activity determined after the addition of ferrozine. Activities are compared with control: ** p < 0.01 (Student's t-test).
Table 5.2. Ferric reductase activities. Activities are shown for lymphocytes, monocytes and macrophages from classical C282Y homozygous GH patients and Cys282Tyr negative control patients.

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Ferric reductase activity (nmol Fe/min/10⁹ cells)</th>
<th>Fe</th>
<th>Mean</th>
<th>S.D.</th>
<th>n</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>lymphocytes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>4.37</td>
<td>N</td>
<td>3.13</td>
<td>3.13</td>
<td>3</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>control</td>
<td>3.98</td>
<td>N</td>
<td>2.51</td>
<td>2.51</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>monocytes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>4.38</td>
<td>N</td>
<td>1.71</td>
<td>1.71</td>
<td>3</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>control</td>
<td>3.75</td>
<td>N</td>
<td>0.71</td>
<td>0.71</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>macrophages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>43.3</td>
<td>N</td>
<td>5.13</td>
<td>5.13</td>
<td>4</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>control</td>
<td>31.3</td>
<td>N</td>
<td>8.67</td>
<td>8.67</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>cell line</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K562</td>
<td>6.93</td>
<td>N</td>
<td>0.88</td>
<td>0.88</td>
<td>3</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>K562 100µM</td>
<td>7.87</td>
<td>100µM</td>
<td>0.77</td>
<td>0.77</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>macrophages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GH 100µM</td>
<td>10.0</td>
<td>100µM</td>
<td>4.87</td>
<td>4.87</td>
<td>5</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>control 100µM</td>
<td>9.2</td>
<td>100µM</td>
<td>9.60</td>
<td>9.60</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Ferric reductase activities for GH and control cultured in N (normalised iron medium); 100µM, iron loaded (additional 100µM ferric citrate in normalized iron medium. p = (Students t-test).
5.4. DISCUSSION

The ferric reductase activity of human K562 cells was demonstrated to have the characteristics of a membrane bound enzyme and was not attributable to the secretion of reducing factors. There was no significant difference in the reductase activity between GH and controls for leukocyte, monocyte or macrophage preparations. This was observed under both normalised iron conditions and in the presence of 100μM ferric citrate. Ferric reductase activity of differentiated macrophages was ~900% that of monocyte activity (p<0.05). This increase most likely reflects the co-ordinated upregulation of proteins of iron metabolism during the transition into macrophages (Cottrell and Jones, 1979).

Macrophage ferric reductase activity was decreased when cultured in 100μM ferric citrate compared to that in normalised iron. From these results it would appear that macrophage ferric reductase activity may be determined by external "free" iron concentrations during macrophage differentiation. Macrophage ferric reductase activity was down regulated by increased "free" iron concentration.

This control of ferric reductase activity could contribute to the low levels of iron observed in the haemochromatotic macrophage. As haemochromatotic macrophages differentiate in high "free" iron conditions in vivo, their ferric reductase activities may be decreased according to the amount of free iron present. Future studies could
address this issue by isolation of differentiated macrophages from GH and control human subjects, to determine the activity of the *in vivo* differentiated cells.

The clonal reference cell line K562 did not show modulation of ferric reductase activity by iron. The K562 cell line originates from an erythroid cell clone derived from a patient with myeloid leukaemia in acute blast crisis (Andersson and Gahmberg 1981). The lack of modulation of K562 ferric reductase activity by iron may reflect either the erythroid lineage or changes in the clonal expansion of this cell line. The identification of the human ferric reductase gene will ultimately establish whether it is transcriptionally repressed by iron, as in *S. cerevisiae*. Future studies will elucidate whether ferric reductase plays a role in the apparent iron deficiency of macrophages.
Chapter 6. Conclusions and Perspectives

This thesis reports two approaches initially directed towards the mapping and identification of the chromosome 6-linked haemochromatosis gene. At the beginning of this thesis, twenty years had elapsed since initial observations of association of haemochromatosis with particular HLA alleles. A positional cloning approach was used to identify potential candidate genes from the chromosome 6 gene region. A functional cloning approach was investigated by analysis of the first reported potential index of disease in vitro.

Before the discovery of the \textit{HFE} gene, chromosome 6-linked haemochromatosis was investigated first by analyses of HLA serotypes and then later by analysis of DNA markers from the HLA region. The availability of highly polymorphic DNA microsatellite markers was an important step in the mapping and the ultimate identification of \textit{HFE}. Four polymorphic microsatellite markers spanning the 6p21.3 'founder haplotype' were originally used in this thesis to characterise the patient and control groups. Some of the patients were homozygous for several of these markers, providing strong indirect evidence for chromosome 6-linked haemochromatosis. Later \textit{HFE} mutation analysis confirmed that most of the patients were homozygous for the Cys282Tyr mutation and that none of the control group had disease related genotypes. Because the link between \textit{HFE} and iron metabolism was not apparent
immediately, it remained a 'candidate' gene for some time. However, HFE mutation analysis is now widely used for confirmation of the diagnosis of haemochromatosis and in family screening and has reduced the need for liver biopsy.

Due to the pressure of commercial gene hunters also searching for the haemochromatosis gene, the mapping information and clones available at that time were used in an attempt to identify candidate genes from around the apparent peak of linkage disequilibrium (D6S1260). This was done by direct hybridisation of genomic PAC clones to a human intestinal cDNA library. This methodology proved effective by the identification of four single copy cDNAs for the zinc finger locus ZNF204.

The gene locus was shown to be intronless and no 5' end could be identified by genomic analysis or by 5'-RACE. Sequence analysis indicated historic mutations that would have inactivated individual zinc finger protein fingers. In vitro coupled transcription and translation studies were unable to demonstrate translation of ZNF204. The locus was therefore considered likely to be an expressed processed pseudogene and unrelated to haemochromatosis.

If the parent locus could be discovered, then the point at which this pseudogene was created could be estimated by looking at the mutation rate compared to the original sequence. However, as yet computer
database screening, somatic cell hybrid panel analysis and FISH has not alluded to a parent locus for ZNF204.

Zinc finger proteins have a potential future role in the manipulation of gene expression via targeted promoter methylation. This targeting can be achieved by creating zinc finger proteins that bind to gene-specific promoter DNA sequences (Latchman, 1995). The fusion of these site-specific zinc finger proteins to methyltransferase results in the targeted methylation of CpG islands upstream of the zinc finger binding site (Xu and Bestor, 1997). An advantage of using this methodology for the selective silencing of genes is that even after short exposure, the cells own methylating pathways will maintain this new pattern of methylation. The action of other regulatory methods such as antisense nucleic acids, ribozymes and other analogous agents all dissipate after time and require constant exposure. The future therapeutic role of zinc finger proteins is still waiting to be fully exploited.

The study of ferric reductase activity in macrophages was conducted in an attempt to identify an in vitro index of the disease. This would allow investigation of potential candidate haemochromatosis genes by functional complementation. Previous studies on duodenal biopsies had shown increased ferric reductase activity in GH compared with controls (Raja et al., 1996). In GH, several lines of evidence indicate that macrophages (like duodenal enterocytes) remain paradoxically iron poor
despite iron overload (Powell et al., 1990; Cox 1990; Olynyk 1995). Macrophages were therefore considered likely to express any altered functional index. Ferric reductase activity of K562 cells was demonstrated to have properties consistent with a membrane bound protein. In contrast to earlier criticisms, the activity was not attributable to a secreted reducing factor. Ferric reductase activity was demonstrated for the first time in white blood cells (leukocyte, monocyte and macrophage preparations), but there was no significant difference between GH and control cultures. Macrophage ferric reductase activity was decreased following culture under iron loaded conditions, although the mechanism of this iron response was not known.

The gene Dcytb has been cloned since these studies were completed and confirmed as the mammalian plasma membrane ferric reductase that catalyzes the reduction of ferric to ferrous iron as part of the process of iron absorption (McKie et al., 2000). The cloning was via a subtractive cloning strategy designed to identify intestinal genes involved in iron absorption. One of the cDNAs obtained by this strategy had a sequence homology to a plasma membrane di-haem protein, and was named duodenal cytochrome b (Dcytb) and shared 40-50% homology to cytochrome b561. Immunohistochemical staining for the protein localized it to the duodenal brush border membrane (McKie et al., 2000), although expression was also found in other tissues. Expression of Dcytb in Xenopus oocytes and cultured mammalian cells led to induction of ferric
reductase activity, providing confirmation of function. The level of Dcytb mRNA and protein expression in duodenum was shown to be highly upregulated by three conditions that stimulate iron absorption: chronic anaemia (hypotransferrinaemic mouse), iron deficiency and hypoxia, consistent with the proposed role of ferric reductase in the pathway of iron absorption. It was postulated that DcytB may utilize cytoplasmic ascorbate as an electron donor for the transmembrane reduction of iron, prior to transport into the cell by DMT1 (McKie et al., 2000). This proposal was supported by the demonstration that, in cells transfected with Dcytb, iron and iron chelators modulate the cellular ascorbate levels (Latunde-Dada et al., 2002).

Ascorbate-stimulated reduction of both iron(III), and interestingly copper(II), was demonstrated by a putative rabbit homologue of Dcytb, purified from duodenal brush border membranes (Knopfel and Solioz, 2002). If there is a relationship between Dcytb and copper, this is not clear at present. Han and Wessling-Resnick (2002) were not able to identify changes in the expression of Dcytb mRNA in Caco-2 cells in response to copper supplementation. However, the measurement of apical ferric reductase activity did show modulation via copper status. It was proposed that multiple ferric reductases other than Dcytb may exist, which may have different specificities and distribution (Han and Wessling-Resnick, 2002).
Unlike some proteins of iron metabolism, Dcytb has no identifiable iron responsive elements for the direct modification of expression in response to iron. Therefore, the sensitivity of macrophage ferric reductase activity to iron loading, reported in this thesis, cannot be attributed to an iron response element-mediated mechanism, and the basis of this regulation remains to be resolved. For example, it may be that the ferric reductase gene contains novel sequences which confer regulation by iron; alternatively, the response to iron may be secondary and be mediated via the products of other genes. It has been proposed, by analogy with the regulation of yeast ferric reductase enzymes FRE1 and FRE2 by the iron-inducible transcriptional regulator, Aft1p, that mammalian ferric reductase activity may be regulated by an as yet unidentified, iron-responsive transcription factor (Latune-Dada et al., 2002).

Dcytb expression has been demonstrated in several cell types other than the duodenal enterocyte, including spleen, liver, brain and cultured cell lines. The finding of high levels in the neutrophil suggested the possibility of a role in host defence. Ferrous iron, from Dcytb activity, may react with superoxide and H$_2$O$_2$ (produced by the respiratory burst oxidase) via Fenton chemistry, to produce a highly reactive hydroxyl radical. The potential role of Dcytb in host defence remains to be explored (McKie et al., 2002).
Ferric reductase has been considered as a potential genetic "modifier" of the rate of iron absorption. In mice, significant differences have been observed between different strains of mice in their predisposition towards iron loading, with DBA/2 mice being more susceptible to iron loading than C57BL/6. Recent quantitative RT-PCR studies have shown that knockout mice on the DBA/2 background have increased levels of duodenal ferric reductase, DMT1 and Ireg1 mRNA, as compared to wild type controls. In comparison, the HFE knockout mice on the C57BL/6 background had no such detectable increase in expression of these genes (Dupic et al., 2002a). Recent studies of duodenal ferric reductase activity in HFE knockout mice (mixed background stain "129yOla - C57Bly6") confirmed a modest increase in nitroblue tetrazolium staining in comparison to wild type mice, in agreement with the earlier study (Simpson et al., 2003). A more detailed analysis of four different mouse strains, C57BL/6, DBA/2, CBA and 129/Sv, showed strain to strain differences not only in serum transferrin saturation and hepatic iron stores, but also in the duodenal mRNA expression of several key genes of iron metabolism: Dcytb, DMT1, Ireg1, and transferrin receptor 1. This study supports some degree of genetic control of the mRNA levels of these molecules (Dupic et al., 2002b)

These mouse studies clearly suggested that functional polymorphisms of duodenal ferric reductase, DMT1 and Ireg1 may be responsible for individual variations in the clinical penetrance of human
haemochromatosis. Lee et al., (2002) investigated a total of 26 potential modifier genes, including Dcytb, DMT1 and Ireg1, by sequencing the coding regions, exon-intron junctions and promoters in DNA from 20 subjects with different HFE genotypes (homozygous C282Y or wild type) and different degrees of clinical penetrance / iron load. However, only one transferrin mutation showed a relationship with iron deficiency anaemia. There was no evidence that polymorphisms of ferric reductase, or any of the other candidate modifier genes, influence the clinical penetrance of haemochromatosis. The putative genes influencing the expression of HFE mutations remain unclear (Lee et al., 2002).

Frazer et al., (2002) investigated the role of the antimicrobial and iron-regulatory peptide hepcidin in relation to iron absorption and hepatic and duodenal gene expression in rats after a switch to an iron deficient diet. They reported that iron absorption increased within 6 days of starting the iron-deficient diet. This was accompanied by increased duodenal expression of the ferric reductase Dcytb, DMT1 and Ireg1, demonstrated by RNAse protection, immunofluorescence and western blotting. These changes correlated with changes in hepatic hepcidin expression and transferrin saturation. There was no alteration in iron stores or haematologic parameters over this period. Frazer et al (2002) proposed a model of the role of hepcidin in intestinal iron absorption based on these observations, in which ferric reductase forms part of the regulatory cycle (Figure 6.1).
Figure 6.1. A model showing the proposed role of hepcidin in intestinal iron absorption (from Frazer et al., 2002). This model incorporates ferric reductase (Dcytb) as a component of this regulatory "loop".

The amount of iron absorbed in the intestine determines the percentage saturation of circulating transferrin with iron. The liver is proposed to detect variation in transferrin saturation, possibly mediated by transferrin receptor 2 (TfR2) and / or the HFE - transferrin receptor 1 (TfR1) complex. This results in modulation of hepatocyte expression of hepcidin: when transferrin saturation decreases, hepcidin production is decreased; when transferrin saturation increases, hepcidin production increases.

Circulating hepcidin levels determine the level of duodenal iron absorption by regulating the expression of Ireg1, DMT1 and Dcytb. It is not known if this is achieved by programming the crypt cells or by direct action on the mature enterocytes. This model permits the amount of iron absorbed to be modulated (via transferrin saturation and hepcidin levels) before body iron stores are altered.
Recently, Muckenthaler et al., (2003) investigated Cybrd (mouse ferric reductase) mRNA expression along with other important genes in iron metabolism using a custom microarray 'Iron Chip'. The mouse types analysed for duodenal expression were HFE knock-out, HFE transgenic mice homozygous for the murine equivalent of the human C282Y mutation, secondary iron overload and iron deficient. The results were interpreted to show that firstly, the duodenal gene expression patterns in the mouse models of haemochromatosis were more similar to secondary iron overload than to iron deficiency. This is controversial, as it is opposite to many lines of evidence suggesting that in haemochromatosis, the duodenal enterocyte is paradoxically iron-deficient in relation to body iron stores. Models of haemochromatosis pathogenesis to date have generally incorporated this concept of a paradoxically iron-deficient duodenum, so that challenging this concept may have far-reaching implications. Secondly, Muckenthaler et al, (2003) observed that Cybrd (ferric reductase) mRNA was an exception to this pattern, being significantly upregulated in the mouse models of haemochromatosis as well as in the iron deficient (rather than iron overloaded) duodenum. Parallel studies of liver expression patterns identified that only hepcidin differed between secondary iron overload in wild type mice (increased hepcidin) versus secondary iron overload in HFE knock-out or "C282Y" transgenic mice (decreased hepcidin levels). The interpretation of this was that hepcidin expression in response to iron overload is HFE-dependent. It was proposed that, in mouse models of
haemochromatosis, decreased hepatic hepcidin levels may lead to increased duodenal ferric reductase levels, contributing to increased iron absorption. As well as challenging the long-held concept of "paradoxical iron deficiency" of the duodenal enterocyte in haemochromatosis, these results also suggest that ferric reductase activity may be directly involved in the increased iron absorption of haemochromatosis. However, these experiments due to methodology are open to interpretation. It is important that the microarray method is a quantitative indication of gene expression. Although some evidence is presented that microarray expression levels vary in approximate accordance with expression data from Northern blot or quantitative PCR methods, this comparison is only provided for three genes. Expression was studied in pools rather than replicate experiments, which may confound the analysis. There are also significant omissions, such as the transferrin receptor, which does not appear in this analysis. Most importantly, the argument that mouse models of haemochromatosis have duodenal expression patterns similar to secondary iron overload, rather than iron deficiency, relies heavily on the expression of genes which are peripheral to the pathway of iron absorption, and so the observations of Muckenthaler et al., (2003) are as yet of unclear significance.

The results of Muckenthaler and colleagues are however in contrast to those results obtained by Stuart et al., (2003) using an RNAse protection assay approach to look at Dcytb expression. RNAse protection assays
have been widely accepted to give a quantitative estimate of RNA levels. Human Dcytb expression was found to be similar in iron deficient and iron replete subjects and was not increased in untreated haemochromatosis patients. In comparison with the microarray studies, duodenal expression of DMT1 (IRE-containing form) and Ireg1 was similar in haemochromatosis patients and iron replete subjects, but in haemochromatosis patients with increased serum ferritin concentrations, both DMT1 (IRE) and Ireg1 were inappropriately increased relative to serum ferritin concentration. These findings are consistent with the two transporters, DMT1(IRE) and Ireg1, playing primary roles in the adaptive response to iron deficiency (Stuart et al., 2003).

The RNAse protection methodology was also used to investigate the role of Dcytb and other genes of iron metabolism in the "mucosal block" phenomenon of iron absorption (Frazer et al., 2003). This is the ability of a large oral dose of iron to reduce the absorption of a dose administered several hours later. After rats were given an oral dose of iron, a rapid decrease in iron absorption was detected, associated with increased enterocyte iron levels and a rapid decrease in the mRNA levels of Dcytb and DMT1(IRE). No such change was seen in expression of DMT1 (non-IRE), Ireg1 or hephaestin. These data implicate the brush border iron ferric reductase (Dcytb) and transporter (DMT1(IRE)) in the rapid "mucosal block" response, probably mediated by the response to local enterocyte iron levels. It was noted that this local and rapid response is
distinct from systemic signals, from either the level of stored iron or of erythropoiesis, to the intestine. Although the mechanism whereby the intestine is signalled of these systemic requirements remains unclear, it is held that they require 2-3 days to influence iron absorption, probably reflecting the time take for the immature crypt cell to migrate up the villus, differentiate and influence iron absorption. Ferric reductase appears to be important in the rapid "mucosal block" phenomenon, where it is probably regulated by intracellular iron levels (Frazer et al., 2003). This observation is consistent with the observation reported in this thesis that macrophages cultured in high iron conditions demonstrate reduced ferric reductase compared to those grown under normalised iron conditions.

Future research on ferric reductase may identify novel applications in biotechnology, for instance in biosensor arrays. By analogy with glucose sensors used in bioreactor fermentation vessels, an oxygen probe could be coupled with ferric reductase bound to a membrane via glutaraldehyde, enabling direct detection of in vitro iron levels.

The identification of the HFE gene and subsequent reporting of DMT1, ferroportin, hephaestin, hepcidin and ferric reductase has enabled new approaches to understand the molecular and cellular mechanisms underlying iron homeostasis. The integration of these pathways and their regulation remain a challenge for future research.
References


Hanot, V. & Schachmann, M. (1886). *Archives de Physiologies Normale et Pathologique* **7**: 50-.


Marx J. (1979) Mucosal uptake transfer and retention of iron, measured by whole-body counting. *Haematol.* **23**: 293-302


*Hepatol* 24: 43-46.

*Eur J Gastroenterol Hepatol* 7: 1203-1208.

*Eur J Gastroenterol Hepatol* 7: 1203-1208.

*Gastroenterology* 114: 996-1002.

*Hepatol* 28: 1105-1109.


Pratiwi, R., Porter, L., Powell, L.W. & Anderson, G.J. (1999). The transferrin receptor is not essential for HFE synthesis or cell surface expression in CHO cells. World Congress on Iron Metabolism, Sorrento 220 [Abstract]


Appendix 1
Appendix 1

The conversion between allele size, as determined by microsatellite analysis, and allele number, as shown in the Table. (Raha-Chowdhury et al 1995, and Worwood and Raha-Chowdhury unpublished data)

<table>
<thead>
<tr>
<th>Allele No.</th>
<th>b.p</th>
<th>Allele No.</th>
<th>b.p</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6S265 (Blue)</td>
<td>D6S1621 (Green)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>119</td>
<td>5</td>
<td>292</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>6</td>
<td>290</td>
</tr>
<tr>
<td>3</td>
<td>123</td>
<td>7</td>
<td>288</td>
</tr>
<tr>
<td>4</td>
<td>125</td>
<td>8</td>
<td>286</td>
</tr>
<tr>
<td>5</td>
<td>127</td>
<td>9</td>
<td>284</td>
</tr>
<tr>
<td>6</td>
<td>129</td>
<td>10</td>
<td>282</td>
</tr>
<tr>
<td>7</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>133</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Allele No.</th>
<th>b.p</th>
<th>Allele No.</th>
<th>b.p</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6S105 (Yellow)</td>
<td>D6S1260 (Green)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>114</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>10</td>
<td>116</td>
<td>2</td>
<td>148</td>
</tr>
<tr>
<td>9</td>
<td>118</td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td>8</td>
<td>120</td>
<td>4</td>
<td>152</td>
</tr>
<tr>
<td>7</td>
<td>122</td>
<td>5</td>
<td>154</td>
</tr>
<tr>
<td>6</td>
<td>124</td>
<td>6</td>
<td>156</td>
</tr>
<tr>
<td>5</td>
<td>126</td>
<td>7</td>
<td>158</td>
</tr>
<tr>
<td>4</td>
<td>128</td>
<td>8</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>130</td>
<td>9</td>
<td>162</td>
</tr>
</tbody>
</table>

*Not seen.*
Appendix 2
Appendix 2

Publications directly related to this thesis:


INTRODUCTION

Iron is of vital importance in all living organisms, being an essential cofactor of enzymes involved in respiration, cell replication and electron shuttling. The homeostasis of iron is maintained primarily by its absorption in the duodenum and the proximal region of the jejunum [1]. Normally the daily loss of iron (1–2 mg) is matched by absorption from the diet. This fine balance between iron excretion and absorption is lost in genetic haemochromatosis (GH). This autosomal recessive disease leads to increased absorption of iron from the diet despite increased body iron stores. Iron is deposited in the parenchymal tissue of the liver, pancreas, heart and other tissues. If untreated, the iron overload eventually leads to tissue damage. The prevalence of GH in Europeans is approximately 1 in 300, which is higher than the combined prevalence of cystic fibrosis, phenylketonuria and Duchenne muscular dystrophy [2–4].

Recently a candidate gene for haemochromatosis (HFE) was identified [5]. In the United Kingdom, 91 % of GH patients were found to be homozygous for a single mutation of HFE, Cys<sup>282</sup>Tyr (Cys<sup>282</sup>→Tyr) [6]. Although the metabolic role of the HFE protein is not fully understood, recent advances support a role in the regulation of iron absorption [7–10].

Abnormalities of iron metabolism in the intestine, liver and macrophages have all been implicated in the pathogenesis of GH [11–13]. In the haemochromatotic macrophage, despite the elevation of serum iron and transferrin saturation, concentrations of iron are minimal until late in the disease [14]. This is in contrast with transfusional siderosis, where there is progressive accumulation of iron by the macrophage. This difference exists despite similar increases in serum-ferritin levels [15]. Cultured GH monocytes also have elevated levels of ferritin release, and abnormal kinetics of iron release, compared with control monocytes [16,17].

The uptake of ‘free’ iron provided in the form of ferric citrate by mononuclear cells has been established [18,19]. Both free iron uptake and ferric reductase activity have been shown to be increased in duodenal biopsies from haemochromatosis patients as compared with controls [20,21].

Ferric reductase activity has been described in both prokaryotic and eukaryotic cells, including Escherichia coli, Saccharomyces cerevisiae, mouse, rat, rabbit and human duodenum, rat liver endosomes, rabbit reticulocytes and various cultured cell lines [21–24]. Its conservation across species indicates a functional role. The S. cerevisiae gene for ferric reductase (FRE1) has been cloned and sequenced [22,23]. The ferric iron uptake mechanism in the yeast, S. cerevisiae, utilizes the ferric reductase enzyme to reduce Fe(III) before absorption. The ferrous Fe(II) iron is transported through the membrane, and is thought to be reoxidized back to Fe(III) as it passes through the membrane. The human homologue of FRE1 has yet to be cloned. However, partial purification of human ferric reductase from both human Hutu 80 duodenal adenocarcinoma cells and human duodenal microvillus membranes has shown that the activity is NADH-dependent and the protein is membrane bound [24].

We demonstrate ferric reductase activity in the K562 human erythroleukaemic cell line and primary cultures of human leucocytes, monocytes and macrophages. We report ferric...
reductase activity in cells from GH patients compared with controls, under conditions of high free iron and during monoocyte-macrophage differentiation.

MATERIALS AND METHODS

Patients
Ten male patients were investigated: five GH and five control patients with no family history of iron disorders. The diagnosis of GH was made from clinical and biochemical data and confirmed by a hepatic iron index > 1.9 and/or removal of > 5 g of iron during initial phlebotomies [3,6,25]. The groups were age matched: GH, 49 ± 17.4 years; control, 55 ± 5.3 years, (mean ± S.D.).

Mutation analysis

The Cys<sup>64</sup>Tyr and His<sup>68</sup>Asp mutations of the haemochromatosis gene, HFE, were determined by PCR and restriction enzyme digestion [6]. All haemochromatosis patients studied were homozygous for the Cys<sup>64</sup>Tyr mutation and negative for His<sup>68</sup>Asp. Two controls carried a single copy of His<sup>68</sup>Asp; all controls were negative for Cys<sup>64</sup>Tyr.

Cell culture

White blood cells were extracted from peripheral blood using Lymphoprep<sup>®</sup> (Nycomed). Monocytes were enriched by their adherence to plastic. Adherent monocytes were cultured for 14 days for in vitro differentiation to macrophages [26]. Cell numbers were determined using a haemocytometer and by fluorescence activated cell sorter analysis (Coulter EpicXL). Cell viability was confirmed by Trypan Blue exclusion (> 95%). The purity of the primary cell cultures was assessed by morphological criteria (Leishmanns stain) and CD14 fluorescence-activated cell sorter analysis; typical purities were 80% for monocytes and 95% for macrophages. The clonal erythroleukaemic cell line K562 was included in all experiments as an internal control. Cells were cultured at 37°C either in conditions of 'normalised' iron (24 μM iron, transferrin saturation 30%), or in normalised medium supplemented with 100 μM ferric citrate, to reflect the elevated iron concentrations of serum iron and free iron of a heavily iron-loaded haemochromatotic [27,28]. The 100 μM ferric citrate solution was freshly prepared as a mixture containing 100 μM ferric chloride and 150 μM sodium citrate, as described previously [29]. Cell cultures supplemented with ferric citrate were protected from light and fresh medium was substituted after 7 days. In addition to the protection from light and chelation by citrate, stability of Fe(III) was also conferred by the RPMI 1640 culture medium itself, which contains many other chelating agents at high concentrations. These include amino acids, sugars, vitamins, bicarbonate and phosphate, which also prevent the hydrolytic polymerization of iron [30]. The amino acids and glucose alone provide an additional 150-fold molar excess of chelating agent over iron [30].

Ferric reductase assay

Cells were harvested, rinsed, then incubated at a concentration of 2 × 10<sup>6</sup> cells/ml in 200 μl of oxygenated physiological buffer (125 mM NaCl/3.5 mM KCl/1 mM CaCl<sub>2</sub>/10 mM MgSO<sub>4</sub>/10 mM D-glucose in 16 mM Hepes/NaOH, pH 7.4) in 96-well tissue culture plates at 37°C. The reaction was started by the addition of 100 μM Fe(III), as a ferric chelate of nitrilotriacetate in a 1:2 ratio (Fe:nitrilotriacetate), and ferrozine (1 mM). The reaction was performed at 37°C. The rate of reaction was determined spectrophotometrically at 562 nm by measuring the reduction of the yellow Fe(II) to the purple coloured stable Fe(II)-ferrozine complex [21].

RESULTS

Preliminary experiments investigated the characteristics of the ferric reductase activity of the clonal control K562 erythro-leukaemic cell line. In each experiment the rate of ferric reductase activity was determined by linear regression analysis of at least three time courses; each individual time point was performed in duplicate. All time courses were linear for at least 60 min. After mild trypsinization, ferric reductase activity was decreased to 38.7% of the control (Table 1). Heating the cells to 65°C for 5 min prior to assay decreased activity to 26.7% of the control. When the assay was performed at 10°C, 17% of the control activity was observed. These observations were consistent with the assay of a membrane bound protein. When K562 and macrophage cell-free supernatants were taken for determining reductase activity at the end of the incubation period, only background levels of reductase activity were detected (Table 1). This indicated that the activity was not due to a soluble, released

Table 1 Human ferric reductase activity has the characteristics of a membrane bound enzyme

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Treatment</th>
<th>Ferric reductase activity (%)&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>K562</td>
<td>0.1 mg/ml trypsin, 37 °C, 60 min</td>
<td>38.7 ± 1.5&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>K562</td>
<td>65 °C, 5 min</td>
<td>26.7 ± 1.5&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>K562</td>
<td>Assay performed at 10 °C</td>
<td>17.0 ± 1.7&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>K562</td>
<td>Supernatant activity&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.33 ± 0.58&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Macrophage</td>
<td>Supernatant activity&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.02 ± 0.02&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Activities are shown for lymphocytes, monocytes and macrophages from classical Cys<sup>64</sup>Tyr homozygous GH patients and Cys<sup>64</sup>Tyr negative control patients. Statistical probabilities were assessed using the Student's t-test. There was a significant increase in activity for both GH (†) and control (‡) preparations (P<0.05 in both cases). N, normalized-iron culture medium (see the Materials and methods section); 100 μM, iron loaded (additional 100 μM ferric citrate in normalized-iron medium).

Table 2 Ferric reductase activities

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Fe</th>
<th>Ferric reductase activity (nmol of Fe·min&lt;sup&gt;-1&lt;/sup&gt;·(10&lt;sup&gt;6&lt;/sup&gt; cells)&lt;sup&gt;-1&lt;/sup&gt;)&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphocytes</td>
<td>GH</td>
<td>N</td>
</tr>
<tr>
<td>Monocytes</td>
<td>GH</td>
<td>N</td>
</tr>
<tr>
<td>Macrophages</td>
<td>GH</td>
<td>N</td>
</tr>
<tr>
<td>Cell line</td>
<td>K562</td>
<td>N</td>
</tr>
<tr>
<td>Macrophages</td>
<td>GH</td>
<td>100 μM</td>
</tr>
<tr>
<td>Macrophages</td>
<td>Control</td>
<td>100 μM</td>
</tr>
</tbody>
</table>

<sup>1</sup>Mean ± S.D., n P

- P > 0.05
- P < 0.05
factor. These observations indicate that ferric reductase is a membrane bound enzyme.

Investigation of primary cultures of lymphocytes, monocytes and macrophages demonstrated ferric reductase activity in these blood-derived cells (Table 2). No significant difference \((P < 0.05)\) was observed between \(\text{GH} (\text{Cys}^{38} \text{Tyr} \text{homozygous})\) and control \((\text{Cys}^{38} \text{Tyr} \text{negative})\) in either lymphocyte, monocyte or macrophage preparations. However, ferric reductase activity of macrophages was approx. 900\% of the monocyte activity, for both \(\text{GH}\) and control preparations, when cultured in normalized iron.

When \(\text{K562}\) cells were cultured in normal or high iron \((100 \mu \text{M ferric citrate})\) concentrations, no difference in ferric reductase activity was observed \((P > 0.05); \text{Table 2}\).

When macrophages were cultured in media containing 100 \(\mu \text{M ferric citrate}\) for 2 weeks, no significant difference \((P > 0.05)\) was observed between \(\text{GH}\) and the control. However, the ferric reductase activity of both \(\text{GH}\) and control macrophages cultured under these conditions was approx. 25\% of the activity when grown at normalized iron concentrations \((P < 0.05); \text{Table 2}\).

**DISCUSSION**

The ferric reductase activity of human \(\text{K562}\) cells was demonstrated to have the characteristics of a membrane bound enzyme and was not attributable to the secretion of reducing factors. There was no significant difference in the reductase activity between \(\text{GH}\) and controls for leucocyte, monocyte or macrophage preparations. This was observed under both normalized iron conditions and in the presence of 100 \(\mu \text{M ferric citrate}\). Ferric reductase activity of differentiated macrophages was approx. 900\% that of monocyte activity \((P < 0.05)\). This increase most likely reflects the co-ordinated up-regulation of the proteins of iron metabolism during the transition into macrophages [31]. Macrophage ferric reductase activity was decreased when cultured in 100 \(\mu \text{M ferric citrate}\) compared with that in normalized iron. From these results it would appear that macrophage ferric reductase activity may be determined by external free-iron concentrations during macrophage differentiation. Macrophage ferric reductase activity was down-regulated by increased free-iron concentration.

This control of ferric reductase activity could contribute to the low levels of iron observed in the haemochromatotic macrophage. As haemochromatotic macrophages differentiate in high free-iron conditions in \textit{vivo}, their ferric reductase activities may be decreased according to the amount of free iron present. Future studies could address this issue by isolating differentiated macrophages from \(\text{GH}\) and control human subjects, and measuring the activity of the \textit{in vivo} differentiated cells. The clonal reference cell line \(\text{K562}\) did not show modulation of ferric reductase activity by iron. The \(\text{K562}\) cell line originates from an erythroid cell clone derived from a patient with myeloid leukaemia in acute blast crisis [32]. The lack of modulation of \(\text{K562}\) ferric reductase activity by iron may reflect either the erythroid lineage or changes in the clonal expansion of this cell line. The identification of the human ferric reductase gene will eventually establish whether it is transcriptionally repressed by iron, as in \textit{S. cerevisiae}.

Future studies will elucidate whether ferric reductase plays a role in the apparent iron deficiency of macrophages in \(\text{GH}\).

**REFERENCES**

suggesting that the 5' region of ZNF204 may have been derived from coding sequence. Taking the mapping and homology analyses together, it seems likely that the putative parent locus is in 6p21.3 and possibly related to ZNF184. Indeed, ZNF204 and ZNF184 both hybridize to a 3.2-kb transcript on Northern analysis (data not shown and Ref. 3), consistent with a shared ancestry. The processed appearance of ZNF204 suggests a retrotransposed intermediate. Localized enrichment of retrotransposed elements has not been previously demonstrated. Interestingly, the first 90 bp of cDNA sequence has homology with the TH1c repeat element family. Further studies would be required to investigate the potential involvement of this sequence in a mechanism of local retrotransposition.

To analyze the evolutionary conservation of ZNF204, a zoo blot of HindIII-digested DNA was screened using an internal PCR fragment probe (Fig. 1c). This showed a ladder of hybridizing bands increasing uniformly in size by approximately 80 bp in all species tested (pig, cow, hamster, and human; Fig. 1d). This step-wise increase in the size of fragments corresponds to the distance between the phenylalanine-residues of adjoining zinc finger domains of 84 bp. The sequence Lys-Ala-Phe occurs commonly in zinc fingers and can be encoded by sequences that include a HindIII site. For the 10 zinc finger genes with highest homology to ZNF204, 15 of the 20 HindIII sites present were positioned at the conserved

FIG. 1. Mapping, isolation, and characterization of ZNF204 (GenBank Accession No. AF033199). (a) Fine-resolution genomic map around D6S1260. PAC clones used to screen the small intestinal cDNA library are shown; dots indicate the presence of markers (4, 5, 8, 9). (b) Overlapping cDNA clones showing the position of EST449 and STSG-9945; a double slash indicates the junction of ZNF204 cDNA clones with chimeric sequence. (c) Schematic representation of ZNF204 showing the 1.1 kb open reading frame consensus sequence and the positions of its start ATG (arrow at 5' end) and polyadenylation consensus sequence (pA). Blocks indicate the positions of individual zinc finger motifs with asterisks representing termination codons and a crooked arrow indicating the location of the frameshift mutation. The position of the PCR hybridization probe is indicated above the consensus. Below the consensus, the three bars represent the three overlapping PCR products amplified from genomic DNA using cDNA primers; cDNA and genomic DNA products were of identical size. (d) ZNF204 cross-hybridizes with multiple zinc finger loci. Southern blot of HindIII-digested DNA (10 µg) from pig, cow, hamster, and human probed with an internal PCR fragment from ZNF204 (c). Southern blots were rinsed once with 4x SSC at room temperature, twice for 20 min in 0.5x SSC, 0.1% SDS at 65°C, and once for 20 min in 0.5x SSC, 0.1% SDS at 42°C, and once for 20 min in 0.5x SSC, 0.1% SDS at 65°C. A ladder of bands increasing uniformly by approximately 80 bp was seen in all lanes reflecting cross-hybridization of ZNF204 to multiple zinc finger loci across species.
phenylalanine residue. Two zinc finger genes contained 5 HindIII sites, all at the phenylalanine residue of the consensus, giving rise to a ladder of HindIII fragments as detected by hybridization in Fig. 1d. Conservation of zinc finger probes has been previously reported, with one or several cross-hybridizing bands detected on zebrafish (7). The demonstration of a conserved ladder of zinc finger fragments by ZNF204 results from the combination of HindIII as a restriction enzyme, formamide hybridization solution, moderate stringency washing, and possibly the partial degeneracy of the ZNF204 probe.

We conclude that ZNF204 most likely arose by a local retrotransposition event within the 6p21.3 ZNF184 cluster. We observe a remarkable regular ZNF204 hybridization pattern conserved across all species tested, which we have explained in terms of the periodicity of HindIII sites within the consensus sequence of multiple zinc finger proteins. At least 35 individual bands could be resolved in human and bovine DNA, suggesting that the zinc finger genes, and therefore proteins, of these species can contain at least 35 motifs. Indeed, the Xfin protein of Xenopus contains 37 C2H2 zinc finger motifs, consistent with these observations (6). Most zinc finger proteins are transcriptional regulators involved in cell growth and differentiation. The cross-hybridization of ZNF204 to multiple zinc finger loci across species may provide a tool for the analysis of C2H2 zinc fingers throughout evolution.

ACKNOWLEDGMENTS

We thank David Latchman, Arian Smit, Dariusz Górecki, Marita Pohlschmidt, Priyal de Zoysa, and Kevin Moore for helpful discussion. The UK HGMP Resource Centre provided PAC clones and the monocochromosomal hybrid panel; the MRC HGMP funded the CCD FISH equipment. This work was funded by grants from The Peter Samuel Charitable Trust (J.S.D., A.P.W.), and the Nuffield Foundation (A.P.W.). The support of the MRC for the Ph.D. studentship of J.P. is acknowledged with gratitude.

REFERENCES