Human Ceruloplasmin Gene and cDNA

by

William Robert Wager
B. Sc, University of British Columbia, 1986.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
Masters of Science

in

THE FACULTY OF GRADUATE STUDIES
Department of Biochemistry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
December
© William Robert Wager, 1992
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **BIOCHEMISTRY**

The University of British Columbia
Vancouver, Canada

Date **Dec 14/92**

DE-6 (2/88)
ABSTRACT

Ceruloplasmin is a 132Kda serum protein found in all vertebrates. The protein is multi-functional. Functions include copper transport, serum and tissue antioxidant and anti-coagulation modulation. The cDNA of ceruloplasmin share homology with the cDNAs of factors V and VIII of the blood coagulation cascade, suggesting that the three genes evolved from a common ancestral gene. There are two ceruloplasmin specific mRNA transcripts, 3.7 and 4.4 Kb, produced by the liver. Expression of the gene is regulated transcriptionally and post-transcriptionally.

The initial aim of this study was to isolate the 3' end of the gene and to determine the evolutionary relationship with the factor V and VIII genes. Further aims were to isolate and characterize the 5' and 3' ends of the cDNA in order to determine their role in the regulation of gene expression.

The gene for human ceruloplasmin is 38 Kbp in length and is composed of at least 19 exons separated by 18 introns. The exact number of 5' and 3' untranslated exons is unclear at the present. The gene structure of ceruloplasmin shares homology with the genes for factor V and VIII. The 4 amino acid addition reported by Yang et al., (1986) is produced by use of an alternate splice junction at the 3' end of exon 18. A 5' cDNA clone that contained 114 bp prior to the initiator methionine was isolated. The clone is produced by splicing of the primary heteronuclear transcript. The origin of the 5' end of the clone is unmapped at the present. A series of 3' cDNA clones was isolated. All the clones encode the COOH-terminus of the protein, but differ from the 3' genomic sequence at points beyond the stop codon. The clones diverge from the genomic sequence at splice junction consensus site.
The origin of the 3' untranslated sequence is unmapped at the present. One 3' clone hybridizes only to the 4.4 Kb transcript.

The ceruloplasmin gene probably diverged from the ancestral gene before the divergence of the genes for factor V and VIII. The two different length mRNA transcripts are a result of both 5' and 3' untranslated sequence additions to the coding mRNA. The role of these untranslated sequences in regulation of ceruloplasmin gene expression is unclear at the present.
TABLE OF CONTENTS

ABSTRACT .................................................................................................................. ii
TABLE OF CONTENTS ................................................................................................. iv
LIST OF TABLES ........................................................................................................... iv
LIST OF FIGURES ......................................................................................................... vii
LIST OF ABBREVIATIONS ......................................................................................... viii
ACKNOWLEDGMENTS ................................................................................................. ix

INTRODUCTION ............................................................................................................. p.1
    1) Ceruloplasmin is Related to Factors V and VIII ................................................. p.2
    2) Functions of Ceruloplasmin .............................................................................. p.2
        A) Ferroxidase activity .................................................................................. p.2
        B) Serum anti-oxidant activity ..................................................................... p.3
        C) Amine oxidase activity ........................................................................... p.3
        D) Angiogenesis ......................................................................................... p.3
        E) Coagulation ............................................................................................ p.3
        F) Copper transport .................................................................................... p.4
    3) Copper loading ................................................................................................. p.5
    4) Different Molecular Forms of Ceruloplasmin .................................................... p.5
    5) Induction of Ceruloplasmin .............................................................................. p.6
    6) Tissue specific expression .............................................................................. p.7
    7) RNA processing and function ....................................................................... p.8
    8) Objectives of This Study ................................................................................ p.10

MATERIALS AND METHODS ....................................................................................... p.11
    1) Reagents .......................................................................................................... p.11
    2) Bacterial strains and media ............................................................................. p.12
    3) Mini preparation of DNA and Double stranded DNA Sequence Analysis .... p.13
    4) Gel electrophoresis ......................................................................................... p.14
    5) Kinase labeling of Oligonucleotides ................................................................. p.14
    6) Klenow labeling of DNA fragments ................................................................ p.15
    7) Screening of Human Leukocyte Genomic Library ........................................... p.15
    8) Screening of λgt-11 Library ............................................................................. p.16
    9) Northern blot Analysis ................................................................................... p.18
    10) Southern blot Analysis .................................................................................. p.18
    11) RNA extraction from Human Liver Tissue .................................................... p.19
    12) Primer Extension Analysis ............................................................................ p.20
    13) Production of 3' Clones ................................................................................ p.24

RESULTS ....................................................................................................................... p.28
    1) Screening of Genomic Library for the 3' end of the Gene ............................ p.28
    2) 5' cDNA Clone ............................................................................................... p.34
    3) 3' cDNA Clones ............................................................................................. p.39
        A) 59 Clone .................................................................................................. p.39
        B) λgt-11 3' Clone ...................................................................................... p.48
### DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Ceruloplasmin Gene Structure</td>
<td>p.49</td>
</tr>
<tr>
<td>2) Comparison of Ceruloplasmin with Factors V and VIII</td>
<td>p.50</td>
</tr>
<tr>
<td>3) 5' cDNA Clone</td>
<td>p.53</td>
</tr>
<tr>
<td>4) 5'g Clone</td>
<td>p.55</td>
</tr>
<tr>
<td>5) Mechanisms of 5'g Production</td>
<td>p.55</td>
</tr>
<tr>
<td>6) Northern Blot Analysis</td>
<td>p.57</td>
</tr>
<tr>
<td>7) Southern Blot Analysis</td>
<td>p.58</td>
</tr>
<tr>
<td>8) 3' λgt-11 Clone</td>
<td>p.58</td>
</tr>
<tr>
<td>9) Other 3' Clones</td>
<td>p.59</td>
</tr>
<tr>
<td>10) Future Studies</td>
<td>p.59</td>
</tr>
<tr>
<td>11) Conclusion</td>
<td>p.60</td>
</tr>
<tr>
<td>12) References</td>
<td>p.62</td>
</tr>
<tr>
<td>TABLE 1</td>
<td>Oligonucleotides used in the study........................ p.23</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------</td>
</tr>
<tr>
<td>TABLE 2</td>
<td>Nucleotide sequence of the intron/exon junctions of the Human Ceruloplasmin Gene....... p.32</td>
</tr>
<tr>
<td>TABLE 3</td>
<td>Size and Position of the exons and introns in the Human Ceruloplasmin Gene................... p.33</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE 1 Positions of Oligonucleotides used in the Primer Extension ........................................ p.22

FIGURE 2 Procedure for Production of the 5' Series of 3' cDNA Clones........................................ p.26

FIGURE 3 Partial Restriction Map and Intron/Exon Organization of the Human Ceruloplasmin Gene .......................................................... p.29

FIGURE 4 Autoradiography of Primer Extension of Ceruloplasmin mRNA .................................. p.35

FIGURE 5 Sequence of 5' cDNA Clone of Human Ceruloplasmin ............................................... p.37

FIGURE 6 Comparison of the 3' Ends of Different Ceruloplasmin cDNA Clones .......................... p.40

FIGURE 7 5' Clone Sequence ........................................................................................................ p.42

FIGURE 8 Sequence Alignment of 5' and the 3' cDNA ..................................................... p.45

FIGURE 9 Northern Blot Analysis of the 3' End of the Ceruloplasmin Gene ............................. p.46

FIGURE 10 Comparison of the Domain Structures and Intron/Exon Junctions of Ceruloplasmin, Factors V and VIII ........................................ p.51

FIGURE 11 Proposed Secondary Structure of 5' RNA ....................................................... p.54
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>Cytidine</td>
</tr>
<tr>
<td>cpm</td>
<td>Counts per minute</td>
</tr>
<tr>
<td>dH2O</td>
<td>Distilled water</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>Dideoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>DEP</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>ETOH</td>
<td>Ethanol</td>
</tr>
<tr>
<td>G</td>
<td>Guanosine</td>
</tr>
<tr>
<td>hnRNA</td>
<td>Heteronuclear ribonucleic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropylthio-β-D-Galactoside</td>
</tr>
<tr>
<td>Kbp</td>
<td>Kilobase pair</td>
</tr>
<tr>
<td>Kda</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>LMP</td>
<td>Low melting point</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>MMLV-RT</td>
<td>Moloney murine leukemia virus reverse transcriptase</td>
</tr>
<tr>
<td>OLB</td>
<td>Oligonucleotide labeling buffer</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>pfu</td>
<td>Plaque forming unit</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-Tetramethylethylenediamine</td>
</tr>
<tr>
<td>tPA</td>
<td>Tri(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>T</td>
<td>Thymidine</td>
</tr>
<tr>
<td>UTS</td>
<td>Untranslated sequence</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra violet light</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-Bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank my supervisor, Ross MacGillivray, for his support of my research. I would also like to thank Marlys Koschinsky for her work on the majority of the gene structure and the complete cDNA sequence without which my work could not have been done. Credit and thanks must go to Heather Kirk for her work on the 3' end of the gene. Her assistance was invaluable. Special thanks go to Jeff Hewitt for his constant support and unmatched technical advice. I would also like to thank the entire lab for their ideas and companionship during my stay. Lastly, I would like to thank my parents for their unconditional support.
Introduction

Ceruloplasmin is an α2-glycoprotein found in the vertebrate serum where it transports 90-95% of the copper. Ceruloplasmin is synthesized as a single polypeptide chain of 132kda. The secreted holoenzyme contains 6 copper atoms (Kingston, et al., 1977). The complete cDNA sequence of ceruloplasmin has been reported for human (Koschinsky et al., 1986; Yang et al., 1986) and rat (Aldred et al., 1987). The human cDNAs encode two forms of the polypeptide, 1046 or 1050 amino acids. An extended form with a four amino acid addition near the carboxyl terminal of the molecule, has been reported (Yang et al., 1986). Ceruloplasmin is produced in a variety of tissues including: liver, choroid plexus, mammary gland, fetal lung, testes, uterus, placenta, and circulating macrophages (Gitlin et al., 1977; Skinner and Griswold, 1983; Aldred and Grimes, et al., 1987; Thomas and Schreiber, 1988; Fleming and Gitlin, 1990; Fleming et al., 1991).

Ceruloplasmin has been implicated in multiple functions including: mixed function oxidase, copper transport, copper metabolism, iron metabolism, angiogenesis, biogenic amine synthesis, and blood coagulation (Freiden and Hseih, 1976; Goldstein et al., 1979; Frieden, 1981; Walker et al., 1990).
1. Ceruloplasmin is Related to Factors V and VIII

Ceruloplasmin is composed of three homology units called the A domains. Each unit is approximately 350 amino acids in length and the three together make up the entire protein. The homology units share approximately 40% identity with each other and 30-40% with factors V and VIII respectively (Church et al., 1984). These homologies have led to the hypothesis that the three proteins shared a common ancestral gene. Through the process of gene duplication, the ceruloplasmin gene evolved. In addition to the triplicated A domain, factors V and VIII also contain a duplicated C domain and a single B domain. The B and C domains share no sequence homologies with ceruloplasmin and probably represent recombination events after the ceruloplasmin gene evolved into a separate gene.

2. Functions of Ceruloplasmin

Along with the copper transport functions, ceruloplasmin possesses several other functions. The different functions ascribed to the protein are possibly contained within different portions of the three homology units which comprise the entire ceruloplasmin molecule.

A. Ferroxidase activity

The ferroxidase activity of ceruloplasmin helps regulate the conversion of Fe (II), bound to ferritin, to Fe (III), which is bound to and transported by transferrin (Frieden, 1986).
B. Serum anti-oxidant activity

The free radical and superoxide scavenging activity of ceruloplasmin has been reported by several groups (Goldstein et al., 1979 and Samokyszyn et al., 1989). These activities support the hypothesis that ceruloplasmin is the main serum anti-oxidant.

C. Amine oxidase activity

Ceruloplasmin possesses oxidase activity towards a variety of amines and therefore has a possible function in regulating the levels of biogenic amines of the CNS (Frieden, 1981). The oxidase activity is copper dependent (Sato and Gitlin, 1991).

D. Angiogenesis

Ceruloplasmin exhibits angiogenic activity when injected into corneas (Folkman and Klagsbrun, 1987). The activity is dependent on the presence of copper (Raju et al., 1984).

E. Coagulation

A recent report has demonstrated the association of ceruloplasmin with protein C of the anti-coagulation cascade (Walker and Fav, 1990). Ceruloplasmin was added to a reaction containing activated protein C and either factor Va or VIIIa. The proteolytic inactivation of the clotting factors was inhibited by the presence of ceruloplasmin. A synthetic peptide that corresponds to residues 1028-1037 of ceruloplasmin, and shared the highest sequence identity to the protein C binding sites of factor V and VIII, was
added to inactivation assays. The peptide also inhibited the inactivation of both clotting factors by protein C. The peptide increased the amidolytic activity of protein C by 5-fold. Protein C, when added in a 2 molar excess to ceruloplasmin, enhanced the oxidase activity of ceruloplasmin 2-fold. This increase was blocked by addition of the synthetic peptide. Together, these results suggest ceruloplasmin interacts with protein C in vivo to help regulate the anti-coagulation activity.

F. Copper Transport

Ceruloplasmin concentrations in the blood range from 15-60mg/dl in normal individuals. The protein contains 6-7 tightly bound copper ions, although the exact locations of the protein ligands are unknown. Ceruloplasmin also reversibly binds up to 10 additional copper ions (McKee and Frieden, 1971). Presumably, these additional ions are delivered to the tissues. Ceruloplasmin was demonstrated as a copper transport protein in several reports (Campbell et al., 1981; Dameron and Harris, 1987; Marceau and Aspin, 1973; Terao and Owen, 1976). This transport activity is responsible for 90-95% of copper delivery in the body. The remaining 5-10% of circulating copper is bound to serum albumin, a high molecular weight protein called transcuprein and the tripeptide glycine-histidine-lysine (Frieden, 1981, 1985, Pickart et al., 1980). Ceruloplasmin receptors have been reported on chicken heart and aorta tissue (Stevens et al., 1984), on rat liver endothelial cell (Kataoka and Tavassoli, 1984), on red and white blood cells (Barnes and Frieden, 1984; Kataoka and Tavassoli, 1985), and on K-562 cells (Percival and Harris 1988).
Recent reports have suggested that ceruloplasmin releases its copper at the cell surface and does not enter the cell (Percival et al., 1990). This differs from the mechanism of iron transfer from transferrin which involves endocytosis of the entire holoprotein and recycling of the apo-protein to the circulation. Along with its role in copper transport in the blood, ceruloplasmin may be involved in transport of copper across natural barriers in the body (see below).

3. Copper Loading

In normal individuals 90% of the circulating ceruloplasmin is the holoprotein (Holtzman et al., 1970). The remaining 10% represents ceruloplasmin that was secreted without copper incorporation during biosynthesis. The process of copper loading appears to occur early on in the biosynthetic process and is independent of glycosylation of the protein (Gitlin et al., 1991). Recent reports have demonstrated the absolute requirement of bound copper for oxidase activity of ceruloplasmin (Sato et al., 1991).

4. Different Molecular Forms of Ceruloplasmin

The ceruloplasmin molecule appears in two forms in serum. The predominant form is the 132kda first described by Takahashi et al., 1984. More recently, a novel 200kda form has been identified by Sato et al., 1990. Experiments indicate this novel form of ceruloplasmin is a dimer of the 132kda. The function of this dimer is not know at present. Other reported forms of ceruloplasmin appear to be results of incomplete copper loading of the molecule (Sato and Gitlin, 1991).
5. Induction of Ceruloplasmin

The induction of ceruloplasmin gene expression with epinephrine and estradiol was first demonstrated in rats (Meyers et al., 1958). Adrenocorticotropic hormone and hydrocortisone increased ceruloplasmin expression in chickens (Starcher et al., 1965). Together, these reports suggest the involvement of adrenal steroids in ceruloplasmin gene expression. Soon afterwards, lipopolysaccharides and interleukin-1 were shown to increase ceruloplasmin expression (Wannemacher et al., 1975). These compounds are elevated during inflammation and may represent the signals responsible for increased ceruloplasmin expression. It is interesting that the inflammation induced levels of ceruloplasmin transcription do not correlate with the timing of the increase of the protein concentration in the blood (Gitlin, 1988).

The effect of copper concentration on ceruloplasmin induction is unclear at the present. Several reports show increases in ceruloplasmin expression with copper additions to copper deficient rats (Linder et al., 1979; Weiner and Cousins, 1980). More recent experiments show no increase in ceruloplasmin expression following copper addition to copper deficient rats or human cell cultures (Gitlin et al., 1992; McArdle et al., 1990)

Levels of ceruloplasmin also increase during pregnancy (Terao et al., 1977), tissue injury and certain malignancies (Cousins, 1985). The tissues responsible for the increased ceruloplasmin levels are not clear at the present.
6. Tissue Specific Expression

Ceruloplasmin expression is highly tissue specific. The primary site of expression is the liver (Koschinsky et al., 1986; Yang et al., 1986). Rat ceruloplasmin is expressed in the liver, choroid plexus, yolk sac, placenta, testes, uterus and epithelial cells of the mammary glands (Aldred et al., 1987; Jaeger et al., 1991). Expression of ceruloplasmin in human and rat fetal lung tissue has recently been reported (Fleming et al., 1990). The expression in the lung is terminated at or shortly after birth. Expression has also been reported in several cell lines including HepG2, Hep3B, and an adenocarcinoma.

Aside from the liver, the sites of ceruloplasmin expression fall into two main types. Natural barriers of the body including the blood/brain barrier, the placenta and mammary tissue form one group. It has been proposed that ceruloplasmin helps to facilitate the transport of copper across these natural barriers (Aldred et al., 1987). The other type of tissues are sites of development including the fetal lung, the testis. The role of ceruloplasmin in these tissues likely involves the anti-oxidant activity to protect the developing lung tissue and increased copper delivery to the mitochondrial enzymes of the developing spermatocytes.
7. RNA Processing and Function

Tremendous progress has recently been made in understanding the complexities and functions of untranslated RNA sequences. Some of the areas include: alternative 5' and 3' splicing, alternative polyadenylation site selection, 3' untranslated expression regulatory sequences, developmental switching of polyadenylation, poly A tail-translation initiation interactions, 3' untranslated sequence directed transcript stabilization and destabilization (reviewed by Jackson and Standart, 1990).

As early as 1979, Alt et al., reported different patterns of mRNA from a single IgM gene. There have also been numerous reports of multiple polyadenylation site utilization in one gene (reviewed by Leff et al., 1986). In 1988, Virtanen et al., reported the use of different polyadenylation sites in immature and mature lymphocytes produced membrane bound and secreted forms of immunoglobulins. In this paper, sequences close to but not inclusive of the poly A signal, AAUAAA, or "GU" region were involved in the switching of processing during development. The Ubx locus of Drosophila produces a 3.2 and a 4.4kbp transcript that differ in location of the poly A tail (Kornfeld et al., 1989). Other genes with multiple poly A sites include: the rat liver mannan-binding protein (Wada et al., 1990), human liver cytosolic aspartate aminotransferase (Bousquet-Lemercier et al., 1990), the bovine homolog to myoD1 (Clark et al., 1990), and the human fetal liver gamma-globulin genes (Dmitrenko et al., 1991). In most of the reports the selection of the poly A site requires a consensus AAUAAA signal, a "GU" rich region and other
uncharacterized 3' untranslated sequences. Some genes require these additional sequences to be within 30 bases of the poly A signal (Brown et al., 1991), while other genes have regulatory sequences over a kilobase away (Jackson and Standart, 1990). Some of the proteins that bind these sites have been purified, as detailed below.

The poly A binding protein is a 70 kilodalton monomer. Cloning of this gene from a variety of species revealed homology in the N-terminal two thirds of all the proteins (Adam et al., 1986; Sachs et al., 1986; Grange et al., 1987; Zelus et al., 1989). It has been proposed that the PABP binds to the poly A tail and helps stabilize the transcript and protect it from 3' to 5' degradation (Bernstein et al., 1989).

Other proteins include the cleavage/polyadenylation factor which binds the AAUAAA polyadenylation signal and cleavage factors 1 (Bienroth et al., 1990). The poly A tail together with the poly A binding protein increases translation rates. Present theories hold that the interaction involves initiation or re-initiation of translation (Munroe et al., 1990).

The length of the poly A tail has also been shown to influence translation. Most of the modulation of poly A tail length occurs in developmental regulation of gene expression. Examples include mouse tissue plasminogen activator and protamine of spermatocytes. The tPA mRNA is stored in the immature oocyte with a poly A tail of 10-40 nucleotides. Upon fertilization, the poly A tail is increased to 400-600 bases and translation begins (Strickland et al., 1988). Maturation-dependent polyadenylation of tPA is controlled by the 3' untranslated sequence (Vassalli et al., 1989).
Other 3' untranslated sequences are involved in destabilization of mRNA's. An "AU" rich sequence causes destabilization of mRNA molecules. These sequences have been reported in heat shock transcripts (Yost et al., 1990) and lymphokine mRNAs (Jones and Cole, 1987).

8. Objectives Of This Study

The aims of this study are as follows: (i) clone and sequence the remainder of the 3' end of the human ceruloplasmin gene and determine the evolutionary relationships with the genes for factor V and VIII, (ii) clone and characterize 5' and 3' untranslated regions in order to help determine their role in the regulation of ceruloplasmin expression.
Methods and Materials

1. Reagents

Yeast extract, bacto-tryptone and bacto-agar were purchased from Difco Laboratories. Acrylamide, bisacrylamide, ammonium persulfate, agarose, urea and TEMED were purchased from Bio-Rad Laboratories. Ammonium sulfate and phenol were purchased from British Drug Houses Ltd. Deoxy and dideoxy-ribonucleotides were purchased from Pharmacia. DTT, RNase A, β-mercaptoethanol, DNase I and lysozyme were purchased from Sigma Chemical Co. α-[thio-35S]-dATP was purchased from New England Nuclear. TA Cloning kit was purchased from Invitrogen. All other reagents were of reagent grade or higher and were purchased from Bio-Rad, BDH, Pharmacia or Sigma.

Phage T4 polymerase and Klenow fragment of E.coli DNA polymerase, MMLV-RT and all the restriction enzymes were purchased from Bethesda Research Laboratories or Pharmacia. T7 DNA polymerase was purchased from Pharmacia. Recombinant Taq polymerase (Amplitaq) was purchased from Perkin Elmer Cetus.

Kodak X-Omat and Kodak XAR film were used for autoradiography.
2. Bacterial Strains and Media

The bacterial medium used to grow lambda phage was NZY (Maniatis et al., 1982) (10g NZ amine, 2g MgCl₂, 5g Yeast extract, adjusted to pH 7.5 with NaOH). The phage libraries were plated on NZY-agar (1.5% w/v) plates, with an overlay of NZY-agarose (0.7% w/v). The medium used to grow all transformed bacteria was Lauria broth (LB) (Maniatis et al., 1982)(5g yeast extract, 10g bactotryptone and 10g NaCl per liter). The selection medium used for transformed bacteria contained the addition of 50-100μg/ml ampicillin or kanamycin.

<table>
<thead>
<tr>
<th>Stains</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>F-,φ80d, lacZΔM15, endA1, recA1, hsdR17(rK-,mK+), supE44, thi-1, d-, gyrA96, Δ(lacZYA-argF), U169</td>
</tr>
<tr>
<td>Y1090</td>
<td>Δ(lacU169), proA+, Δ(ion), araD139, strA, supF, [trpC22::Tn10], (pMC9), (rK-,mK+)</td>
</tr>
<tr>
<td>LE392</td>
<td>F-, hsdR574 (rK-,mK+), supE44, supF58, lacY1</td>
</tr>
</tbody>
</table>
3. Mini Preparation of DNA and Double Stranded DNA Sequence Analysis

An overnight culture of bacteria was pelleted in a 1.5 ml Eppendorf tube. The pellet was resuspended in 100 µl of STET buffer (8% sucrose, 5% Triton-X100, 50mM EDTA, 0.5 mg/ml lysozyme). The suspension was boiled for 2 minutes and then centrifuged for 15 minutes. The supernatant was removed and added to 100 µl of isopropanol and centrifuged for 5 minutes. The pellet was washed with 70% ETOH and resuspended in 40 µl of dH2O. The resulting solution was suitable for sequence analysis and/or enzyme digests.

Double stranded DNA sequence analysis involved mixing 8 µl of the DNA solution with 11 µl dH2O, 10 picomoles of primer, and 2 µl of 2N NaOH. The solution was boiled for 2 minutes. One tenth volume of 3M NaOAc and 2 volumes of 95% ETOH were added. The solution was centrifuged for 15 minutes. The pellet was washed with 70% ETOH and air dried. The pellet was resuspended in 8 µl dH2O and 2 µl of 5X sequencing buffer (200 mM Tris-Cl pH7.7, 100 mM MgCl2, 250 mM NaCl). The primer was annealed at 37°C for 15 minutes. To the solution, the following were added; 1 µl 100mM DTT, 2 µl labeling mix (7.5 µM dGTP, 7.5 µM dTTP, 7.5 µM dCTP), 1 µl 35S-dATP (1200Ci/mM), and 2 µl T7 polymerase (diluted 1:6). After 5 minutes extension at 37°C, 3 µl of the reaction was added to 2.5 µl of each termination mix (80µM dNTP, 8µM ddNTP, 50mM NaCl). After 5 minutes of termination, 4 µl of stop buffer (90% formamide, 20 mM EDTA, 0.3% Bromophenol blue and Xylene cyanol) was added to each reaction. The reactions were boiled for 2 minutes prior to loading on
a denaturing polyacrylamide gel. After electrophoresis, the gels were dried in a Bio-Rad gel dryer and placed on X-ray film.

4. Gel Electrophoresis

DNA fragments were separated according to size by electrophoresis in agarose or polyacrylamide gels. The buffer for agarose gels was 1X TAE (50X TAE contains 2M Tris pH 7.5, 1M glacial acetic acid, and 100mM EDTA; Maniatis et al., 1982). The buffer for polyacrylamide gels was 1X TBE (10X TBE contains 890mM Tris, 890mM Boric acid, and 25mM EDTA, pH 8.3; Maniatis et al., 1982). Denaturing polyacrylamide gels contained 8.3 M Urea. Bis:Acrylamide stocks contain bisacrylamide:acrylamide in H2O (2%:38%, w/v).

Polymerization was initiated with the addition of ammonium persulphate and TEMED to final concentrations of 0.066% (w/v) and 0.04% (w/v), respectively. Electrophoresis of agarose gels containing 2μl of EtBr (10mg/ml) and varying concentrations of agarose, (0.6%-2%), were used to separate the different size DNA fragments. The DNA was visualized under a UV lamp (260nm).

5. Kinase Labeling of Oligonucleotides

A typical reaction contained 50-100 nanograms of oligonucleotide suspended in 5 μl of dH2O. To the oligonucleotide, the following were added; 1 μl 10X kinase buffer (500mM Tris-Cl pH 7.6, 100mM MgCl2, 50mM DTT, 1mM spermidine, 1mM EDTA), 2 μl γ32P-ATP (3000Ci/mmole), 11 μl dH2O and 1 μl T4 polynucleotide kinase (10 units). The reaction was incubated for 1-4 hours at 37°C. The
reaction was terminated with the addition of 2μl of 500mM EDTA and 80 μl of dH2O.

6. **Klenow Labeling of DNA Fragments**

DNA was labeled by the procedure of Feinberg and Vogelstein, 1982. The typical reaction contained 50-100 nanograms of DNA suspended in 20 μl of dH2O. The DNA solution was boiled for 5 minutes and then snap cooled on ice for 2 minutes. To a separate Eppendorf tube 20 μl of oligonucleotide labeling buffer (OLB) was added to 4 μl of BSA (10mg/ml), 4 μl α³²P-dATP(3000Ci/mmole), 4 μl α³²P-dCTP and 1 unit Klenow (Large fragment DNA Polymerase). After cooling, the DNA was added to the reaction buffer solution. After a pulse spin in the Eppendorf centrifuge, the reaction was incubated at 37°C for 4-12 hours. After the reaction was complete, 150 μl of 10mM EDTA was added. One microliter of the diluted reaction was precipitated with TCA (see Maniatis et al., 1982) to measure the specific activity of the probe. Only probes with specific activity of greater than 10⁷ cpm/μg DNA were used.

7. **Screening of Human Leukocyte Genomic Library**

The amplified library was purchased from Clontech (Palo Alto, Ca.). The library was constructed with DNA (partially digested with MboI) cloned into the BamHI sites of EMBL-3 lambda phage. The genomic inserts were excised from the phage by digestion with Sal I. The average insert size was 16 kb. The initial screen was carried out on 1x10⁶ plaques, representing approximately 5 genomic equivalents. Phage were plated at a density of 5x10⁴ plaques per 150mm petri
The phage were incubated with 2ml of overnight culture of *E. coli*, strain P2-392, in NZY media (supplemented with 0.2% maltose) at 37°C for 15 minutes to allow attachment. The phage were then plated on NZY-agarose. The plates were incubated at 37°C for 5-6 hours until the plaques were visible but not confluent. Replica plates were produced. Two sets of plaques were transferred to nitrocellulose filters (132mm, Schleicher and Schuell). The filters were placed in denaturing solution (0.5 N NaOH, 1.5 M NaCl) for 5 minutes. The filters were then transferred to a neutralizing solution (1 M Tris-Cl pH 7.5, 1.5 M NaCl) for 5 minutes. DNA was fixed to the filters by baking the filters at 68°C for 2 hours. Ceruloplasmin specific clones were identified by plaque-hybridization (Benton and Davis, 1977) to 3' ceruloplasmin cDNA restriction fragments labeled with 32P by Klenow extension (see Klenow Labeling section). Recombinant clones of ceruloplasmin were detected by autoradiography and purified to homogeneity by successive rounds of dilution plating and hybridization.

8. Screening λgt11 Library

An aliquot of a HepG 2 λgt11 library containing approximately 10^6 phage was absorbed to 2ml of an overnight culture of *E. coli*, strain Y1090, by incubation at 37°C for 15 minutes. The library was plated out at a concentration of 50,000 plaques per 150mm petri dish. The plates were incubated for 5-6 hours until the individual plaques were visible but separate. Phage were transferred to duplicate nitrocellulose filters (see genomic phage screen), denatured and fixed. The filters were incubated in prehybridization solution (see
Southern procedure) for two hours. An EcoRI/HincII restriction fragment from the 5' end of λhCP-1 (a cDNA clone containing the 5' end of the ceruloplasmin cDNA) was labeled with $^{32}$P using Klenow large fragment polymerase. This probe was incubated with the filters for 16 hours at 50°C. The filters were washed under high stringency and placed on X-ray film overnight. Duplicate positive areas of the phage plates were excised. The agarose plugs were submerged in 500μl of SM buffer and placed at 4°C overnight. One microliter of the adsorbed phage solutions were used as template for a PCR based second screen. The PCR conditions were: 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 1 minute. A 3' to 5' directed oligonucleotide, located (+110-90), was combined with oligonucleotides specific for the forward or reverse arms of the λgt11 phage to give ceruloplasmin specific amplification. The PCR products were ligated into the TA cloning vector (Invitrogen) at 15°C for 4 hours. The ligation products were used to transform competent E.coli, strain DH5α. The transformation was plated out on LB-kanamycin plates (containing 0.1M IPTG and 2% X-gal) and grown for 24 hours at 37°C. Several white colonies were picked and used to inoculate 1.5ml of LB-kanamycin medium. DNA from the 1.5 ml cultures was produced using the mini prep technique. DNA from several different clones were sequenced with forward and reverse universal primers. Agarose plugs from the λgt11 library that contained ceruloplasmin cDNA clones were dilution plated in order to isolated individual plaques. The plaques were screened a second time with a probe from the 5' end of the ceruloplasmin cDNA clone (see above). A duplicate positive phage plaque was amplified in
liquid culture. The DNA was isolated with a scaled down large scale lambda phage isolation procedure (Maniatis et al., 1989). The insert was released with EcoRI digestion and ligated into an EcoRI digested Bluescript vector. The ligation mixture was used to transform competent DH5α bacteria. The bacteria were plated on LB-ampicillin plates (containing 0.1M IPTG and 2% X-gal) overnight. White colonies were picked and grown up in 1.5ml of LB-ampicillin (50µg/ml) overnight. The DNA was prepared by the mini prep technique.

9. Northern Blots

The RNA was prepared as follows. Ten micrograms of total RNA (liver) was added to 2µl of 5X formaldehyde running buffer (100mM MOPS pH7.0, 40mM NaOAc, 5mM EDTA pH8.0), 3.5µl deionized formaldehyde, 10µl deionized formamide and 1µl ethidium bromide (10mg/ml). The solution was incubated at 68°C for 15 minutes. The solution was cooled on ice for two minutes and 2µl of formaldehyde loading buffer was added prior to loading into a 1% formaldehyde agarose gel (Maniatis et al., 1989) and electrophoresed at 100 volts for 30 minutes. No denaturation was required. The RNA was transferred to nitrocellulose filters by the method of Southern (Maniatis et al., 1982). After 12 hours of transfer the RNA was immobilized to the filters by baking for 2 hours at 68°C.

10. Southern Blots

After the DNA was separated according to size by agarose gel electrophoresis, the gel was placed in denaturing solution (1.5 M
NaCl/0.5 N NaOH) for 30 minutes. The gel was placed on a sheet of 3mm Whatman paper saturated with 10X SSC. A nitrocellulose sheet was wetted with 2X SSC (Maniatis et al., 1982) prior to overlaying the gel. The DNA was transferred in a buffer of 10X SSC in a sandwich apparatus (Maniatis et al., 1982). The DNA was fixed to the filter by exposure to UV light (260nm) for 2 minutes. The filter was placed in a prehybridization solution of 6X SSC, 2X Denhardt's (0.02% BSA, 0.02% ficoll, and 0.02% polyvinylpyrrolidone), 0.1% SDS, and 30µg/ml sheared single stranded salmon sperm DNA. The prehybridization mixture was incubated at 68°C for 2 hours. A probe labeled with $^{32}$P (see Klenow labeling section) was boiled for 5 minutes prior to incubating with the filter. Hybridization conditions were overnight at the appropriate temperature for the given probe length. The filter was washed in 1X SSC and then under increasingly more stringent conditions (lower concentrations of SSC and increasing temperature) until the background was low. The filter was air dried and placed on X-ray film overnight at -70°C.

11. RNA Extraction From Human Liver

The procedure used was described by Chomczynski and Sacchi, 1987. Briefly, one gram of previously frozen (-70°C) human liver was placed in a mortar filled with liquid nitrogen. The liver was ground with a pestle and transferred to a new 50ml polypropylene centrifuge tube. The liquid nitrogen was allowed to evaporate and 10 ml of solution D (4 M guanidinium thiocyanate, 25mM sodium citrate pH 7.0, 0.5% sarkosyl w/v, and 100mM β-mercaptoethanol) was added. The tissue was disrupted with a polytron for 10-20
seconds. Next, 1 ml of 2 M NaOAc pH 4.0, 10 ml water saturated phenol, and 2 ml of chloroform:isoamyl alcohol [49:1] were added and mixed by inversion. The solution was placed on ice for 10 minutes. The tube was centrifuged at 10,000g for 20 minutes at 4°C. The aqueous phase, containing the RNA, was transferred to a new tube and mixed with 10 ml of isopropanol and placed at -20°C for 1 hour. The RNA was pelleted by centrifugation at 10,000g and resuspended in 3 ml solution D. The RNA solution was then mixed with 6 ml ETOH and precipitated RNA was collected by centrifugation. The pellet was washed with 70% ETOH, dried and dissolved in 500 μl of 0.5% SDS in DEP treated H₂O at 65°C for 10 minutes. The RNA was stored at -70°C. The integrity of the RNA was determined by electrophoresis of an aliquot on a formaldehyde agarose gel, containing 1μl (10mg/ml) EtBr, and looking for 18S and 28S bands under UV light.

12. Primer Extension Analysis

One hundred nanograms of each oligonucleotide (20-mer) were labeled at the 5' ends with γ³²P-ATP using T4 polynucleotide kinase. The reaction was diluted with 100 μl of dH₂O. The diluted reaction was passed through a charged SEP-PAK (C18) cartridge. The cartridge was washed with 10 ml of dH₂O and the oligonucleotide was eluted with 20% acetonitrile in dH₂O. The eluate was dried down in a speed-vac desicator. The labeled oligonucleotide was resuspended in 10 μl of dH₂O. Approximately 100,000 cpm of labeled oligonucleotide was mixed with 10 μg of total liver RNA. One tenth volume of 3 M NaOAc and 2.5 volumes of 95% ETOH were
added. The solution was frozen at -70°C for 1 hour. The solution was centrifuged at 14000g for 15 minutes. The pellet was washed with 70% ETOH and air dried. The pellet was resuspended in 30 µl of hybridization buffer (40 mM PIPES pH 6.4, 1 mM EDTA pH 8.0, 400 mM NaCl, 80% formamide). The hybridization mixture was incubated at 85°C for 10 minutes and quickly transferred to 37°C for 12 hours. To the hybridization mixture, 170 µl dH2O and 400 µl ETOH were added and the solution was placed on ice for 1 hour. The mixture was centrifuged for 15 minutes. The pellet was washed with 70% ETOH and air dried. The pellet was resuspended in 20 µl reverse transcription buffer (50 mM Tris-Cl pH 7.6, 60 mM KCl, 10 mM MgCl2, 1 mM dNTP, 1 mM DTT, 20 units placental RNase inhibitor, 1 µg actinomycin D). Fifty units of Superscript reverse transcriptase were added and the reaction was incubated at 37°C for 2 hours. The reaction was terminated with the addition of 1 mM EDTA. The RNA was degraded with the addition 5 nanograms of DNase-free RNase and incubated at 37°C for 30 minutes. The reaction was diluted with 150 µl TE pH 7.6 (containing 100 mM NaCl) and 200 µl phenol:chloroform. After centrifugation, the aqueous phase was transferred to a new Eppendorf tube. To this tube, 500 µl of 95% ETOH was added and the tube was placed on ice for 1 hour. The precipitated ssDNA was pelleted by centrifugation at 14000g for 20 minutes. The pellet was washed with 70% ETOH. The pellet was resuspended in 4 µl of TE pH 7.4, and 6 µl of formamide loading buffer. The solution was heated to 95°C for 5 minutes prior to loading on a 6% polyacrylamide:urea gel. To estimate the size of the
Figure 1

Positions of Oligonucleotides used in the Primer Extensions

The arrows represent the positions of the oligonucleotides with respect to the published cDNA sequence (Koschinsky et al., 1986). The +1 site represents the first amino acid in the mature protein. The shaded bars represent the single stranded DNA products of the primer extension experiments.
### TABLE 1

**Oligonucleotides used in this study**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Location</th>
<th>C/NC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo 1a</td>
<td>5'-GGAGACAAAGTCAAAATTAT-3'</td>
<td>exon 14</td>
<td>C</td>
<td>P.C.R.</td>
</tr>
<tr>
<td>Oligo 2a</td>
<td>5'-CCCTAAATTCATCAAGTGT-3'</td>
<td>intron</td>
<td>N</td>
<td>NC P.C.R.</td>
</tr>
<tr>
<td>Oligo 3a</td>
<td>5'-TACAATATAAACCAAAAT-3'</td>
<td>intron</td>
<td>N</td>
<td>C Probe</td>
</tr>
<tr>
<td>Oligo 23.1</td>
<td>5'-GGAGCCTGAGAAGAAATGAAG-3'</td>
<td>5' UTS</td>
<td>NC</td>
<td>Primer Extension</td>
</tr>
<tr>
<td>Oligo 17.1</td>
<td>5'-GCATAATCCCAAGTCTGTTTC-3'</td>
<td>exon 1</td>
<td>NC</td>
<td>Primer Extension</td>
</tr>
<tr>
<td>Oligo dT&lt;sub&gt;17&lt;/sub&gt;-XSP</td>
<td>5'-T17TCTAGAGAGCTCCTGCAG-3'</td>
<td></td>
<td></td>
<td>Extension Sequence</td>
</tr>
<tr>
<td>Oligo 21.1</td>
<td>5'-GACCACATTCTATGCTGGAATG-3'</td>
<td>exon 18</td>
<td>C</td>
<td>3' Clone Production</td>
</tr>
</tbody>
</table>

<sup>a</sup>C-coding strand  
NC-non-coding strand
extended product, the reactions were electrophoresed alongside previously characterized sequencing reactions.

13. Production of 3' Clones

Total RNA from human liver was used as template for the first strand synthesis. First strand synthesis of single strand DNA was primed with oligo dT17-XSP. One microgram of total RNA was added to 10 µl of DEP-treated water. To this, 100 nanograms of oligo dT was added and the mixture was heated to 65°C for 15 minutes. The mixture was cooled to room temperature. To the mixture, the following were sequentially added: 10 µl of 5X RT buffer, 2.5 µl of 10mM dNTP, 1 µl of human placental RNase inhibitor, 17 µl DEP water, 1 µl 100mM DTT and 50 units of MMLV-RT. The reaction was incubated at 37°C for 2 hours and then the reaction was extracted once with phenol/chloroform. After extraction, the solution was incubated with 1 µl of RNase (10mg/ml) at 37°C for 30 minutes. The reaction was loaded on a 1% LMP agarose gel. After electrophoresis, the lane was cut out of the gel. The gel-lane was cut into slices corresponding to approximately 200 base pair increments. The gel slices were added to 500µl of water and heated to 68°C for 15 minutes. These solutions became the template for the PCR reactions.

To 39 µl of water, 1 µl of template, 5 µl of 10X PCR buffer, 1 µl 100mM MgCl₂, 1 µl 10mM dNTP, 1 unit of Taq polymerase, and 1µl (20picograms) of oligonucleotides were added. The oligonucleotides used were dT17-XSP and #21.1 (cDNA position 3130-3149). The PCR was carried out under the following conditions; 94°C for 30
seconds, 45°C for 30 seconds, 72°C for 3 minutes per cycle, and 35 cycles. Varying concentrations of MgCl₂ ranging from 500μM to 5mM were tried to determine the optimum. The PCR reactions were stored at 4°C. The reactions were analyzed by electrophoresising 5 μl (10%) of each reaction on a 1% agarose gel. The gel was denatured and the DNA was transferred to nitrocellulose by the Southern method. The filter was probed with an end labeled oligonucleotide (3270-3289) to show ceruloplasmin specific amplification. Product bands were cut out of the gels and the DNA was electroeluted and ETOH precipitated. The DNA was resuspended in 10 μl of dH₂O. Half (5μl) of the product was ligated into the Invitrogen TA Cloning vector as described in TA Cloning Kit protocol. White colonies were picked and grown overnight in LB-Kanamycin media. DNA from the cultures was prepared for DNA sequence analysis by the mini prep procedure. Prior to sequence analysis, the insert sizes were determined by EcoRI/KpnI digestion and electrophoresis. Initial screens for positive clones were by sequencing with an oligonucleotide primer directly 3' to #21.1 in the coding cDNA. All positive clone inserts were further sequenced completely on both strands.
Procedure for production of the 59 series of 3' cDNA clones.

Reverse transcription of human liver RNA was primed with an oligo dT primer. The RNA was digested with ribonuclease, and the single stranded DNA was size separated by agarose gel electrophoresis. The lane containing the cDNA was cut into slices of approximately 200bp increments. The slices were heated to melt the agarose and the resulting solutions were used as templates for the P.C.R.
RNA → MMV-RT → DNA

Gel Electrophoresis

1St. Strand Synthesis

M. Wt. Marker

Excise Gel Slices

M. Wt. Marker

Gel Slice Template for PCR

M. Wt. Marker

Clone and Sequence

PCR Products
Results

1. Screening the Genomic Library for the 3' End of the Gene

The isolation and characterization of the genomic phage clone was carried out with the assistance of Heather Kirk.

One million recombinant clones from a human genomic library were screened at high density for the 3' end of the ceruloplasmin gene. Eleven duplicate positive areas were removed, diluted, replated and rescreened. One duplicate positive area was excised, diluted and plated in order to isolate a single plaque. The phage was amplified in liquid culture and purified. The DNA from the phage clone was digested with Sall to release the insert from the lambda arms. The insert was separated from the arms by electrophoresis in an agarose gel. A gel slice containing the insert was removed and the DNA was recovered by electro-elution. The insert was approximately 13.6 kb in length. The insert was digested with EcoRI and the resulting fragments were cloned into Bluescript vector either digested with EcoRI or EcoRI and Sall. A portion of the insert EcoRI digest was electrophoresed on an agarose gel to determine the number and sizes of the different restriction fragments. Twenty white colonies were picked and grown overnight in liquid culture. DNA isolated from the cultures was digested with either EcoRI or
Figure 3

Partial restriction map and intron/exon organization of the human ceruloplasmin gene.

A complete map for Bam HI (B), SalI (S), HindIII (H), and EcoRI (E) is shown at the bottom of the figure. A partial map for AccI (A), XbaI (X), KpnI (K), SstI (T), and BalI (L) is also given. Exons are shown as black boxes on the line which represents genomic DNA. The location of the 10 phage clones that together contain the gene are labeled wt. No phage was isolated that spanned the intron between exons 14 and 15.
EcoRI and Sall. The phage insert contained exons 15-19 within 8 kb, (which completed the cDNA sequence, 2656-3321) and 5 kb of 3' untranslated DNA. The subclones of the phage insert were probed with $^{32}\text{P}$ labeled oligonucleotide probes derived from the 3' end of the cDNA to determine the positions of the exons within the restriction fragments. All the intron/exon junctions were sequenced in both directions. All the intron junctions were spliced according to the GT..AG rule (Breathnach and Chambon, 1981; Cech, 1983). The exons varied in length from 129 to 255bp. Exon 18 contained two different 3' ends that differed by 12 nucleotides. Use of a second cryptic 5' splice junction 12 nucleotides downstream produced the different transcript. The result of the addition of the 12 nucleotides was a mature protein containing 4 additional amino acids. These findings agree with the report of Yang et al., 1986 of two different cDNAs coding for the human ceruloplasmin protein.

Because the 5' end of WT10 did not overlap with the 3' end of previously isolated genomic phage clones, the exact length of intron N was determined by PCR. Oligonucleotides derived from cDNA and genomic DNA sequences were produced to amplify the gap between exons 14 and 15. The result of the PCR amplification of genomic DNA was a single band. A Southern blot of the PCR product was probed with an oligonucleotide that corresponded to the known intronic region between the two exons, confirming the intronic length of 2.2kb.
## TABLE II

**Nucleotide sequence of intron/exon junctions in the human ceruloplasmin gene**

Exon sequence is shown in upper case; intron sequence is shown in lower case. The codon phase refers to the position of the intron in the codon triplet. 0 intron occurs between codons, I intron occurs after the first nucleotide, and II intron occurs after the second nucleotide of the codon.

<table>
<thead>
<tr>
<th>Exon number</th>
<th>5' splice donor</th>
<th>Intron</th>
<th>3' splice acceptor</th>
<th>Codon phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>L4b</td>
<td>n.d.²</td>
<td>LD</td>
<td>cgctttctccccctcgaagAAGG</td>
<td>--</td>
</tr>
<tr>
<td>1b</td>
<td>ACACgtaaga</td>
<td>A</td>
<td>tgtcttttttctttgcagGGAA</td>
<td>II</td>
</tr>
<tr>
<td>2b</td>
<td>GAGGgtaagt</td>
<td>B</td>
<td>cattgcatgtgcttcctagGGC</td>
<td>I</td>
</tr>
<tr>
<td>3b</td>
<td>AAGGgtacat</td>
<td>C</td>
<td>aatagtaactttaactccagATTC</td>
<td>I</td>
</tr>
<tr>
<td>4b</td>
<td>CTCCgtaaga</td>
<td>D</td>
<td>actctgctcttgactcacagCTGT</td>
<td>I</td>
</tr>
<tr>
<td>5b</td>
<td>AAGGgtagga</td>
<td>E</td>
<td>cttctttgttcatttcagCCGG</td>
<td>I</td>
</tr>
<tr>
<td>6b</td>
<td>GAAGgtaatt</td>
<td>F</td>
<td>acacccccctccccctagTGAC</td>
<td>II</td>
</tr>
<tr>
<td>7b</td>
<td>CTGGgtgagt</td>
<td>G</td>
<td>aattctctctttgctccagGTC</td>
<td>I</td>
</tr>
<tr>
<td>8b</td>
<td>AGAAGtgagt</td>
<td>H</td>
<td>taatgtgacctctctcacagCTGT</td>
<td>I</td>
</tr>
<tr>
<td>9b</td>
<td>ACAGgtaatt</td>
<td>I</td>
<td>tcattttgtttttattacagGATG</td>
<td>O</td>
</tr>
<tr>
<td>10b</td>
<td>CACTgtaatt</td>
<td>J</td>
<td>ttatccaaacttttacagCCAT</td>
<td>I</td>
</tr>
<tr>
<td>11b</td>
<td>ACAGgtaatt</td>
<td>K</td>
<td>ttctcttatctctccagGACC</td>
<td>II</td>
</tr>
<tr>
<td>12b</td>
<td>AGAGgtaatt</td>
<td>L</td>
<td>gttgAAAAAAAAAAAAAAgTTGTT</td>
<td>I</td>
</tr>
<tr>
<td>13b</td>
<td>CTAGgtgagt</td>
<td>M</td>
<td>ttctctctccactcttccagGTC</td>
<td>I</td>
</tr>
<tr>
<td>14b</td>
<td>CCAggtactc</td>
<td>N</td>
<td>aaacactttgtgatgatatgTAG</td>
<td>I</td>
</tr>
<tr>
<td>15</td>
<td>TAAGgtaaa</td>
<td>O</td>
<td>aagatctctctctctccagGACC</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>CATGgtatgt</td>
<td>P</td>
<td>gatatgctcttcttagCTAT</td>
<td>I</td>
</tr>
<tr>
<td>17</td>
<td>CAAGgtaaga</td>
<td>Q</td>
<td>ttgtcatttttcataattagCACA</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>GAAGgtaaat</td>
<td>R</td>
<td>aaaaatgctttttttctatagACAC</td>
<td>I</td>
</tr>
<tr>
<td>18*</td>
<td>CCAggtagta</td>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>19</td>
<td>GCAT  3' end</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹not determined  
* alternate 3' end of exon  
²taken from M. Koschinsky thesis, 1988
TABLE III

Sizes and positions of exons and introns in the human ceruloplasmin gene.

Sizes of all introns were estimated from restriction enzyme analysis and P.C.R. (see text for details). Exon sizes were determined from the DNA sequence analysis of the gene.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Nucleotide Position</th>
<th>Length (bp)</th>
<th>Intron</th>
<th>Nucleotide Position</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>-12 - 146</td>
<td>158</td>
<td>A</td>
<td>159 - 2459</td>
<td>2300</td>
</tr>
<tr>
<td>2a</td>
<td>2460 - 2708</td>
<td>248</td>
<td>B</td>
<td>2709 - 4759</td>
<td>2050</td>
</tr>
<tr>
<td>3a</td>
<td>4760 - 4973</td>
<td>213</td>
<td>C</td>
<td>4974 - 5774</td>
<td>800</td>
</tr>
<tr>
<td>4a</td>
<td>5775 - 5949</td>
<td>174</td>
<td>D</td>
<td>5950 - 7400</td>
<td>1450</td>
</tr>
<tr>
<td>5a</td>
<td>7401 - 7656</td>
<td>255</td>
<td>E</td>
<td>7657 - 8957</td>
<td>1300</td>
</tr>
<tr>
<td>6a</td>
<td>8958 - 9130</td>
<td>172</td>
<td>F</td>
<td>9131 - 12231</td>
<td>3100</td>
</tr>
<tr>
<td>7a</td>
<td>12232 - 12372</td>
<td>140</td>
<td>G</td>
<td>12373 - 14273</td>
<td>1900</td>
</tr>
<tr>
<td>8a</td>
<td>14274 - 14427</td>
<td>153</td>
<td>H</td>
<td>14428 - 15578</td>
<td>1150</td>
</tr>
<tr>
<td>9a</td>
<td>15579 - 15791</td>
<td>212</td>
<td>I</td>
<td>15792 - 25192</td>
<td>9400</td>
</tr>
<tr>
<td>10a</td>
<td>25193 - 25344</td>
<td>151</td>
<td>J</td>
<td>25345 - 26645</td>
<td>1300</td>
</tr>
<tr>
<td>11a</td>
<td>26646 - 26859</td>
<td>213</td>
<td>K</td>
<td>26860 - 28110</td>
<td>1250</td>
</tr>
<tr>
<td>12a</td>
<td>28111 - 28319</td>
<td>208</td>
<td>L</td>
<td>28320 - 29920</td>
<td>1600</td>
</tr>
<tr>
<td>13a</td>
<td>29921 - 30061</td>
<td>140</td>
<td>M</td>
<td>30062 - 31312</td>
<td>1250</td>
</tr>
<tr>
<td>14a</td>
<td>31312 - 31441</td>
<td>129</td>
<td>N</td>
<td>31443 - 32243</td>
<td>2100</td>
</tr>
<tr>
<td>15</td>
<td>32243 - 32351</td>
<td>107</td>
<td>O</td>
<td>32352 - 33451</td>
<td>1100</td>
</tr>
<tr>
<td>16</td>
<td>33452 - 33669</td>
<td>217</td>
<td>Q</td>
<td>34212 - 35461</td>
<td>1250</td>
</tr>
<tr>
<td>18</td>
<td>34462 - 35625</td>
<td>163</td>
<td>R</td>
<td>35626 - 37875</td>
<td>2250</td>
</tr>
<tr>
<td>19</td>
<td>37876 - 38016</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) taken from M. Koschinsky thesis, 1988
2. 5' cDNA Clone

A lambda gt11 library was screened for a clone that contained the entire 5' cDNA. One clone was purified to homogeneity and the insert was cloned into the Bluescript vector. The DNA was sequenced with the 17.1 reverse direction primer (110-90 nucleotide numbers of cDNA). Sequencing of the clone revealed the presence of the 5' end of the coding sequence, the 59 base signal sequence and an additional 154 bases prior to the EcoRI linker and λgt11 vector arm (see Figure 5). The sequence of the clone agreed with a previously reported 5' cDNA clone isolated by Koschinsky et al., 1986. Both clones diverge from the genomic sequence 13 bases 5' to the initiator methionine. There is a 3' consensus splice junction at this site in the genomic DNA.

Primer extension experiments were carried out to determine the transcription start site. A primer directed 3' to 5' from nucleotides -75 to -95 was end labeled with $^{32}$P and used in the extension experiment to determine the extent of the 5' end of the hnRNA transcript. By electrophoresing the primer extension reaction out along side the products of a known sequencing reaction the exact number of base pairs in the extended product was determined (see Figure 4). When the primer extension product length was compared to the 5' clone sequence, a transcription start site was mapped 114 bases upstream of the signal peptide initiator methionine.

Two other oligonucleotides (5'.1 and 17.1) were used in primer extension experiments. The majority of the extensions terminated at position -95, thereby making the reading of these primer extension
Figure 4

Autoradiography of Primer Extension of Ceruloplasmin mRNA
Top panel: The primer extension was primed with the oligonucleotide #23.1. The arrow marked A points to the single strand DNA product of the extension. The product was electrophoresed in a 6% polyacrylamide, 8M urea gel along side a known DNA sequencing reaction to determine the exact number of nucleotides in the product.
Bottom panel: The reaction was primed with the oligonucleotide #17.1. The arrow marked B points to the largest single stranded product. The left side of the panel shows end labeled marker (bp)
Figure 5

Sequence of 5' cDNA Clone of Human Ceruloplasmin

The horizontal arrow shows the position of the 23.1 oligonucleotide primer. The ATG codon for the initiator methionine is underlined. The +1 denotes the codon for the first amino acid of the mature protein. The vertical arrow identifies the transcription start site.

1 - CACTGAAATTGGAAAAATATTGCTTTAATGAAACAAATTACTCTTTGTGCA
49-ACACTAAATTGTGTCATCAATCAAGCAAAATAGGAAGAAAGTCTTATTTAT
97-AAAATTGCCTGCTCCTGAIii i ACTTCATTTCTTCTCAGGCTCCAAGAA
146-GGGGAAAAAAATGAAGATTTTGATTTGGTAiiiii CTG I I Hi ATGT Met
195-AGTCCCCAGCCTGGGCGAAAGAAAGCATTATT
Lys Glu Lys His
+1
experiments extremely difficult. This position represents the longest previously isolated 5' cDNA clone. Secondary structure in the RNA at this position was probably responsible for poor primer extensions.
3. 3' cDNA Clones

First strand cDNA from total human liver RNA was size separated by electrophoresis on an agarose gel. The gel was cut into slices corresponding to 200 base pair increments of the marker ladder. The single stranded DNA from these slices was used as template in the PCR. Using the T17XSP and 21.1 (corresponding to cDNA positions 3129-3150) oligonucleotides in the PCR produced a bank of clones. All the clones were analyzed for their insert size by restriction enzyme digestion and gel electrophoresis. The identities of ceruloplasmin cDNA clones were confirmed by DNA sequence analysis of their 5' ends. Figure 6 shows the sequences of the 3' clones compared to the published sequence of ceruloplasmin. All the clones share 100% sequence identity with ceruloplasmin until nucleotide 3298 of the published sequence (Koschinsky et al., 1986). All the clones share near 100% sequence identity with each other. However the clones share no sequence homology with the published ceruloplasmin 3' untranslated sequence after nucleotide 3298.

A. 59 Clone

The cDNA clone, designated 59, was 710 nucleotides in size. The 5' end of the clone contained the last 71 nucleotides of coding sequence, the stop codon and 78 nucleotides of 3' untranslated sequence. Following the 78bp the sequence diverges from previously reported 3' untranslated sequence (Koschinsky et al., 1986; Yang et al., 1986) and the 3' genomic sequence.
**Figure 6**

Comparison of the 3' ends of different Ceruloplasmin cDNA clones.

The hatched bars represent the segments of DNA sequence that are identical to the 3' end of the genomic sequence. The vertical line represents the position of the stop codon in the gene and cDNA clones. The numbers represent the points of sequence divergence from the Koschinsky cDNA sequence. The position of polyadenylation is shown as AAAAA.

The light and dark bars represent unmapped, unique sequences.
59 Clone Sequence

The arrow indicates the point of divergence from the genomic sequence. The stop codon of the cDNA and the restriction sites used to produce the probe for the Northern and Southern blots are underlined. The polyadenylation signals and the GT region are in bold print.
1- GACCACATTCTGCTGGGAATGGAAACCTTTACCCGGTTCTACAAAAATG

51- AAGACACCGGTCTGGCTGAATGGAATAAAAATTTGATATAAGTGAAAAAAA
STOP

101- GAGAAAAACCAATGATTCTACATAACAAATGTATGTGATGAAGATGTAATAATAGAAT

151- GCTACTTTGGGATAGCGCTATAACCTTTTTTAAAAAAAGACTGAAATGCACG
Dral

201- TTTTACATTTTGGTGCAGAAACTGTCACACTCTTCTATACAAAGTCAAT

251- AGACTGTATTTTGGACAAACTGACATATCCTGTAATATAAATCTGA

301- TTAATATAAGTGATTCATATCACCACATATTAGGTAGGCTTACTAAAACT

351- GTATTACAAAAATAATCTATGACTTAGTGGGCTCTGGACAGTCCUTAAA

401- GGTACATCTATTTAAGATTTCAATTCAGATTTTCAGAAATAAAAATTAATAAGATTTCA

451- TGAAAATTTTTCAATCTTTGCTCATTATTATCTATGTACTTACTG

501- ACTTCATTCTAATGGACCAGTGGGAGAGGCGATAGATGTAGAGGAAG

551- TAAAATAAAAATCTAACAAGACAAAAATAAAATAATAAAAAATTTCAAAAACA

601- GTGAGACCAACATTGTAACTTTCTAATTAAAAACTATAAGAAAAATTTGTGT

651- GTTTTGAGCCACATATTGATTGGTTTACAGCAGCAATAAAAAAA
HindIII

701- TCAATAAGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Analysis of the entire 59 clone showed the presence of 6 distinct polyadenylation signals. The signal ATAAAA was identified as the putative polyadenylation signal.

A probe was produced from the 3' end of the 59 clone spanning from the DraI site (177) to the HindIII (677). This probe was labeled with $^{32}$P using Klenow large fragment polymerase and hybridized to a total human RNA Northern blot and two genomic Southern blots. The Northern blot showed a single band at 4.4 kb which hybridized to the probe (see Figure 9). A control Northern blot was analyzed with a 3' cDNA probe (EcoRI fragment 2450-2850), and showed both the 3.7 and 4.4 kb bands of ceruloplasmin.

The genomic Southern blots contained both EcoRI and EcoRI/HincII digested DNA. The Southern blots were hybridized and washed at high and low stringencies. The high stringency Southern blot did not show any hybridizing bands. The low stringency Southern showed a smear over several kilobases, but no clear bands. This seemed to indicate the probe had homology to many different genomic sequences and therefore was not useful for Southern analysis.
Figure 8

Sequence Alignment of the 59 Clone with the 3' cDNA

The sequences of the 5' end of the 59 clone (59) and the 3' end of the ceruloplasmin cDNA (3P) are shown. Mismatches are marked with a star (*). The TGA stop codon is in bold print. The underlined sequence is the putative 5' splice junction consensus sequence.

59 GACCACATTCATGCTGGAATGGAAACCACTTTACACCGTTCTACAAAAATG
3P GACCACATTCATGCTGGAATGGAAACCACCTTTACACCGTTCTACAAAAATG

59 AAGACACAAATCTGGCTGAATGAAATAAATTGGTGATAAGTGGAAAAA
3P AAGACACAAATCTGGCTGAATGAAATAAATTGGTGATAAGTGGAAAAA

59 AGAGAAAAACCAATGATTCATAACAATGTATGTGAAAGTGTAAAATAGA
3P AGAGAAAAACCAATGATTCATAACAATGTATGTGAAAGTGTAAAATAGA

59 ATGTGTAATAGAATGCTACTTTTGGAAATGGCTATAACC
3P ATGTGTAATAGAATGCTACTTTTGGAAATGGCTATAACA
Northern blot analysis of the 3' end of the ceruloplasmin gene.

Panel A: The sample of total human liver RNA (10μg) was electrophoresed in a denaturing agarose-formaldehyde gel and transferred to nitrocellulose. The filter was hybridized with a $^{32}$P-labeled DraI/HindIII fragment derived from the 3' end of the 59 clone.

Panel B: The probe used in the hybridization was a 450bp EcoRI fragment derived from the 3' end of the coding cDNA. The arrows show the relative sizes of the bands compared to RNA standards.
B. Screen of λgt11 Library for 3’ Clones

A HepG 2 cDNA library was screened for a 3’ cDNA clone. The primary screen used a 450 base pair EcoRI fragment from the 3’ end of a previously isolated cDNA clone to probe a λgt11 library. The second screen was performed by PCR using an oligonucleotide 2986-3006 and a T17XSP oligonucleotide. The resulting product was cloned into the Bluescript vector and named λgt-11.14. When the clone was sequenced, complete agreement with the published cDNA clones was found in the coding region and 3’ UTS until nucleotide 3338. The sequence of λgt-11.14 diverged from all known sequences, including those isolated and reported in this thesis, at this point. This clone continued past the polyadenylation site utilized in the Koschinsky clone but diverged prior to the polyadenylation site utilized by the Bowman clone. The clone continued for another 110 base pairs before its poly A tract (see Figure 6). A probe from the 3’ end of this clone was labeled and used as a probe against the 3’ genomic clone which contained 5 kb of sequence beyond the last coding exon. There was no homology between the probe and the 3’ genomic fragment.
Discussion

1. Ceruloplasmin Gene Structure

Previous work by Koschinsky (1986) isolated and characterized nine recombinant phage clones that contained the first 14 exons of the ceruloplasmin gene. These clones spanned 30 kbp of genomic DNA; the remaining exons were isolated in one phage clone that spanned 13 kbp. This phage clone contained the last 5 exons within 8 kb and 5 kb of 3' genomic sequence. The exons range in size from 129bp-255bp, with a calculated median size of 175bp. This is consistent with published exon size distributions (Naora and Deacon, 1982). The intronic regions of the gene comprise more than 34 kb, and vary in size from 0.8kbp to over 9kbp. Although the intronic region size is larger than predicted by Naora and Deacon, the intron/exon size ratios resembles those of the two related genes of factor V and VIII as they have a high percentage of the gene consisting of introns. All the splice donor/acceptor sequences conform to the GT/AG rule for nucleotides immediately flanking the boundaries (Breathnach and Chambon, 1981; Cech, 1983). This splice junction consensus sequence has recently been modified to include the GC/AG variation (Shapiro and Senepathy, 1987).
2. **Comparison of Ceruloplasmin Gene with factors V and VIII**

The ceruloplasmin gene spans approximately 38 kbp of chromosome 3. This is smaller than the genes of either factor V (70 kbp, Cripe et al., 1992) or factor VIII (>180kbp, Gitschier et al., 1984). One of the major differences is the presence of a large B domain in both factor V and factor VIII. The 30% identity of the A domains of ceruloplasmin with those of factors V and VIII (Ortel et al., 1984), supports the theory that the three genes evolved from a common ancestral gene. Comparisons of the intron/exon junctions in the A domains of the three genes showed ceruloplasmin shares only moderate similarities with the junctions of factors V and VIII, which share 19 of 24 positions (see Figure 10). This would suggest that the ceruloplasmin gene diverged from the ancestral gene first. The duplicated C domains of factors V and VIII share approximately 40% identity, while the B domains share no homology also suggesting the addition of the C domains to the genes occurred after ceruloplasmin diverged and before the divergence of factors V and VIII from each other. The lack of homology in the B domains suggest that the insertion events of the B domains occurred after the two clotting factors diverged. Both B domains are each encoded by a single exon, suggesting the B domains were produced by reverse transcription of mRNA molecules (Cripe et al., 1992). If the B domains are of different origins, an interesting question arises. Why did the two different B domains insert into the same location of the two genes? Is there a sequence at this site that favors insertion of DNA fragments?
Comparison of the domain structures and intron/exon junctions of ceruloplasmin, factor V and factor VIII.

The boxes represent the domain structures of the three proteins and are drawn to scale (one inch equals 100 amino acids). The three different domains are numbered. The arrows represent the positions of the intron/exon junctions.
3. 5' cDNA Clone

Isolation of 5' cDNA clones that contain a transcription start site has been difficult. Marlys Koschinsky isolated several clones that contained 38 nucleotides upstream to the codon for the initiator methionine (Koschinsky, 1986). Evidence from her thesis suggested that there were more nucleotides in the 5' untranslated transcript. The isolation of a clone from a HepG 2 λgt11 expression library contained an additional 82 base pairs of 5' untranslated sequence. The primer extension experiment with the 23.1 oligonucleotide identified a transcription start site 114 base pairs 5' to the initiator methionine codon. Another primer extension experiment primed with an oligonucleotide further 3' (110-90) showed a product that would place its start site at or near the above mentioned site. Difficulties in interpretation of primer extension results and inability to isolate full length 5' clones due to premature termination during first strand synthesis probably reflect secondary structure in the 5' untranslated region (see Figure 11). The use of a primer directly adjacent to the putative secondary structure gave a single clear product that allowed identification of a transcription start site.

The human 5' untranslated sequence is much longer than the 30 base pair rat ceruloplasmin 5' untranslated sequence (Fleming and Gitlin, 1992). Analysis of the two sequences showed no homology (data not shown). The transcription start sites also begin at different nucleotides. The more common guanosine nucleoside is the first nucleoside in the human sequence, while an adenosine is the first nucleoside in the rat transcript. Another difference between the two UTS's is the discontinuous nature of the human 5' UTS.
Proposed Secondary Structure of 5' RNA
A proposed secondary structure of the 5' RNA was generated by the method of Zuker in PC/Gene. Gaps were inserted (-) to generate the secondary structure with the highest free energy of formation. The free energy of formation of the structure is given in kilocalories (kcal).

Energy of structure = -32.4 kcal

---CACUGAA AA UUA A UCUA AU GA ACAAUUUA UG
UUGGA AUA UGCUU AU GA ACAAUUUA UG
AACC UAU ACGAA UA CU UGUUAAAUA AC C
--- C A G --- C AA

54
At least one splice site has been demonstrated by comparison of the cDNA and genomic sequences (Koschinsky, 1986). Until a large 5' genomic clone from the human gene is isolated and sequenced, the exact number of splices involved in producing the human transcript(s) remains unknown.

4. 59 Clone

The longest 3' untranslated clone, designated 59, has mapped and unmapped segments. The 5' end of the clone agrees with the coding sequence of the cDNA that encodes 1046 amino acids. No clone was isolated that encoded the additional 4 amino acids reported by Yang et al. The 59 clone includes 71 bases of coding sequence and 78 bases of identical 3' untranslated sequence to ceruloplasmin. The 59 clone diverges from this point on.

The 59 clone contains 7 polyadenylation signals, however no "GU" polyadenylation selection signals are present 3' to the first 4 polyadenylation signals. The polyadenylation signal used may use the "GU" signal 36 bp upstream but until the genomic sequence that generated this sequence is cloned and characterized this can not be confirmed.

5. Mechanisms of 59 Production

There are several possible origins of the 59 cDNA clone: (i) the ceruloplasmin gene is a tandem duplicate gene pair, (ii) the 59 clone is a product of trans-splicing or (iii) the unmapped sequence originates from cis-splicing of a larger heteronuclear transcript that
extends beyond the 3' end of the presently characterized genomic clone. Each of these possibilities has merits.

The evolution of the ceruloplasmin gene involved the process of gene duplication. If a further duplication event occurred to produce two tandem ceruloplasmin genes, then the 59 clone may have originated from the second gene. A problem with this hypothesis is the requirement for preservation of the restriction patterns in the genes. Many of the restriction sites used to subclone and map the gene fall within introns. Aside from splice junction and branchpoint consensus sequences, there appear to be no functional constraints to conserve these restriction sites over time. Therefore, the possibility of a second, identical gene is doubtful.

Trans-splicing has recently been reported in mammals (Fu and Maniatis, 1992). With this recent finding, the possibility that the 59 clone is a product of trans-splicing must be considered. Most, if not all, reported trans-spliced transcripts involve the splicing of a short 5' leader or coding sequence to a much longer 3' RNA transcript. The production of the 59 clone by trans-splicing would involve the splicing of a long 5' cis-spliced RNA molecule to a 3' untranslated RNA transcript. This discrepancy in splicing mechanism argues against trans-splicing involvement in the production of the 59 clone.

The possibility that the 59 clone was derived from cis-splicing of a much larger hnRNA transcript that comes from the characterized ceruloplasmin gene and additional 3' genomic sequences seems the most plausible. Sequencing of 600 base pairs of the 3' untranslated genomic region showed several features that argue against this hypothesis. Approximately 400 base pairs downstream of the stop
codon a tandem series of 5 polyadenylation signals followed by a "GT" rich region suggest the primary transcript would be cleaved at this point (Humphrey and Proudfoot, 1988). Southern blots of the remaining 4.5kbp of 3' genomic DNA showed no homology with the 3' end of the 5' clone.

Isolation of genomic clones that continue along the 5' and 3' directions of the chromosome would give evidence to support the duplicate gene hypothesis or the extended 3' alternate splicing hypothesis. Support for the trans-splicing hypothesis would involve cloning the genomic region responsible for the 3' end of the 5' clone and mapping it to another genomic location.

6. Northern Blot Analysis

The presence of only the 4.4 kb transcript hybridization with the 3' end of the clone suggests that this clone contains the 3' end of only the larger of the two ceruloplasmin transcripts. Although ceruloplasmin has two different length transcripts, the larger 4.4 kb transcript has only been reported in the liver and choroid plexus. The reasons why this transcript is only produced in these tissues may involve tissue specific expression and or post-transcriptional regulation. There are numerous reports of 3' untranslated sequence involvement in gene expression regulation. The function of the larger 3' untranslated sequence is unknown at present.
7. **Southern Blot Analysis**

The probe that hybridized to the 4.4 kb RNA transcript did not hybridize to any unique DNA sequence. The Southern blot produced under low stringency was smeared suggesting there are several to many sequences that are homologous to the probe in the genome. However, because only one band was seen on the Northern blot, most of the sequences that are homologous are not transcribed. The sequence of the genomic 3' region shows a series of 5 tandem polyadenylation signals followed closely by a "GT" secondary poly A signal. The presence of these strong transcription termination signals suggests that the 5' clone did not come from alternative splicing of a long 3' untranslated sequence that includes the immediate 3' sequence.

8. **3' λgt11 cDNA Clone**

The sequence comparison of the λgt-11.14 clone isolated from the HepG 2 library showed it was completely identical with the published sequences of the 3' coding region of ceruloplasmin. The sequence identity continued beyond the polyadenylation site of the Koschinsky clone (3321) but diverged at nucleotide 3338 which is prior to the polyadenylation site of the Bowman clone. The 3' region of the clone was not homologous to any of the reported cDNA clones or the 3' genomic sequence. The point of divergence was at a consensus 5' splice junction sequence. The presence of the splice junction implies that the clone was spliced differently than any previously isolated clones. The location of the 3' splice acceptor site has not been determined. This suggests the unique sequence is a
result of differential splicing of a hnRNA molecule. Another possibility as to the origin of the clone may involve the transformed nature of the cell line used as the source of the RNA. The process of transformation of a cell into an immortal cell line undoubtedly changes many of the normal cellular functions. The production of this clone being a result of altered cellular splicing functions cannot be ruled out at this time.

9. Other 3' Clones

The procedure that produced the 59 clone also produced a series of shorter clones. All of these clones encoded the 3' region of the coding cDNA and differing lengths of 3' untranslated sequence. Analysis of the sequences showed complete identity of all the clones up to the polyadenylation site. All the clones also diverged from the published 3' untranslated sequences at the same nucleotide (data not shown). These results imply that the differential splicing of the 59 clone was not a procedural artifact. It is interesting that none of these shorter clones were spliced similar to the λgt-11.14 clone or published cDNA clones. Again, the location of origin of the 3' end of these clones is unknown at present.

10. Future Studies

Isolation of the human 5' genomic flanking region will permit a variety of experiments. S1 mapping and primer extension experiments will determine the transcription start site(s). Characterization of this genomic DNA will help to identify promoter
and enhancer sequences and trans-acting factors involved in the different forms of induction regulation.

Utilizing PCR technology, an evolutionary survey of the vertebrata subphylum with respect to ceruloplasmin may help to determine how and when the genes of factors V, VIII and ceruloplasmin diverged from each other.

Cloning of the 3' flanking region of genomic DNA will help to determine if the human ceruloplasmin locus is one gene or a tandem duplication and the source of the 59 series and λgt1 clones. A series of experiments, with the 3' end of the 59 clone ligated to a reporter gene, could be designed to explore the different possible forms of gene expression regulation. The 3' UTS may be involved in regulating tissue specific, developmental specific and induction specific expression. Specific experiments can be designed to explore each of these different forms of expression regulation.

**Conclusion**

Ceruloplasmin is a 132Kda serum protein found in the serum of all vertebrates. The protein is multi-functional. Functions include copper transport, serum and tissue anti-oxidant and anti-coagulation modulation. Expression of the gene is highly regulated. There are tissue, developmental and induction specific forms of regulation.

The human ceruloplasmin gene spans 38 Kbp of genomic DNA. The gene is composed of 19 coding exons separated by 18 introns. The exons vary in length from 129 to 255 bp. The introns vary in length from 400 bp to over 9.5 Kbp. The gene structure is similar to factors V and VIII. The divergence of the ceruloplasmin gene from the
common ancestral gene probably occurred prior to the divergence of the genes of factors V and VIII.

Post-transcriptional regulation has also been demonstrated. It has been proposed that both the 5' and 3' untranslated regions are involved in some forms of regulation. Liver produces two transcripts, 3.7 and 4.4 kb. A 5' cDNA clone has been isolated that contains 114 bp of untranslated sequence prior to the initiator methionine. This cDNA clone was produced by splicing of the hnRNA. The 5' end of the clone remains unmapped at the present. A series of 3' cDNA clones have been isolated. Again, splicing of the untranslated sequences was involved in their production and their 3' ends are unmapped at the present. One clone contains a 500 bp sequence that hybridizes only to the larger 4.4 kb transcript. Evidence from this thesis and the Koschinsky thesis suggest splicing in both the 5' and 3' untranslated regions contribute to the variations in transcript size. The role of the 5' and 3' untranslated sequences in expression regulation is unknown at the present.
References


