Isolation and Characterization of a Processed Gene for Human Ceruloplasmin†

Marlys L. Koschinsky,‡ Billy K.-C. Chow,‡ Jacquie Schwartz,§ John L. Hamerton,§ and Ross T. A. MacGillivray*,‡
Department of Biochemistry, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada, and Department of Human Genetics, University of Manitoba, Winnipeg, Manitoba R3E 0W3, Canada
Received April 8, 1987; Revised Manuscript Received July 7, 1987

ABSTRACT: A processed pseudogene for human ceruloplasmin has been isolated that contains DNA corresponding to the functional gene sequence encoding the carboxy-terminal 563 amino acid residues and the 3′ untranslated region. The pseudogene appears to have arisen from a processed RNA species, since intervening sequences coincident with those of the functional gene have been removed, with the exception of a short segment of intronic sequence which denotes the 5′ boundary of the pseudogene. The nucleotide sequence of the pseudogene is highly homologous (97% sequence identity) with that of the wild-type gene, suggesting that pseudogene formation was a relatively recent evolutionary event. In addition to single base substitutions, there is a large 213 base pair (bp) deletion in the pseudogene sequence which corresponds to the location of an intron–exon junction in the functional gene. A 4 bp duplication that occurs at amino acid residue 683 of the wild-type coding sequence results in a frameshift mutation and introduces a premature translational termination codon at this point. This is concordant with the inability to detect a human liver transcript corresponding to the pseudogene by nuclease S1 mapping analysis. The 3′ end of the pseudogene is characterized by a 62 bp segment composed mainly of repeated TC dinucleotides. On the basis of genomic Southern blot analysis performed under high-stringency conditions, the pseudogene that we have identified seems to comprise the only sequence in the human genome that is closely related to the wild-type gene. Using somatic cell hybridization, we have mapped the pseudogene to human chromosome 8. This differs from the site of the wild-type ceruloplasmin locus, which has been assigned to chromosome 3.

Higher eukaryotic genes commonly exist in multigene families that include both functional genes and also closely related sequences that appear to be inactive and are therefore termed pseudogenes (Jacq et al., 1977). Two distinct categories of pseudogenes have been identified. The first category includes those genes that have retained the intervening sequences found in their functional counterparts [e.g., see Lacy and Maniatis (1980) and Cleary et al. (1981)]. Frequently, these pseudogenes are located adjacent to the respective wild-type gene, suggesting that they have arisen from a gene duplication event. The second, more abundant category of pseudogenes contains processed genes [see Vanin (1985) for a recent review]. These genes are characterized by the precise loss of introns, suggesting that they are derived from processed RNA molecules after reverse transcription and insertion into the genome. This model is supported by the fact that processed pseudogenes and their functional counterparts are not located on the same chromosome (Battey et al., 1982; Czosnek et al., 1984). In the majority of cases, sequence homology between processed pseudogenes and the functional genes ceases at points corresponding to the beginning and the end of the functional transcripts. Exceptions to this include the human immunoglobulin ϵ (Ueda et al., 1982) and $\lambda\psi 1$ pseudogenes (Hollis et al., 1982) and the mouse corticotropin β-lipotropin precursor pseudogene (Notake et al., 1983) which appear to be DNA copies of only a portion of the wild-type mRNA transcripts.

Ceruloplasmin [ferroxidase; iron(II):oxygen oxidoreductase, EC 1.16.3.1] is a glycoprotein from the α_2-globulin fraction of vertebrate plasma (Frieden, 1981). This protein is synthesized in the liver as a single polypeptide chain of $M_r \simeq 132\,000$ (Kingston et al., 1977; Takahashi et al., 1984) and is the principal copper-transport protein in plasma, binding 90–95% of blood copper in vertebrates (Frieden, 1981). The ceruloplasmin gene has been mapped to human chromosome 3q25 by using both somatic cell hybridization and in situ hybridization (Yang et al., 1986; Royle et al., 1987). This same region of chromosome 3 also contains a linkage group consisting of transferrin (Yang et al., 1984), pseudocholinesterase 1 (Robson et al., 1966; Eiberg & Mohr, 1979), and the gene for α2-HS glycoprotein (Eiberg et al., 1983).

We have recently reported the complete cDNA sequence of human preceruloplasmin (Koschinsky et al., 1986). In this paper, we describe the isolation and characterization of a human ceruloplasmin pseudogene. In addition, we have mapped the pseudogene to chromosome 8 by using somatic cell hybridization.

Materials and Methods

Materials

All restriction endonucleases and DNA-modifying enzymes were purchased from Bethesda Research Laboratories, Pharmacia-PL Biochemicals, New England Biolabs, or Boehringer Mannheim and were used according to the manufacturers' recommendations. ^{32}P-Labeled nucleotides were purchased from Amersham. Nitrocellulose was obtained from Schleicher & Schuell, and Zetaprobe was purchased from Bio-Rad Laboratories.

Methods

Human Genomic Libraries. A human genomic phage library constructed in Charon 4A (Lawn et al., 1978) was generously provided by Dr. T. Maniatis and amplified as described (Maniatis et al., 1982). In addition, a human lymphocyte genomic library in λEMBL 3 (Geddes, 1987) was provided by Val Geddes (University of British Columbia) and was screened prior to amplification.

† This work was supported in part by grants from the British Columbia Health Care Research Foundation (to R.T.A.M.), the Medical Research Council of Canada (MA-4061 to J.L.H.), and the Children's Hospital of Winnipeg Research Foundation Inc. (to J.L.H.). M.L.K. and B.K.-C.C. were supported by Studentships from the Medical Research Council of Canada.
‡ University of British Columbia.
§ University of Manitoba.

0006-2960/87/0426-7760$01.50/0

Screening of Human Genomic Phage Libraries. Five genome equivalents [1×10^6 plaque-forming units (pfu)[1]] of both human genomic phage libraries were screened by plaque hybridization (Benton & Davis, 1977; Woo, 1980) using human ceruloplasmin cDNAs (Koschinsky et al., 1986) as hybridization probes. Probes were labeled with ^{32}P by using random-sequence hexadeoxyribonucleotides as primers for elongation with the Klenow fragment of DNA polymerase I (Feinberg & Vogelstein, 1982). Positive phage were plaque-purified, and DNA was extracted according to the procedure of Maniatis et al. (1982).

DNA Sequence Analysis. A 1.9 kbp *Bam*HI–*Hin*dIII restriction fragment containing DNA sequences hybridizing to the ceruloplasmin cDNA probe phCP-1 (Koschinsky et al., 1986) was isolated from the pseudogene clone 3ψ10 (see Figure 1) and subcloned into pUC12 for analysis. DNA sequence analysis of this subclone (designated pψ10) was carried out essentially as described by Deininger (1983). Plasmid DNA was randomly sheared by sonication. Fragments ranging from 300 to 500 bp in length were recovered by electroelution from a 5% polyacrylamide gel, and the ends were made blunt by using T4 DNA polymerase. These fragments were then ligated into the *Sma*I site of M13mp18 and used to transform *Escherichia coli* strain JM103 (Messing, 1983). Subclones containing human ceruloplasmin pseudogene inserts were identified by hybridization to the 1.9 kbp insert isolated from the pψ10 subclone. The DNA sequence of an overlapping 0.8 kbp *Eco*RI fragment, isolated from the 3ψ10 phage clone, was also determined. All DNA sequence analysis was performed by using the chain termination method (Sanger et al., 1977). DNA sequence data were analyzed by the DBUTIL program of Staden (1982).

Genomic Southern Blot Analysis. Human genomic DNA was isolated from liver by the method of Blin and Stafford (1976). Samples containing 10 μg of DNA were digested with *Eco*RI. The fragments were separated by electrophoresis on 1% agarose gels and transferred to nitrocellulose by the method of Southern (1975). The cDNA clones (Koschinsky et al., 1986) corresponding to nucleotide residues 1–1200 (λhCP-1) or residues 660 through the poly(A) tail (phCP-1) were used as hybridization probes. These cDNA clones were labeled to specific activities of 1×10^8 cpm/μg either by nick translation (Maniatis et al., 1975) or by random priming with synthetic oligonucleotides and the Klenow fragment of DNA polymerase I as described above. Hybridization and washing conditions were performed as described by Kan and Dozy (1978).

Nuclease S1 Mapping Analysis. A single-stranded probe corresponding to nucleotide residues 1927–2213 of the human ceruloplasmin cDNA was prepared from a recombinant M13 template as described (Kay et al., 1986). For the nuclease S1 protection assay, 100 000 cpm of probe (specific activity 10^8 cpm/μg) was mixed with 1 or 0.35 μg of human liver poly(A+) RNA, which was isolated by the guanidine hydrochloride method (Chirgwin et al., 1979) followed by chromatography on oligo(dT)–cellulose (Maniatis et al., 1982). Subsequent hybridization and nuclease S1 digestion were performed according to the conditions described by Kay et al. (1986). Reaction products were separated by electrophoresis on a 6% denaturing polyacryiamide gel and were visualized by autoradiography.

Chromosome Mapping. Chromosome localization of the human ceruloplasmin pseudogene was performed by using human–hamster somatic cell hybrids which have been previously characterized by cytogenetic and isozyme analysis (Donald et al., 1983; Riddell et al., 1985). DNA from cultured cell lines was isolated as described (Riddell et al., 1986). DNA (5 μg) from each hybrid line, as well as control human placental and hamster DNA, was digested with *Eco*RI, electrophoresed on 1% agarose gels, and transferred to Zetaprobe according to Southern (1975). A unique 1.1 kbp *Eco*RI insert derived from the 3ψ10 pseudogene phage clone was subcloned into pUC12 and isolated for use as a hybridization probe (designated probe A). Probe A was labeled with ^{32}P by nick translation as described above to a specific activity of 7×10^8 cpm/μg. Hybridization was carried out overnight at 42 °C in 50% formamide, 3× SSPE, 1% $NaDodSO_4$, 0.5% nonfat powdered milk, 10% dextran sulfate, 200 μg/mL salmon sperm DNA, and labeled probe at a concentration of 20 ng/mL. The blots were washed twice in 2× SSC at room temperature for a total of 10 min. They were then washed for 15 min in 2× SSC and 0.1% $NaDodSO_4$ at 55 °C and finally for 15 min in 0.2× SSC and 0.1% $NaDodSO_4$ at 55 °C. Filters were then exposed to film for 18 h at −70 °C with intensifying screens.

RESULTS

Screening of Human Genomic Libraries. The Maniatis human genomic phage library was screened with the phCP-1 cDNA clone (Koschinsky et al., 1986) as a hybridization probe. Two independent positive clones were isolated. One of these clones (designated 3ψ10) corresponded to a pseudogene for human ceruloplasmin. This was initially determined by restriction enzyme mapping and Southern blot analysis (data not shown) and subsequently confirmed by using DNA sequence analysis and nuclease S1 analysis (see following sections). The other phage contained part of the wild-type ceruloplasmin gene (M. L. Koschinsky, unpublished results). To obtain additional pseudogene clones, 10^6 phage from a second genomic library (provided by Val Geddes) were screened with both λhCP-1 and phCP-1 cDNA clones (Koschinsky et al., 1986) as hybridization probes. From this screen, 11 different clones were obtained. Eight of these clones were from the wild-type ceruloplasmin locus while three clones, designated 3ψ21, 3ψ29, and 3ψ9, were identified as human ceruloplasmin pseudogene clones. Restriction enzyme analysis showed that these phage clones overlapped the 3ψ10 clone isolated from the Maniatis library (Figure 1). A total of ~21 kbp of continuous genomic DNA is represented by these four phage clones. The human ceruloplasmin pseudogene maps to a region of ~1.7 kbp which is also shown in Figure 1.

DNA Sequence Analysis. The nucleotide sequence (see Figure 2) of the human ceruloplasmin pseudogene was determined by using the strategy shown in Figure 1. Each nucleotide was determined an average of 3.4 times, and 54% of the sequence was determined on both strands. Nucleotides 53–1644 of the pseudogene sequence are very similar to nucleotides 1502–3318 of the ceruloplasmin cDNA, extending through the region corresponding to the 3′ untranslated region of the phCP-1 clone (Koschinsky et al., 1986). The pseudogene, however, is not characterized by a poly(A) tract at the expected polyadenylylation site. DNA sequence analysis in a 3′ direction revealed the presence of an unusual 54 bp segment, composed primarily of repeated CT dinucleotides (nucleotides 1867–1920, Figure 2).

Following the sequence corresponding to the 3′ untranslated region of phCP-1, the next 43 bp of pseudogene sequence (nucleotides 1645–1687) correspond to the 3′ untranslated

[1] Abbreviations: pfu, plaque-forming unit; bp, base pair(s); kbp, kilobase pair(s); SSPE, 0.01 M sodium phosphate buffer, pH 7.4, 0.15 M NaCl, and 0.001 M EDTA; SSC, 0.015 M sodium citrate buffer, pH 7.0, and 0.15 M NaCl; $NaDodSO_4$, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid.

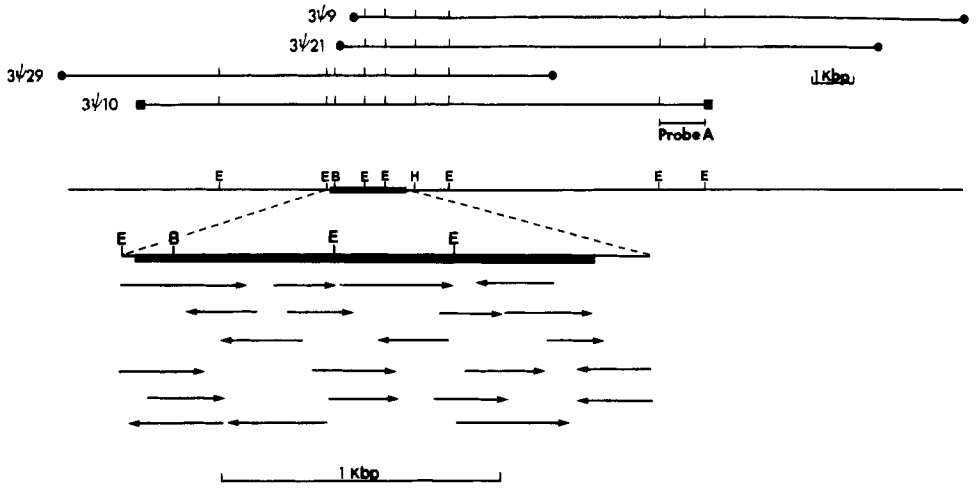

FIGURE 1: Partial restriction map and sequencing strategy for the human ceruloplasmin pseudogene. The complete restriction map for *Eco*RI (E) is shown. The lines above the restriction map represent the four overlapping phage clones 3ψ9, 3ψ21, 3ψ29 (isolated from the Geddes library), and 3ψ10 (from the Maniatis library). The solid bar represents the ceruloplasmin pseudogene that is homologous to the ceruloplasmin cDNA sequence. Solid circles at the ends of EMBL3 phage clones represent *Sau*3A sites, while solid squares at the end of the 3ψ10 phage clone isolated from the Maniatis library represent *Eco*RI linkers. The locations of the *Bam*HI (B) and *Hin*dIII (H) sites used for subcloning and sequence analysis are shown within the 3ψ10 clone; the remainder of the map is incomplete for these two enzymes. The region containing the *Bam*HI/*Hin*dIII fragment has been expanded below the restriction map. Arrows below this region indicate the direction and extent of nucleotide sequence obtained from various M13 clones. Probe A was used as a hybridization probe for chromosomal localization studies (see text). The scale represents 1 kbp.

sequence of the ceruloplasmin cDNA clone (designated CP-1) characterized by Yang et al. (1986). This clone differs from the phCP-1 clone in that the 3′ untranslated region extends for a further 120 bp prior to the site of poly(A) addition. Following this 43 bp segment, the remainder of the pseudogene sequence prior to the poly(CT) segment shares little homology with the 3′ untranslated region from the cDNA clone described by Yang et al. (1986).

A comparison of the nucleotide sequences of the pseudogene and ceruloplasmin cDNA (Koschinsky et al., 1986) is also shown in Figure 2. A deletion of 213 bp was observed in the pseudogene sequence, corresponding to nucleotides 1865–2077 of the human ceruloplasmin cDNA sequence (see Figure 2). Prior to this deletion, the pseudogene sequence contains an open reading frame corresponding to amino acid residues 483–602 of the ceruloplasmin coding sequence. Occasional base changes within this open reading frame result in the occurrence of amino acid substitutions (see Figure 2). The 213 bp deletion in the pseudogene sequence maintains this open reading frame, which then continues for an additional nine amino acid residues. At this point, an insertion of four duplicated nucleotide residues (AGCT) (nucleotides 445–448, Figure 2) causes a frameshift mutation, such that a premature TGA termination codon occurs immediately following this insertion. The remainder of the pseudogene sequence is homologous to the phCP-1 cDNA clone and the cDNA clones characterized by Yang et al. (1986) but contains a number of single base substitutions when compared with the cDNA sequence. There is also a small deletion of 17 bp corresponding to nucleotide residues 2941–2958 of the cDNA (see Figure 2). However, this does not result in the resumption of an open reading frame in the pseudogene sequence.

The 5′ end of the pseudogene sequence (nucleotides 10–52, Figure 2) corresponds to the intronic sequence found in the wild-type gene (M. L. Koschinsky, unpublished results). A consensus 3′ splice acceptor sequence (Cech, 1983) is located immediately prior to nucleotide 53 (see Figure 2). This also corresponds to the location of an intron–exon junction in the wild-type gene (M. K. Koschinsky, unpublished results). On the basis of DNA sequence analysis, nucleotide residue nine of the pseudogene marks the point of divergence from the wild-type gene. Southern blot analysis confirms that the 2.6 kbp *Eco*RI fragment, located directly 5′ to the 0.8 kbp *Eco*RI fragment (see Figure 1), does not hybridize to the corresponding wild-type gene (data not shown). In contrast to most processed pseudogenes (Vanin, 1985), the points of divergence of the ceruloplasmin pseudogene are not flanked by short direct repeats. Accordingly, the particular integration event that gave rise to the ceruloplasmin pseudogene is unclear.

Genomic Southern Blot Analysis of the Human Ceruloplasmin Gene. To determine whether the human genome contained pseudogene sequences corresponding to the 5′ end of the wild-type ceruloplasmin gene, human liver DNA was cleaved with *Eco*RI and subjected to Southern blot analysis using either the phCP-1 or the λhCP-1 cDNA clone as hybridization probe. When phCP-1 (which contains sequences corresponding to the 3′ portion of the ceruloplasmin cDNA) was used as a probe, several hybridizing DNA fragments were observed (Figure 3A). The 0.45, 0.8, and 1.5 kbp fragments (indicated by arrows in Figure 3A) correspond to the pseudogene that we have characterized (see Figure 1). The remainder of the fragments correspond to the wild-type gene (M. L. Koschinsky, unpublished results). The hybridization pattern obtained by using the λhCP-1 cDNA probe (which corresponds to the 5′ end of the ceruloplasmin transcript) is shown in Figure 3B and contains only bands that have been assigned to the wild-type gene (unpublished results). Additionally, λhCP-1 hybridizes to only a single genomic *Pst*I fragment (data not shown), further suggesting the absence of pseudogene sequences corresponding to the 5′ end of the wild-type ceruloplasmin gene.

Human Ceruloplasmin Transcript Analysis. The presence of the 213 bp deletion in the pseudogene DNA sequence compared to the cDNA sequence was utilized to detect the presence of pseudogene-specific transcripts. A single-stranded DNA fragment (286 nucleotides in size) derived from the human ceruloplasmin cDNA sequence was used as a hybridization probe for nuclease S1 analysis. This probe corresponds to a region spanning nucleotide residues 1927–2213 of the cDNA sequence, thereby containing 150 bp within the deleted region identified in the pseudogene sequence (see Figure 4b). Protection of human liver poly(A+) RNA with this probe

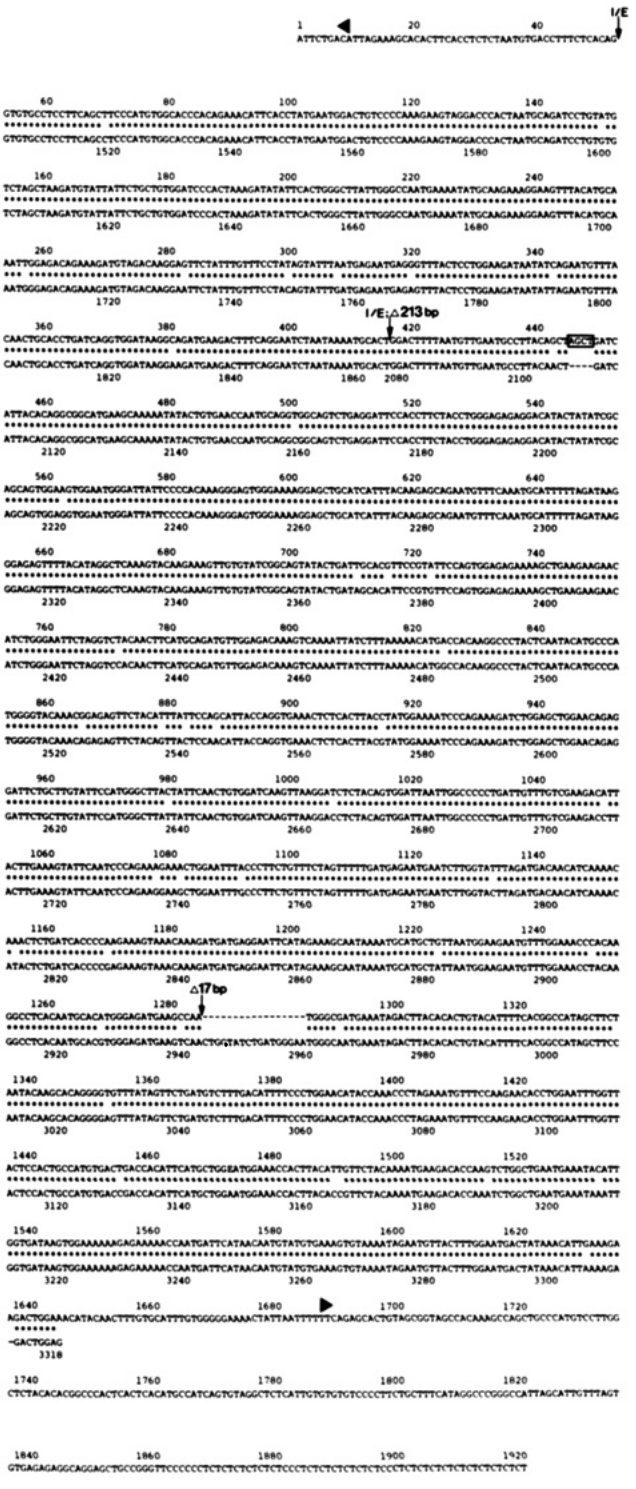

FIGURE 2: Nucleotide sequence of the human ceruloplasmin pseudogene and comparison with the corresponding region of the ceruloplasmin cDNA sequence. The pseudogene sequence was determined by analysis of the overlapping clones shown in Figure 1. I/E denotes positions of intron–exon boundaries in the wild-type gene (M. L. Koschinsky, unpublished results). The position of the AGCT insertion that results in a frameshift mutation is enclosed in a box. The sizes and positions of deletions (Δ) relative to the ceruloplasmin cDNA sequence are also shown; the region of the cDNA corresponding to the 213 bp deletion has been omitted. The places (5′ and 3′) where the pseudogene sequence diverges from the wild-type sequence are indicated by arrowheads (see text for details). The cDNA sequence represents nucleotides 1502–1864 and 2078–3318 of phCP-1 (Koschinsky et al., 1986). Asterisks indicate identical nucleotides in corresponding positions in the two sequences. Dashes were introduced at points of insertion or deletion in order to maximize homology.

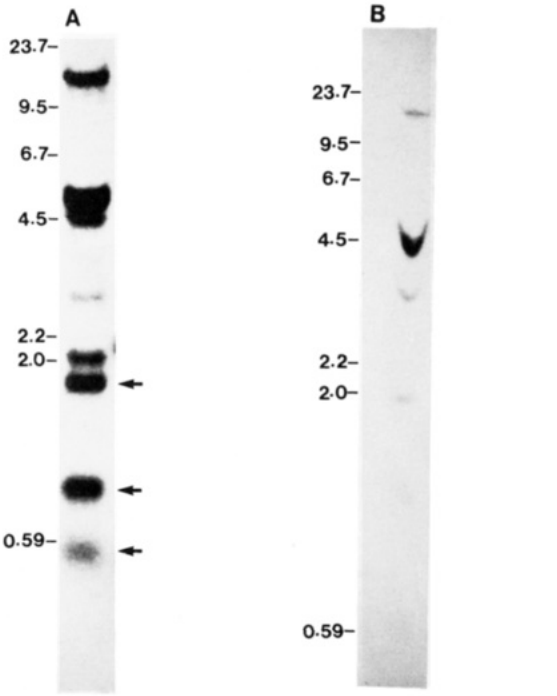

FIGURE 3: Genomic Southern analysis of the human ceruloplasmin gene and related sequences. *Eco*RI-digested genomic DNA was hybridized with either the phCP-1 (panel A) or the λhCP-1 (panel B) ceruloplasmin cDNA clones, under conditions of high stringency. Arrows indicate bands that have been assigned to the human ceruloplasmin pseudogene (see text for details). The remaining bands localized the wild-type gene (M. L. Koschinsky, unpublished results). Positions of ^{32}P-labeled *Hin*dIII fragments of phage λ DNA used as size markers are shown. Fragment sizes are given in kilobase pairs.

resulted in a single protected band of 286 bp, corresponding to the wild-type transcript (Figure 4a). In addition, a band corresponding to the full-length probe (FLP) was observed (Figure 4a) which contains M13 sequences in addition to ceruloplasmin sequences. If the pseudogene was transcribed in liver, and assuming that nuclease S1 cleaves all single-base mismatches between the probe and the putative transcript, protected fragments of 47 and 55 bp would be expected (corresponding to nucleotides 453–499 and 501–555, respectively; see Figure 2). However, no protected fragment of a smaller size, which would correspond to an RNA species containing the 212 bp deletion, was observed at either of the RNA concentrations used (Figure 4a).

Chromosomal Localization. The chromosomal location of the human ceruloplasmin pseudogene was determined by somatic cell hybrid analysis. For this purpose, an isolated 1.1 kbp *Eco*RI restriction fragment (probe A, Figure 1), located 3′ to the pseudogene, was used as a hybridization probe. Southern blot analysis of *Eco*RI-digested human–hamster somatic cell hybrid DNA (Figure 5) showed that all 22 hybrid cell lines tested (lanes 1–22) were concordant for the presence or absence of the 1.1 kbp band and human chromosome 8 (Table I). Control lanes containing either *Eco*RI-digested human placental DNA (lane HP, Figure 5) or hamster DNA (lane H, Figure 5) were included in the Southern blot analysis. No cross-hybridization was detected in the lane containing hamster DNA, while the expected 1.1 kbp *Eco*RI band corresponding to probe A was observed in the human placental DNA.

DISCUSSION

Using previously characterized cDNA clones for human ceruloplasmin as hybridization probes, we have isolated and

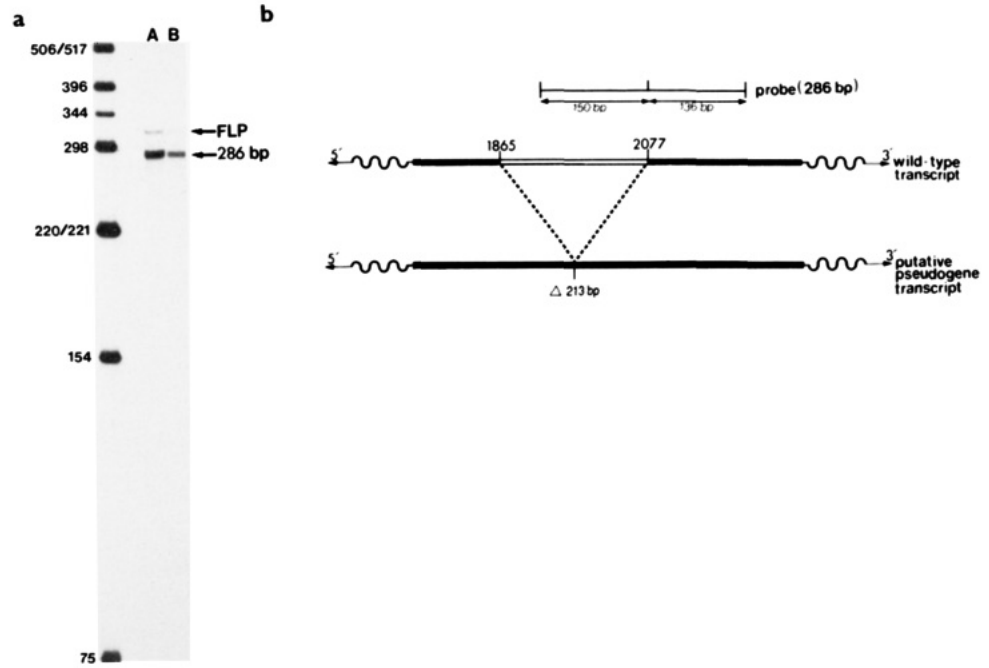

FIGURE 4: Nuclease S1 transcript analysis for the ceruloplasmin pseudogene. Panel b shows the location of the 286 bp probe used for S1 nuclease protection analysis relative to the 213 bp deletion observed in the pseudogene sequence. Part of this probe (150 bp) is within the deleted region (corresponding to nucleotides 1865–2077 of the cDNA), while 136 bp of the probe are 3′ to the deletion. Following hybridization to 1 μg (lane A) or 0.35 μg (lane B) of poly(A+) RNA and nuclease S1 digestion, S1-resistant products were separated on a denaturing polyacrylamide gel and visualized by autoradiography (panel a). A band corresponding to the size of the undigested full-length probe (FLP) was observed, as well as a protected band of 286 bp corresponding to the wild-type transcript. A *Hin*fI digest of pBR322 was used for markers. The sizes of resulting fragments are given in base pairs.

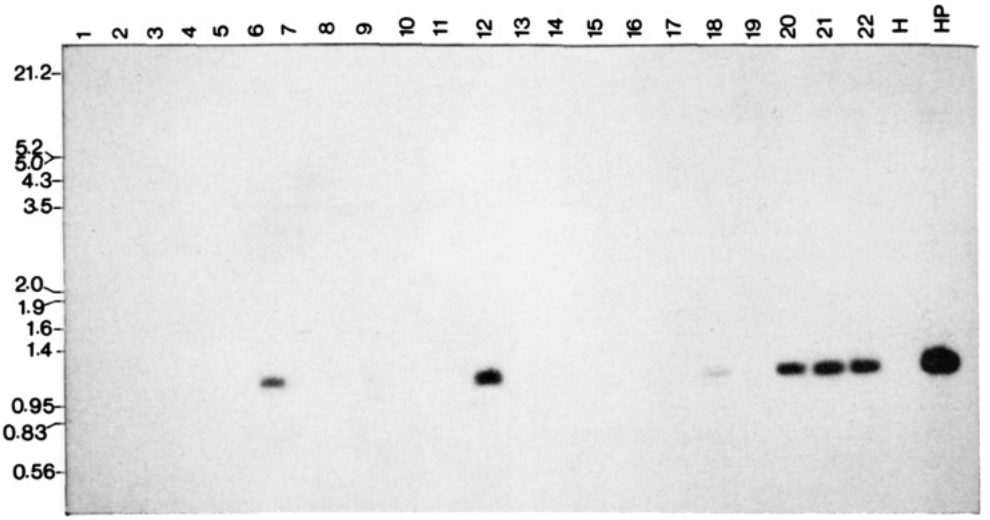

FIGURE 5: Chromosome mapping of the human ceruloplasmin pseudogene using somatic cell hybrid analysis. A 1.1 kbp pseudogene-specific probe (probe A; see Figure 1) was hybridized to *Eco*RI-digested DNA from human–hamster hybrid cell lines (lanes 1–22). Numbering of the cell lines corresponds to that shown in Table I. *Eco*RI-digested hamster DNA (lane H) and human placental DNA (lane HP) were included as controls. The positions of ^{32}P-labeled *Hin*dIII/*Eco*RI fragments of λ phage DNA used as size markers are shown. Fragment sizes are given in kilobase pairs.

characterized four overlapping recombinant phage clones that together encode approximately 21 kbp of contiguous human genomic DNA. Within this region, we have identified a pseudogene for human ceruloplasmin corresponding to nucleotides 1502–3198 of the ceruloplasmin cDNA coding sequence (Koschinsky et al., 1986). Additionally, the pseudogene extends through the 123 bp of 3′ untranslated sequence that is present in the phCP-1 clone previously described (Koschinsky et al., 1986) and continues for a further 40 bp, the sequence of which corresponds to the 3′ untranslated region of a ceruloplasmin cDNA clone described by Yang et al. (1986). On this basis, it appears that the pseudogene has been derived from an mRNA species corresponding to the cDNA clone described by Yang et al. (1986), as opposed to the shorter mRNA species represented by phCP-1. The derivation of the ceruloplasmin pseudogene from this particular mRNA species is interesting, since it has been shown that for different rat cytochrome *c* mRNA species occurring at the same intracellular concentration, multiple pseudogenes arose predominantly from one mRNA (Scarpulla, 1984). It has been speculated that this may be due to secondary structure in the 3′ end of the mRNA which facilitates binding of enzymes involved in reverse transcription, or subsequent integration into the genome (Scarpulla, 1984).

As is typical of processed pseudogenes, the human ceruloplasmin pseudogene is characterized by a lack of introns

Table I: Segregation of the Human Ceruloplasmin Pseudogene with Human Chromosomes in Human–Hamster Somatic Hybrids

	cell line	response to probe A[b]	identifiable, intact human chromsomes[a]																							
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
1	41.06	−	−	−	+	−	+	−	−	−	−	+	−	−	+	+	+	+	+	+	−	+	−	−	+	−
2	45.01	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
3	45.43	−	+	−	−	−	−	−	−	−	+	+	+	−	−	−	−	+	+	+	+	+	+	−	+	−
4	76.14	−	−	−	+	−	+	+	−	−	−	+	+	−	+	+	−	+	+	+	−	+	+	−	+	−
5	76.31	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+
6	76.33	+	−	−	+	+	+	−	−	+	−	+	−	−	+	−	+	+	−	+	+	−	+	−	+	−
7	79.05b	−	+	−	+	+	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	+	−	+	−
8	80.05d	−	−	−	−	−	−	+	+	−	−	−	−	+	−	−	−	+	+	−	−	+	−	−	+	+
9	80.14c	−	−	−	−	−	+	+	+	−	−	+	−	−	−	−	−	+	−	−	−	+	+	−	+	−
10	80.17a	−	−	−	−	−	+	+	−	−	+	−	+	−	−	−	+	+	−	+	+	+	−	+	−	−
11	82.82a	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−
12	85.16a	−	+	−	−	+	+	−	−	+	−	+	+	−	+	−	+	+	+	+	−	+	−	−	−	−
13	89.27a	−	−	−	−	−	+	−	−	−	+	−	−	+	−	−	−	−	+	+	−	−	+	+	−	−
14	100.02b	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+	−	+	−	−	−	−
15	102.05b	−	+	−	−	−	+	+	−	−	−	−	−	−	+	−	−	+	+	−	−	+	+	+	−	−
16	103.04	−	−	−	−	+	+	−	−	−	−	−	−	+	−	−	+	+	−	+	−	+	−	−	−	−
17	111.02a	−	+	+	−	−	+	+	−	−	−	−	−	+	+	+	+	+	+	−	+	+	+	+	−	−
18	112.10a	+	+	−	−	+	−	−	−	+	−	−	−	+	+	−	+	+	−	−	+	−	−	+	+	−
19	120.33	−	−	−	−	−	+	+	−	−	−	−	+	+	+	−	−	−	+	−	+	−	+	−	+	−
20	120.35	+	−	−	+	−	+	+	−	+	−	−	−	+	+	+	−	−	+	+	−	−	−	+	+	−
21	133.05	+	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−
22	134.02a	+	+	−	−	+	+	−	+	+	−	+	+	−	+	−	+	+	+	+	+	+	+	−	+	−
	% discordancy		27	18	32	23	45	45	32	0	41	27	27	41	27	45	36	64	50	50	32	55	55	45	41	41

[a] Presence (+) or absence (−) of human chromosomes as determined by cytogenetic analysis and confirmed by isozyme analysis (Donald et al., 1983; Riddell et al., 1985). [b] Presence (+) or absence (−) of human *Eco*RI-digested sequences homologous to the human ceruloplasmin pseudogene probe (probe A).

compared to the wild-type gene. Unlike numerous other examples of processed pseudogenes in which the intervening sequences are precisely removed [e.g., see Vanin et al. (1980)], the ceruloplasmin pseudogene contains a 213 bp deletion (corresponding to nucleotides 1865–2077 of the cDNA sequence) that occurs at the sites of intron–exon junctions in the wild-type gene (M. L. Koschinsky, unpublished results). There is also a small 17 bp deletion in the pseudogene sequence, beginning at nucleotide 2943 of the cDNA sequence. It has not been determined if this deletion also corresponds to the location of an intron–exon boundary in the wild-type gene. It is unclear whether the 213 bp deletion observed in the pseudogene occurred at the time of intron removal or whether the deletion is the result of a subsequent mutation event.

The 5′ boundary of the pseudogene is characterized by the presence of a short sequence that is homologous to an intron in the wild-type ceruloplasmin gene. This is expected since the pseudogene diverges from the wild-type gene prior to the 5′ end of this intron. Therefore, the appropriate 5′ splice donor site (Cech, 1983) required for intron removal is absent.

Although the pseudogene that we have identified appears to have been derived from a processed RNA species, there is no poly(A) tract present in the sequence. While the majority of processed pseudogenes have a poly(A) tail, several exceptions have been reported (Vanin et al., 1980; Notake et al., 1983). We postulate that the absence of a poly(A) tract in the human ceruloplasmin pseudogene may be the result of the mechanism of its formation, possibly involving base pairing between the poly(A) tail of the mRNA and U-rich regions in the 3′ untranslated region. In this case, the six T residues observed following the 3′ untranslated segment corresponding to that described by Yang et al. (1986) may represent the site of mRNA self-priming. This explanation may also account for the divergence that occurs between the pseudogene sequence and the remainder of the 3′ untranslated region characterized by Yang et al. (1986).

The presence of a repeated CT dinucleotide segment at the 3′ end of the ceruloplasmin pseudogene sequence corresponding to the coding strand is interesting. A 116 bp segment, composed mainly of repeated GA dinucleotides corresponding to the coding strand, has been reported at the 3′ end of the mouse corticotropin–β-lipotropin precursor pseudogene (Notake, 1983). In the latter case, this repeated segment occurs immediately following the point at which the pseudogene diverges from the wild-type gene. In the human ceruloplasmin pseudogene, the repeated TC region occurs 172 bp 3′ to the point at which the pseudogene diverges from the wild-type gene. The rat metallothionein pseudogene 1ψb, which has been characterized by Andersen et al. (1986), contains a 42 bp poly(CA) tract, located ~300 bp 3′ to the site of polyadenylylation. While stretches of repeating CA residues have been found in eukaryotic DNA at the site of recombination events such as gene conversion (Shen et al., 1981) and are thought to induce Z-DNA conformational changes (Nordheim & Rich, 1983), the function of significantly long purine or pyrimidine stretches remains unclear. It is possible that such sequences are involved in the process of pseudogene integration into the genome, since little is known about the mechanism mediating this event.

The pseudogene is highly homologous (~97% nucleotide sequence identity) to the wild-type ceruloplasmin sequence, suggesting that it has been formed relatively recently in evolutionary time. This is characteristic of processed pseudogenes that have been analyzed to date (Li et al., 1981; Freytag et al., 1984), all of which have arisen following mammalian

radiation (~100 million years ago) (Vanin, 1985). Genomic Southern blot analysis indicates that there are no pseudogene sequences corresponding to the 5′ end of the wild-type gene. This suggests that the pseudogene may have arisen from an aberrant transcript, as a result of initiation within the gene. Such a model has been proposed for the mouse corticotropin–β-lipotropin pseudogene (Notake, 1983). This pseudogene is similar to the human ceruloplasmin pseudogene in that it is only a partial copy of the functional gene, encoding the carboxy-terminal 143 amino acid residues and the 3′ untranslated region (Notake, 1983).

All processed pseudogenes studied to date are located on different chromosomes than their functional counterparts (Vanin, 1985). Using previously characterized human–hamster hybrid cell lines (Donald et al., 1983), we have mapped the ceruloplasmin pseudogene to human chromosome 8. This differs from the assignment of the functional ceruloplasmin gene to chromosome 3 (Yang et al., 1986; Royle et al., 1987). It has been reported previously by Yang et al. (1986) that a 0.8 kbp *Eco*RI band that can be identified on genomic Southern blots probed with the ceruloplasmin cDNA segregates with human chromosome 11. However, our data suggest that this 0.8 kbp *Eco*RI fragment is part of the human ceruloplasmin pseudogene that we have characterized and mapped to chromosome 8. The reason for this discrepancy is unclear at the present time, since genomic Southern analysis indicates that there is only one pseudogene for human ceruloplasmin. However, since the mapping analysis of Yang et al. (1986) was performed with the ceruloplasmin cDNA to probe human–mouse hybrid cell lines, the 0.8 kbp band may represent a cross-reacting species in the mouse genome. This is in agreement with previous difficulties that we have encountered in chromosome mapping when using cDNA fragments as probes (results not shown).

ACKNOWLEDGMENTS

We thank Dr. T. Maniatis and V. A. Geddes for providing genomic phage libraries and Roland Russnak for technical advice.

Registry No. Ceruloplasmin, 9031-37-2.

REFERENCES

Andersen, R. D., Birren, B. W., Taplitz, S. J., & Herschman, H. R. (1986) *Mol. Cell. Biol. 6*, 302–314.

Battey, J., Max, E. E., McBride, O. W., Swan, D., & Leder, P. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5956–5960.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Blin, N., & Stafford, D. W. (1976) *Nucleic Acids Res. 3*, 2303–2308.

Cech, T. R. (1983) *Cell (Cambridge, Mass.) 34*, 713–716.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Cleary, M. L., Schon, E. A., & Lingrel, J. B. (1981) *Cell (Cambridge, Mass.) 26*, 181–190.

Czosnek, H., Bienz, B., Givol, D., Zakut-Houri, R., Pravtcheva, D. D., Ruddle, F. H., & Oren, M. (1984) *Mol. Cell. Biol. 4*, 1638–1640.

Deininger, P. L. (1983) *Anal. Biochem. 129*, 216–223.

Donald, L. J., Wang, H. S., Kamali, V., Liebert, L., Rudner-Marion, M., Cameron, E. S., Riddell, D. C., Vust, A., Wrogemann, K., & Hamerton, J. L. (1983) *Cytogenet. Cell Genet. 37*, 453A.

Eiberg, H., & Mohr, J. (1979) *Cytogenet. Cell Genet. 25*, 149–150.

Eiberg, H., Mohr, J., & Nielsen, L. S. (1983) *Cytogenet. Cell Genet. 37*, 461.

Feinberg, A., & Vogelstein, B. (1982) *Anal. Biochem. 132*, 6.

Freytag, S. O., Bock, H.-G. O., Beudet, A. L., & O'Brien, W. E. (1984) *J. Biol. Chem. 259*, 3160–3166.

Frieden, E. (1981) *Met. Ions Biol. Syst. 13*, 117–142.

Geddes, V. A. (1987) M.Sc. Thesis, University of British Columbia.

Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D., & Leder, P. (1982) *Nature (London) 296*, 321–325.

Jacq, C., Miller, J. R., & Brownlee, G. G. (1977) *Cell (Cambridge, Mass.) 13*, 109–120.

Kan, Y. W., & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5631–5635.

Kay, R. J., Boissy, R. J., Russnak, R. H., & Candido, E. P. M. (1986) *Mol. Cell. Biol. 6*, 3134–3143.

Kingston, I. B., Kingston, B. L., & Putnam, F. W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5377–5381.

Koschinsky, M. L., Funk, W. F., van Oost, B. A., & MacGillivray, R. T. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5086–5090.

Lacy, E., & Maniatis, T. (1980) *Cell (Cambridge, Mass.) 21*, 545–553.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell (Cambridge, Mass.) 15*, 1157–1174.

Li, W.-H., Gojobori, T., & Nei, M. (1981) *Nature (London) 292*, 237–239.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Nordheim, A., & Rich, A. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1821–1825.

Notake, M., Tobimatsu, T., Wantanabe, Y., Takahashi, H., Mishina, M., & Numa, S. (1983) *FEBS Lett. 156*, 67–71.

Riddell, D. C., Mallonee, R., Phillips, J. A., Parks, J. S., Sexton, L. A., & Hamerton, J. L. (1985) *Somatic Cell Mol. Genet. 11*, 189–196.

Riddell, D. C., Wang, H. S., Beckett, J., Holden, J. J. A., Mulligan, L., Phillips, A., Simpson, N., Wrogemann, K., Hamerton, J. L., & White, B. N. (1986) *Cytogenet. Cell Genet. 42*, 123–128.

Robson, E. B., Sutherland, I., & Harris, H. (1966) *Ann. Hum. Genet. 29*, 325–336.

Royle, N. J., Irwin, D. M., Koschinsky, M. L., MacGillivray, R. T. A., & Hamerton, J. L. (1987) *Somatic Cell. Mol. Genet. 13*, 285–292.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Scarpulla, R. C. (1984) *Mol. Cell. Biol. 4*, 2279–2288.

Shen, S. H., Slighton, J. L., & Smithies, O. (1981) *Cell (Cambridge, Mass.) 26*, 191–203.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Staden, R. (1982) *Nucleic Acids Res. 10*, 4731–4751.

Takahashi, N., Ortel, T., & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 390–394.

Ueda, S., Nakai, S., Nishida, Y., Hisajima, H., & Honjo, T. (1982) *EMBO J. 1*, 1539–1544.

Vanin, E. F. (1985) *Annu. Rev. Genet. 19*, 253–272.

Vanin, E. F., Goldberg, G. I., Tucker, P. W., & Smithies, O. (1980) *Nature (London) 286*, 222–226.
Woo, S. L. C. (1980) *Methods Enzymol. 68*, 389–395.
Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., van Bragt, P. H., Baldwin, W. D., & Bowman, B. H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2752–2756.
Yang, F., Naylor, S. L., Lum, J. B., Cutshaw, S., McCoombs, J. L., Naberhaus, K. H., McGill, J. R., Adrian, G. S., Moore, C. M., Barnett, D. R., & Bowman, B. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3257–3261.

Synthesis and Biological Activity of Amino Terminus Extended Analogues of the α Mating Factor of *Saccharomyces cerevisiae*†

Michael A. Tallon,‡ P. Shenbagamurthi,‡ Stevan Marcus,§ Jeffrey M. Becker,§ and Fred Naider*,‡

Department of Chemistry, The College of Staten Island, City University of New York, Staten Island, New York 10301, and Department of Microbiology and Program in Cellular, Molecular, and Developmental Biology, University of Tennessee, Knoxville, Tennessee 37996

Received March 31, 1987; Revised Manuscript Received June 30, 1987

ABSTRACT: The synthesis and biological activity are reported for extended analogues of the secreted tridecapeptide α-factor (Trp-His-Trp-Leu-Gln-Leu-Lys-Pro-Gly-Gln-Pro-Met-Tyr) from *Saccharomyces cerevisiae*. Peptides with Ala, Glu-Ala, Ala-Glu-Ala, or Glu-Ala-Glu-Ala attached to the amino terminus of α-factor were synthesized by the solid-phase method on a (phenylacetamido)methyl (PAM) resin, using a combination of dicyclohexylcarbodiimide- and 1-hydroxybenzotriazole-accelerated active ester coupling procedures. Free peptides were obtained by hydrogen fluoride (HF) cleavage in the presence of appropriate scavengers. Normal high HF cleavage and "low–high" HF cleavage were equally effective in liberating the desired product from the PAM resin. Yields of pure peptide ranged from 9% to 17%. All of the extended α-factors, which represent sequences of pro-α-factor coded for in the *MFα1* structural gene, caused morphological aberrations (shmoo assay) in strain X2180-1A (*MATa*) the same as those caused by the tridecapeptide. The 14-peptide was equally active compared to the native α-factor whereas the 17-peptide was 5–10-fold less active. The analogues also arrested to various degrees (halo assay) the growth of *S. cerevisiae* RC629 (*MATa sst1*) and *S. cerevisiae* RC631 (*MATa sst2*), two supersensitive mutants, and were converted to pheromones of equal activity by treatment with V8 protease. A temperature-sensitive receptor mutant responded to all the peptides at the permissive but not the restrictive temperature. An α-factor antagonist, des-Trp^1,Ala^3-α-factor, inhibited activity of all extended peptides. These results confirm that all the extended peptides interact with the same receptor and that this receptor can accommodate additional residues at the amino terminus of the α-factor.

Cells of the α mating type of *Saccharomyces cerevisiae* secrete a low molecular weight peptide, termed α-factor, which is required for sexual conjugation between α haploids (*MATα*) and *a* haploids (*MATa*) of this yeast (Thorner, 1981). The α-factor specifically inhibits DNA replication and initiates an aberrant elongation of *MATa* cells, called "shmooing", which forms the basis for a biological assay of the mating factor. The sequence of two peptides responsible for α-factor activity, Trp-His-Trp-Leu-Gln-Leu-Lys-Pro-Gly-Gln-Pro-Met-Tyr and His-Trp-Leu-Gln-Leu-Lys-Pro-Gly-Gln-Pro-Met-Tyr, was determined by Stotzler et al. (1976). The tridecapeptide is the predominant species responsible for mating factor activity. Direct evidence has been obtained that the α-factor encoded by the *MFα1* gene is biosynthesized as a larger precursor polypeptide, prepro-α-factor (Emter et al., 1983; Julius et al., 1983, 1984a,b). This precursor contains four identical tandem repeats of the mature pheromone sequence, each separated from the other by spacer regions of six to eight residues of nearly identical sequence: Lys-Arg-Glu-Ala-Glu-Ala and Lys-Arg-Glu-Ala-Glu(or Asp)-Ala-Glu-Ala.

Previous genetic and biochemical studies of mutations affecting processing and secretion of prepro-α-factor (Achstetter & Wolf, 1985; Julius et al., 1983, 1984a), as well as analysis of the organization of the α-factor structural gene *MFα1* (Kurjan, 1985; Kurjan & Herskowitz, 1982; Singh et al., 1983), provide strong evidence that the first proteolytic processing event in α-factor maturation is cleavage at the carboxyl side of the Lys-Arg pair of the spacer regions. Such cleavage would result in four propheromone fragments containing four to six residues, Glu-Ala-Glu-Ala or Glu-Ala-Glu(or Asp)-Ala-Glu-Ala, attached to the amino terminus of the mature tridecapeptide α-factor. Three of these fragments would also have the Lys-Arg pair attached to their carboxyl ends. The final processing events to yield tridecapeptide would be the result of the combined action of carboxypeptidase ysc (which cleaves Lys-Arg) and dipeptidyl aminopeptidase yscIV (which cleaves X-Ala sequences) (Wolf, 1986).

Nonmating *ste13* mutants of *MATα* cells bearing defects in the dipeptidyl aminopeptidase do not produce normal α-factor (Julius et al., 1983). Rather, α-factor reported to contain Glu-Ala-Glu-Ala or Asp-Ala-Glu-Ala on the amino terminus of a tridecapeptide was isolated from cultures of *ste13*

† This study was supported by Grants GM 22087 and GM 22086 from the National Institute of General Medical Sciences, National Institutes of Health, and a grant from the PSC/HBE grants program of the City University of New York.
‡ City University of New York.
§ University of Tennessee.

0006-2960/87/0426-7767$01.50/0