Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5086–5090, July 1986
Biochemistry

Complete cDNA sequence of human preceruloplasmin

(copper-binding protein/oligonucleotides/amino acid sequence homology/gene evolution)

MARLYS L. KOSCHINSKY, WALTER D. FUNK, BERNARD A. VAN OOST, AND ROSS T. A. MACGILLIVRAY

Department of Biochemistry, University of British Columbia, Vancouver, BC, Canada V6T 1W5

Communicated by Harry B. Gray, March 31, 1986

ABSTRACT **A cDNA for human ceruloplasmin (EC 1.16.3.1) was identified in a human liver cDNA library by screening with two mixtures of synthetic oligodeoxyribonucleotides that were complementary to two regions of ceruloplasmin mRNA as predicted from the amino acid sequence of plasma ceruloplasmin. The resulting clone (phCP1) contained DNA coding for amino acid residues 202–1046 of the protein, followed by a stop codon, a 3′ untranslated region of 123 base pairs, and a poly(A) tail. To isolate cDNAs encoding the 5′ end of ceruloplasmin mRNA, a cDNA library was constructed in λgt10. The cDNA for this library was synthesized by reverse transcription of human liver poly(A)$^+$ RNA, using random oligonucleotides as primers. When this cDNA library was screened by using a 5′ fragment of phCP1 as a hybridization probe, several positive clones were identified. One of these clones (λhCP1) contained DNA coding for a probable signal peptide of 19 amino acid residues followed by DNA coding for residues 1–380 of plasma ceruloplasmin. Blot hybridization analysis showed that ceruloplasmin mRNA from human liver and the human hepatoma cell line HepG2 is 3700 nucleotides in size. Liver contained an additional mRNA species that is like ceruloplasmin mRNA and is 4500 nucleotides in size. Comparison of the complete nucleotide sequences of human ceruloplasmin cDNA and human clotting factor VIII cDNA showed regions of sequence homology, suggesting that these two proteins have evolved from a common ancestor.**

Ceruloplasmin [ferroxidase; iron (II):oxygen oxidoreductase, EC 1.16.3.1] is a blue glycoprotein from the α_2-globulin fraction of vertebrate plasma (1). This protein is synthesized in the liver as a single polypeptide chain of M_r 132,000 (2, 3) and is the principal copper transport protein in plasma, binding 90–95% of the blood copper in vertebrates (1). Each molecule of ceruloplasmin possesses 6 (or 8) copper atoms bound tightly at spectroscopically defined sites (4, 5). Up to 10 additional copper atoms are bound less tightly to the molecule and may be involved in a copper transport function for ceruloplasmin (6). The 6 (or 8) spectroscopically defined copper-binding sites are differentiated into three types with the following stoichiometry: two type I sites, one type II site, and 2 (or 3) type III sites (4, 5). In addition to its primary role in copper transport, at least three other functions have been ascribed to ceruloplasmin, including ferroxidase activity (7, 8), amine oxidase activity (e.g., ref. 9), and superoxide dismutase activity (10). Ceruloplasmin is also an acute phase reactant (11), and its plasma level is increased 2- to 3-fold in response to inflammation (4). It has been suggested that the observed heterogeneity of function may be related to the various catalytic activities provided by the three types of copper-binding sites (1). However, the diverse functional nature of ceruloplasmin remains poorly understood, since the multiple enzymic functions of the protein have not yet been localized to specific areas of the polypeptide chain. In the hereditary disorder Wilson disease (hepatolenticular degeneration), serum ceruloplasmin levels are characteristically decreased (12). This deficiency has been ascribed to a genetic defect that leads to a disruption of normal copper metabolism and subsequent copper deposition in tissues (13).

The complete amino acid sequence of human ceruloplasmin demonstrates an internal threefold homology (3). Each homology unit consists of approximately 350 amino acid residues and shares 40% sequence identity with the other units. The pattern of proteolytic cleavage of ceruloplasmin suggests that each of the three repeat units can be further subdivided into two or three domains (14). These domains may correspond to the different biological activities observed within the ceruloplasmin molecule such that the multiple enzymic functions of the protein are localized to specific areas of the polypeptide chain.

The internal triplication shown in the structure of the human ceruloplasmin molecule poses some interesting evolutionary questions. Ceruloplasmin contains amino acid residues in positions homologous to known type I copper-binding sites in azurin and plastocyanin (15, 16). A different sequence in ceruloplasmin is homologous to a copper-binding site in bovine superoxide dismutase (17) and also to sequences in cytochrome oxidase (18). In addition to this observed relationship to both copper oxidases and multicopper oxidases, it has been demonstrated recently that regions corresponding to the three repeat units in human ceruloplasmin are present in bovine factor V (19) and human factor VIII (20, 21), two accessory proteins in the blood clotting cascade (22). Although factor V probably contains a type II copper-binding site (23), the significance of the extensive sequence homology between these clotting factors and ceruloplasmin is unclear at present.

In this paper, we report the isolation and characterization of two human ceruloplasmin cDNA clones that together encode a leader peptide of 19 amino acid residues, the complete amino acid sequence of plasma ceruloplasmin, a 3′ untranslated region, and a poly(A) tail.

MATERIALS AND METHODS

Materials. All restriction and DNA-modifying enzymes were purchased from Bethesda Research Laboratories, Pharmacia-PL Biochemicals, New England Biolabs, or Boehringer Mannheim. Avian myeloblastosis virus reverse transcriptase was purchased from Life Sciences (St. Petersburg, FL). *Eco*RI-digested λgt10 DNA (treated with calf intestine alkaline phosphatase) and Gigapack packaging extracts were obtained from Vector Cloning Systems. DNase I-digested rat thymus DNA was generously supplied by A. Wallis (Department of Medical Genetics, University of British Columbia).

Synthetic Oligonucleotide Mixtures Encoding Human Ceruloplasmin. Three pools of heptadecadeoxyribonucleotides were used as hybridization probes:

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s).

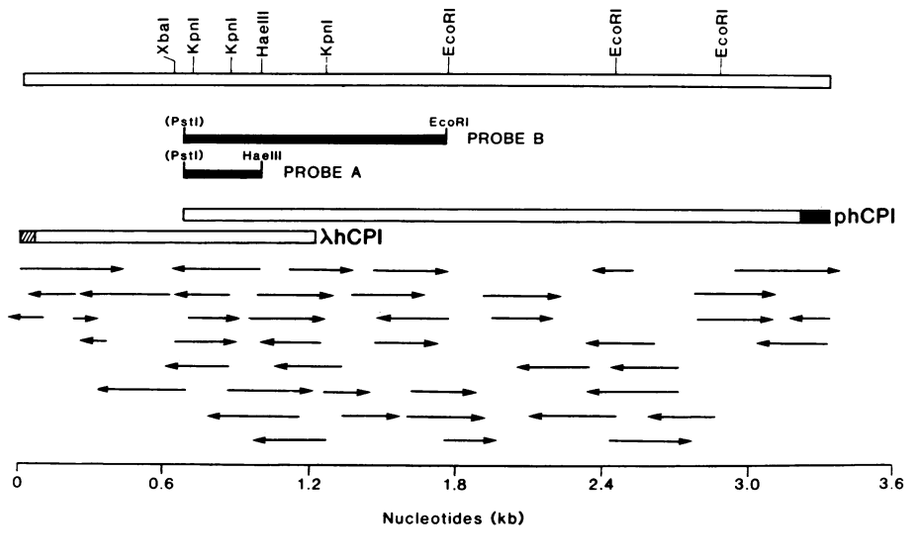

FIG. 1. Restriction map and sequencing strategy for human preceruloplasmin cDNA clones. The longer bars below the restriction map represent the clones phCP1 and λhCP1 and together include regions coding for the leader peptide (hatched bar), the plasma protein (open bar), and the 3′ untranslated sequence (solid bar). Arrows indicate the extent and direction of nucleotide sequence obtained from various M13 clones. Restriction fragment probes A and B, which were used in library screening (see text for details), are indicated directly below the restriction map (solid bars). The *Pst* I sites in parentheses result from the cloning procedure used in the construction of the cDNA library. kb, Kilobases.

Pool I: 5′ d(TARTARTGYTTYTCYTT) 3′
Pool II: 5′ d(ATNGCRTGCATYTTRTT) 3′
Pool III: 5′ d(CCCATNARRTACCARTT) 3′,

in which R represents both G and A, Y represents T and C, and N represents A, G, T, and C. The three oligonucleotide pools are complementary to the mRNA encoding amino acid residues 1–6, 937–942, and 962–967 of ceruloplasmin, respectively, as predicted from the amino acid sequence (3). The oligonucleotide pools were synthesized with an Applied Biosystems 380A DNA synthesizer, and the heptadecanucleotide fractions were purified by polyacrylamide gel electrophoresis in the presence of 8.3 M urea (24). The oligonucleotide mixtures were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (25), and the unincorporated ATP was removed by chromatography on Sephadex G-25. The excluded fraction was added directly to the hybridization mix.

Preparation of RNA. Samples of human liver were obtained from brain-dead organ donors. The liver samples were rinsed in cold saline, immediately frozen in liquid nitrogen, and stored at −70°C. RNA was isolated from the frozen liver and from HepG2 cells by the guanidine hydrochloride method (26). Poly(A)$^+$ RNA was isolated by chromatography on oligo(dT)-cellulose (ref. 27, pp. 197–198).

Human Liver cDNA Libraries. An adult liver cDNA library (28) was provided by S. H. Orkin (Children's Hospital Medical Center, Boston). This library contains cDNA inserts of ≥500 base pairs (bp) inserted into the *Pst* I site of pKT218 by homopolymeric dG·dC tailing. To isolate cDNAs encoding the 5′ end of ceruloplasmin mRNA, a randomly primed human liver cDNA library was constructed in λgt10 (29). Human liver poly(A)$^+$ RNA was used as a template for the synthesis of cDNA by reverse transcriptase. DNase I-digested rat thymus DNA (average length 20 nucleotides) was used as a primer (30). Second strand synthesis was performed as described by Gubler and Hoffman (31), using ribonuclease H, DNA polymerase I, and *Escherichia coli* DNA ligase. After S1 nuclease treatment, the double-stranded cDNA was methylated by using *Eco*RI methylase and *S*-adenosylmethionine, *Eco*RI linkers were ligated to the ends, and the linkers were digested with *Eco*RI. The cDNA was then chromatographed on a column (30 × 0.2 cm) of Bio-Gel A-50m (Bio-Rad) equilibrated with 0.01 M Tris·HCl, pH 7.5/0.3 M NaCl/0.001 M EDTA. Fractions forming the leading edge of the cDNA peak (corresponding to cDNA fragments >1000 bp in size) were pooled, and the DNA (50 ng) was ligated with *Eco*RI-digested, dephosphorylated λgt10 DNA (1 μg). Half of the resulting DNA was packaged into phage particles *in vitro* by using a Gigapack and plated on *E. coli* strain C600 Hfl$^+$. The library contained 400,000 independent recombinants and was screened without amplification.

Screening the cDNA Libraries. The cDNA libraries were screened with ^{32}P-labeled DNA fragments as hybridization probes. Initially, mixtures of synthetic oligonucleotides were used with the hybridization and washing conditions of Fung *et al.* (32). The libraries were subsequently screened with restriction fragments labeled by nick-translation (33) and phage M13 clones labeled by primer extension (34). Hybridization and washing conditions for the latter screens were as described by Maniatis *et al.* (ref. 27, pp. 326–328).

DNA Sequencing Analysis. DNA sequence analysis of phCP1 (see Fig. 1) was carried out essentially as described by Deininger (35). Plasmid DNA was randomly sheared by sonication. Fragments (300–500 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 M13mp8 and used to transform *E. coli* strain JM103 (36). Subclones containing ceruloplasmin cDNA inserts were identified by plaque hybridization (36), using the ^{32}P-labeled phCP1 *Pst* I insert as a probe. Restriction endonuclease fragments subcloned in appropriate M13 vectors were also used in sequence analysis of the cDNA clones. All DNA sequence analysis was performed by using the chain termination method (37).

Computer Analysis. DNA sequence data were analyzed by using the DBUTIL program of Staden (38).

Blot Hybridization Analysis. Samples of poly(A)$^+$ RNA were denatured with formamide and separated by electrophoresis in a 1% agarose gel according to Maniatis *et al.* (ref. 27, pp. 202–203). After transfer to nitrocellulose, the RNA was hybridized to phCP1 plasmid that had been labeled previously by nick-translation (33).

RESULTS AND DISCUSSION

Isolation of Human Ceruloplasmin cDNAs. Two hundred thousand recombinant clones from a human liver cDNA library (28) were screened at high colony density by using the pool II and pool III oligonucleotide mixtures as hybridization probes. One recombinant plasmid, designated phCP1, hybridized specifically to both oligonucleotide mixtures. Restriction endonuclease mapping showed that phCP1 contained a *Pst* I insert of 2700 bp. Subsequent DNA sequence analysis showed that this insert contained DNA coding for amino acid residues 202–1046 of plasma ceruloplasmin (3) in addition to a 3′ untranslated region and a poly(A) tract (see following section). To isolate a clone coding for the 5′ region

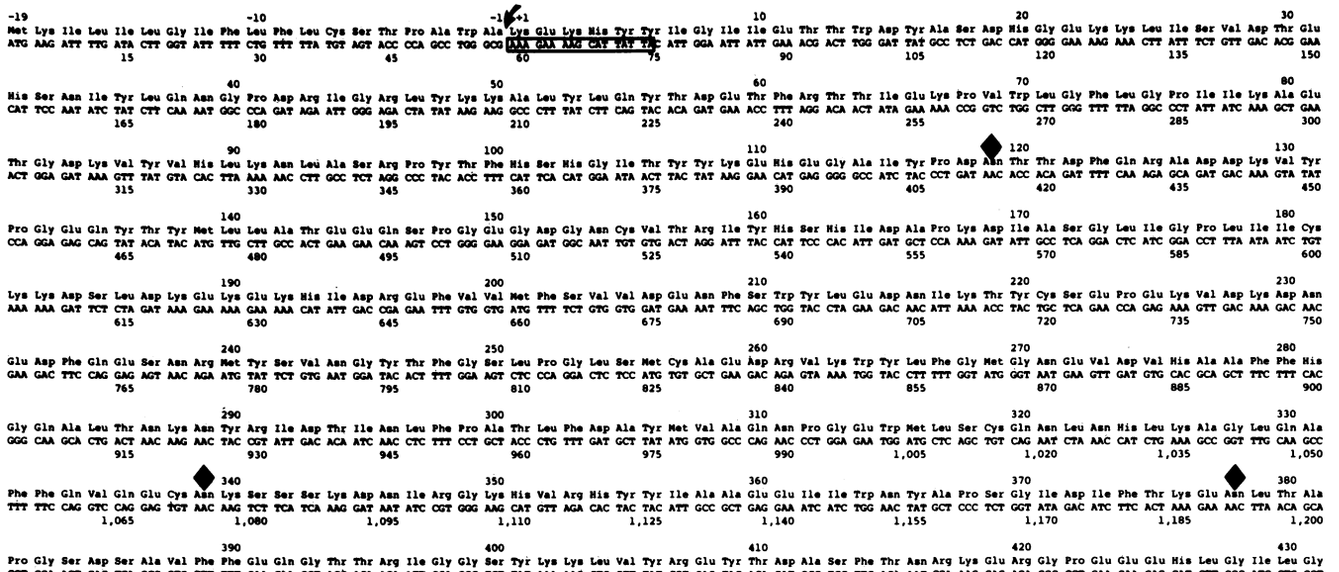

FIG. 2. Nucleotide sequence of human preceruloplasmin cDNA. The sequence was determined by analysis of the overlapping clones shown in Fig. 1. The predicted amino acid sequence of human preceruloplasmin is indicated above the DNA sequence. The putative signal peptidase cleavage site is shown by a solid arrow. Potential carbohydrate attachment sites (3) are represented by solid diamonds. Boxed sequences are complementary to oligonucleotide probes used to screen the cDNA libraries. The polyadenylylation signal ATTAAA is underlined.

of ceruloplasmin mRNA, the cDNA library was rescreened, using a 322-bp *Hae* III–*Pst* I fragment as a hybridization probe. This fragment was derived from the 5′ end of the phCP1 insert (probe A, Fig. 1). However, no clones were identified that extended further 5′ than phCP1.

To isolate a clone containing cDNA coding for the remainder of ceruloplasmin mRNA, a second cDNA library was constructed. To avoid the construction of a library that was enriched for cDNAs coding for the 3′ regions of mRNAs [as a consequence of priming cDNA synthesis from the poly(A) tail], cDNA was synthesized from human liver poly(A)$^+$ RNA with random deoxyribonucleotides (average size 20 nucleotides) used as primers. To select for longer cDNA clones, the double-stranded cDNA was sized by gel filtration prior to ligation into λgt10. Four hundred thousand clones from this library were screened by plaque hybridization using the 1071-bp *Pst* I–*Eco*RI fragment of phCP1 as a probe (probe B, Fig. 1). Of the 16 positive clones that were identified, 13 hybridized to the 322-bp *Hae* III–*Pst* I fragment derived from the 5′ end of the phCP1 insert (probe A, Fig. 1). Of these, only one was found to hybridize to the pool I oligonucleotide mixture, which corresponds to the amino-terminal six amino acids of plasma ceruloplasmin. Only this clone, designated λhCP1, was studied further. Restriction endonuclease mapping indicated that λhCP1 contained a 1200-bp insert and overlapped phCP1 as shown in Fig. 1.

DNA Sequence Analysis. The complete nucleotide sequences of the inserts of phCP1 and λhCP1 were determined by using the strategy shown in Fig. 1. The majority of the sequence was determined by analysis of randomly sheared fragments cloned in M13. The remainder of the sequence was determined by analysis of specific restriction endonuclease fragments cloned in M13. The complete nucleotide sequence of the two ceruloplasmin cDNAs and the predicted amino acid sequence of the protein are shown in Fig. 2. The position of each nucleotide was determined an average of 3.4 times, and 62% of the sequence was determined on both strands. In the region where they overlap, the nucleotide sequences of λhCP1 and phCP1 were identical.

Nucleotides 58–3195 of the cDNA sequence code for the plasma form of ceruloplasmin; the predicted amino acid sequence agrees completely with that determined by Takahashi *et al.* (3), who used protein chemistry techniques. Following the open reading frame is a TGA stop codon (encoded by nucleotides 3196–3198), a 3′ untranslated region of 123 bp (nucleotides 3199–3321), and a poly(A) tail. The 3′ untranslated region contains a putative polyadenylylation signal ATTAAA (39) that is located 14 nucleotides upstream of the poly(A) tail. This polyadenylylation signal is observed in 12% of such 3′-terminal sequences from vertebrates (40) and is a variant of the more commonly observed signal AATAAA. Nucleotides 1–57 code for an amino-terminal extension of 19 amino acid residues that is removed prior to the appearance of ceruloplasmin in plasma. The amino-terminal leader peptide contains a methionine residue at position −19, which may function as the initiator methionine. The leader peptide is rich in hydrophobic residues and thus resembles a signal peptide (41, 42). Such sequences function in the initiation of export of nascent polypeptide chains across the rough endoplasmic reticulum (43). The cDNA sequence predicts that an Ala-Lys bond (encoded by nucleotides 55–60, Fig. 2) is cleaved during the removal of the leader peptide. This is consistent with demonstrated signal peptidase cleavage specificity (44) and suggests that ceruloplasmin is synthesized in liver as a typical preprotein containing a signal sequence of at least 19 amino acids.

The base composition of ceruloplasmin mRNA is somewhat A+U rich (33% A, 26% U, 22% G, 19% C), reflective of the coding region, in which 60% of the codons end in either A or U. This observation is in contrast to the codon usage in other liver mRNAs such as those for prothrombin (45), factor X (46), and factor XII (47), in which approximately 90% of the codons end in G or C. One codon is not used in the coding region of ceruloplasmin mRNA (CGC for arginine), and others are used rarely (2 of 51 alanine residues are encoded by GCG).

Blot Hybridization Analysis. To estimate the size of ceruloplasmin mRNA, samples of poly(A)$^+$ RNA from human liver and total RNA from the human hepatoma cell line HepG2 (48) were separated by electrophoresis in an agarose gel and transferred to nitrocellulose. The RNA blot was then hybridized to the cDNA insert of phCP1. In both the liver and HepG2 samples, the cDNA hybridized to a mRNA species that was 3700 ± 200 nucleotides in size (Fig. 3). However, the cDNA hybridized to an additional mRNA species of 4500 ± 250 nucleotides in the liver RNA sample (Fig. 3, lane 1). The cDNA clones contain a total of 3321 bp plus a poly(A) tract, which is usually 180–200 nucleotides (49). Ceruloplasmin is synthesized and secreted from HepG2 cells (48), which suggests that the 3700-nucleotide RNA species represents a functional ceruloplasmin mRNA. The identity of the 4500-nucleotide RNA species is unclear, but it may represent a highly homologous mRNA or alternate processing of ceruloplasmin heterogeneous nuclear RNA. Because it is not detected in the hepatoma cell RNA, however, it is possible that the 4500-nucleotide RNA represents a homologous mRNA synthesized in nonhepatocyte cells in the liver.

Homology With Factor VIII. Extensive amino acid sequence homology has been reported between the three domains of ceruloplasmin (3) and three domains of blood coagulation factor VIII (20, 21). As expected, this homology extends to the nucleotide sequences when they are aligned according to Vehar *et al.* (20), as shown in Table 1. The three domains of ceruloplasmin exhibit 46–51% nucleotide sequence identity with each other, while the three homologous domains A1–A3 of factor VIII exhibit 40–44% identity with each other. In addition, the ceruloplasmin domains show 40–48% identity with the factor VIII domains, indicating that the sequences of each of the six domains have mutated to approximately the same extent. This similarity in homology may indicate similar functional constraints in the ceruloplasmin and factor VIII molecules.

Summary. We have isolated and characterized cDNA clones coding for human preceruloplasmin. These cDNAs

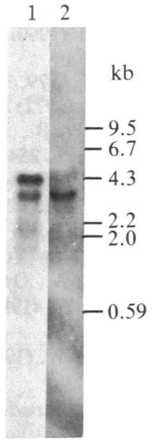

FIG. 3. Blot hybridization analysis of human ceruloplasmin mRNA. RNA was separated by electrophoresis in a denaturing agarose/formaldehyde gel and transferred to nitrocellulose. The filter was hybridized with ^{32}P-labeled phCP1. The filter was exposed to x-ray film for 18 hr at −70°C with Lightning Plus intensifying screens (Du Pont). Lane 1, 10 μg of human liver poly(A)$^+$ RNA; lane 2, 20 μg of total HepG2 cell RNA. The positions of *Hin*dIII fragments of λ phage DNA used as size markers are shown.

Table 1. Comparison of the nucleotide sequences of the homologous regions of ceruloplasmin and factor VIII

	% identity				
	CP-2	CP-3	VIII-A1	VIII-A2	VIII-A3
CP-1	51	48	43	48	42
CP-2		46	41	45	42
CP-3			40	43	46
VIII-A1				41	40
VIII-A2					44

The putative domains CP-1, CP-2, and CP-3 correspond to amino acid residues 1–350, 351–710, and 711–1046 of ceruloplasmin, respectively (see Fig. 2). The regions VIII-A1, VIII-A2, and VIII-A3 correspond to amino acid residues 1–339, 378–721, and 1691–2028 of factor VIII (20), respectively. The nucleotide sequences were aligned by inserting gaps to maximize the corresponding amino acid sequence homologies, as described by Vehar *et al.* (20). The comparisons are expressed as the percentage of identical nucleotides in corresponding positions in two sequences.

can now be used to isolate the human ceruloplasmin gene. Comparison of the organization of the ceruloplasmin and factor VIII genes will allow a detailed analysis of the evolution of these two plasma proteins. In addition, the cDNA clones can be used to investigate the association of ceruloplasmin with Wilson disease by using restriction fragment length polymorphisms.

We thank Dr. Stuart Orkin for sending us a human liver cDNA library; Gordon Vehar, Richard Lawn, Roland Russnak, and Cora-Jean Edgell for helpful suggestions; and Grant Mauk for critically reading the manuscript. This work was supported in part by grants from the Medical Research Council of Canada and the British Columbia Health Care Research Foundation. M.L.K. was supported by a Studentship from the Medical Research Council of Canada.

1. Frieden, E. (1981) in *Metal Ions in Biological Systems*, ed. Sigel, H. (Dekker, New York), Vol. 13, pp. 117–142.
2. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5377–5381.
3. Takahashi, N., Ortel, T. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 390–394.
4. Owen, C. A., Jr. (1982) *Biochemical Aspects of Copper: Copper Proteins, Ceruloplasmin and Copper Protein Binding* (Noyes, Park Ridge, NJ).
5. Ryden, L. & Bjork, I. (1976) *Biochemistry* **15,** 3411–3417.
6. McKee, D. J. & Frieden, E. (1971) *Biochemistry* **10,** 3880–3883.
7. Curzon, G. & O'Reilly, S. (1960) *Biochem. Biophys. Res. Commun.* **2,** 284–286.
8. Curzon, G. (1961) *Biochem. J.* **79,** 656–663.
9. Young, S. N. & Curzon, G. (1972) *Biochem. J.* **129,** 273–283.
10. Goldstein, I. M., Kaplan, H. B., Edelson, H. S. & Weissmann, G. (1979) *J. Biol. Chem.* **254,** 4040–4045.
11. Cooper, E. & Ward, M. (1979) *Invest. Cell Pathol.* **2,** 293–301.
12. Scheinberg, I. H. & Gitlin, D. (1952) *Science* **116,** 484–485.
13. Gitlin, D. & Gitlin, J. D. (1975) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, New York), Vol. 2, pp. 321–374.
14. Takahashi, N., Bauman, R. A., Ortel, T. L., Dwulet, F. E., Wang, C. C. & Putnam, F. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 115–119.
15. Dwulet, F. E. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2805–2809.
16. Ryden, L. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6767–6771.
17. Richardson, J. S., Thomas, K. A., Rubin, B. H. & Richardson, D. C. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1349–1353.
18. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282,** 189–194.
19. Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G. & Fass, D. N. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6934–6937.
20. Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M. & Capon, D. J. (1984) *Nature (London)* **312,** 337–342.
21. Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1984) *Nature (London)* **312,** 342–347.
22. Jackson, C. M. & Nemerson, Y. (1980) *Annu. Rev. Biochem.* **49,** 765–811.
23. Mann, K. G., Lawler, C. M., Vehar, G. A. & Church, W. R. (1984) *J. Biol. Chem.* **259,** 12949–12951.
24. Atkinson, T. & Smith, M. (1984) in *Oligonucleotide Synthesis: A Practical Approach*, ed. Gait, M. J. (IRL, Oxford), pp. 35-81.
25. Chaconas, G. & van de Sande, J. H. (1980) *Methods Enzymol.* **65,** 75–85.
26. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratories, Cold Spring Harbor, NY).
28. Prochownik, E. V., Markham, A. F. & Orkin, S. H. (1983) *J. Biol. Chem.* **258,** 8389–8394.
29. Huynh, T., Young, R. A. & Davis, R. W., in *DNA Cloning: A Practical Approach*, ed. Glover, D. (IRL, Oxford), pp. 49–78.
30. Goelet, P. & Karn, J. (1984) *Gene* **29,** 331–342.
31. Gubler, U. & Hoffman, B. (1983) *Gene* **25,** 263–269.
32. Fung, M. R., Campbell, R. M. & MacGillivray, R. T. A. (1984) *Nucleic Acids Res.* **12,** 4481–4492.
33. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
34. Russnak, R. & Candido, E. P. M. (1985) *Mol. Cell. Biol.* **5,** 1268–1278.
35. Deininger, P. L. (1983) *Anal. Biochem.* **129,** 216–223.
36. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
37. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
38. Staden, R. (1982) *Nucleic Acids Res.* **10,** 4731–4751.
39. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
40. Wickens, M. & Stephenson, P. (1984) *Science* **226,** 1045–1051.
41. von Heijne, G. (1982) *J. Mol. Biol.* **159,** 537–541.
42. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12,** 5145–5164.
43. Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H. & Lingappa, V. R. (1979) in *Secretory Mechanisms*, eds. Hopkins, C. R. & Duncan, C. J. (Cambridge Univ. Press, London), Vol. 33, pp. 9–36.
44. von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
45. MacGillivray, R. T. A. & Davie, E. W. (1984) *Biochemistry* **23,** 1626–1634.
46. Fung, M. R., Hay, C. W. & MacGillivray, R. T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3591–3595.
47. Cool, D. E., Edgell, C.-J. S., Louie, G. V., Zoller, M. J., Brayer, G. D. & MacGillivray, R. T. A. (1985) *J. Biol. Chem.* **260,** 13666–13676.
48. Knowles, B. B., Howe, C. C. & Aden, D. P. (1980) *Science* **209,** 497–499.
49. Perry, R. P. (1976) *Annu. Rev. Biochem.* **45,** 605–629.