JOURNAL OF BACTERIOLOGY, Feb. 1988, p. 1007–1011
0021-9193/88/021007-05$02.00/0

Vol. 170, No. 2

Isolation and Nucleotide Sequence of the Chloroperoxidase Gene from *Caldariomyces fumago*

MARK J. NUELL,†* GUO-HUA FANG, MILTON J. AXLEY,‡ PAUL KENIGSBERG,§ AND LOWELL P. HAGER

Department of Biochemistry, Roger Adams Laboratory, University of Illinois, Urbana, Illinois 61801

Received 27 July 1987/Accepted 16 November 1987

The chloroperoxidase gene from the filamentous fungus *Caldariomyces fumago* has been isolated within a 16.3-kilobase insert in the vector λEMBL3. The DNA sequence of the gene and its immediate flanking regions has been determined, and the start site of transcription has been mapped by primer extension.

Chloroperoxidase catalyzes the insertion of chlorine, bromine, and iodine atoms into organic acceptor molecules. It is intimately involved in the biosynthesis of chlorinated secondary metabolites in the filamentous fungus *Caldariomyces fumago*. The enzyme is glycosylated and secreted into the culture medium at levels of 500 mg/liter when fructose is used as the carbon source for growth. No enzyme is produced if glucose is used to provide carbon to the cells. Previous studies in our laboratory have shown that the chloroperoxidase mRNA pool varies in concert with the level of secreted enzyme activity (1).

To determine the primary structure of chloroperoxidase and to begin to investigate what sequence elements of the chloroperoxidase gene control the expression and secretion of the protein, we have cloned and sequenced the gene and its flanking regions.

Cloning the chloroperoxidase gene. Total cellular DNA from *C. fumago* was partially digested with the restriction endonuclease *Mbo*I. The resulting fragments were fractionated on a low-melting-point agarose gel, and the 15- to 22-kilobase (kb) fragments were eluted. The eluted fragments were ligated into the *Bam*HI site of bacteriophage λEMBL3, and the recombinant phage were packaged in vitro using the Gigapak system from Stratagene Cloning Systems, San Diego, Calif. Fifty thousand recombinant phage were screened by plaque hybridization, using the insert from the cDNA clone pMA340 (1), labeled by nick translation, as the probe. Six positive clones were plaque purified by successive rounds of plaque hybridization. DNA isolated from each positive clone was digested with *Pvu*II and subcloned into the *Sma*I site of M13mp8. Ninety-six subclones from each λ clone were screened by colony hybridization using the oligonucleotide probe T7MC29 (1). Single-stranded DNA templates prepared from three appropriate subclones were sequenced, using the oligonucleotide T7MC29 as the primer in dideoxynucleotide chain termination reactions (14). The sequence of all three clones matched that of the cDNA clone pMA340, proving that they indeed included part of the structural gene for chloroperoxidase.

Nucleotide sequence of the chloroperoxidase gene. To obtain the sequence of the entire structural gene and its flanking regions, attention was directed to a single λ clone designated λMN11. Fragments of the clone spanning the entire gene were isolated in M13mp8 by using the cDNA probes pMA340 (1) and pGFx79 (4). The panel of M13 subclones was rescreened with pGFx79, a nearly full-length cDNA clone (4). Subclones representing the amino-proximal portion of the gene were found in both orientations, while only one subclone representing the carboxyl-proximal portion was found. Double-stranded replicative form DNA was prepared from the carboxyl-proximal subclone, and the insert was excised by digestion with *Sal*I and *Eco*RI and recloned in M13mp9 to provide template DNA in the opposite orientation. The entire DNA sequence of the chloroperoxidase gene and its immediate flanking regions is presented in Fig. 1. The DNA sequence extends from 819 base pairs upstream from the start of transcription to a point 506 base pairs downstream from the putative polyadenylation site. The nucleotide sequence of the chloroperoxidase gene was analyzed using the software available on the BIONET National Computer Resource for Molecular Biology.

A search of the plant, organelle, and unannotated sequences of the GENBANK and EMBL databases on BIONET revealed no significant homologies with known DNA sequences. A comparison of the genomic sequence with the sequence of a cDNA clone spanning almost the entire gene (4) demonstrated that there are no introns in the chloroperoxidase gene.

Transcription initiation site. Figure 2 shows the results of a primer extension experiment in which the TRANSINIT oligonucleotide was hybridized to *C. fumago* poly(A)$^+$ RNA and then extended in the 5′ direction with reverse transcriptase using the conditions of Boel et al. (2). The TRANSINIT oligonucleotide is complementary to bases 45 to 72 of the chloroperoxidase gene. The products of the primer extension reaction are shown in the lane labeled Ext. Three prominent bands were observed in the primer extension reaction, representing the three T residues in the sequence 5′-gtTggaTgaTga-3′. This sequence is the complement of the message analog sequence presented in Fig. 1. Thus the strongest of the three transcription start sites corresponds to nucleotide 0 in the DNA sequence (Fig. 1). Weaker starts are found at positions +3 and +7. All three sites are in agreement with the general pattern of eucaryotic transcription initiation sites, where an adenine residue is embedded in a stretch of pyrimidines (3). The identity of the primer extension products was confirmed by excising them from a preparative gel and subjecting the eluted DNA to Maxam-Gilbert sequencing (10).

Inspection of the DNA sequence in the region approxi-

* Corresponding author.
† Present address: Laboratory of Molecular Genetics, Gerontology Research Center, 4940 Eastern Avenue, Baltimore, MD 21224.
‡ Present address: Department of Biochemistry, National Heart, Lung and Blood Institute, Bethesda, MD 20205.
§ Present address: Département de Biologie, Centre d'Etudes Nucléaires de Saclay, 91191 Gif sur Yvette, France.

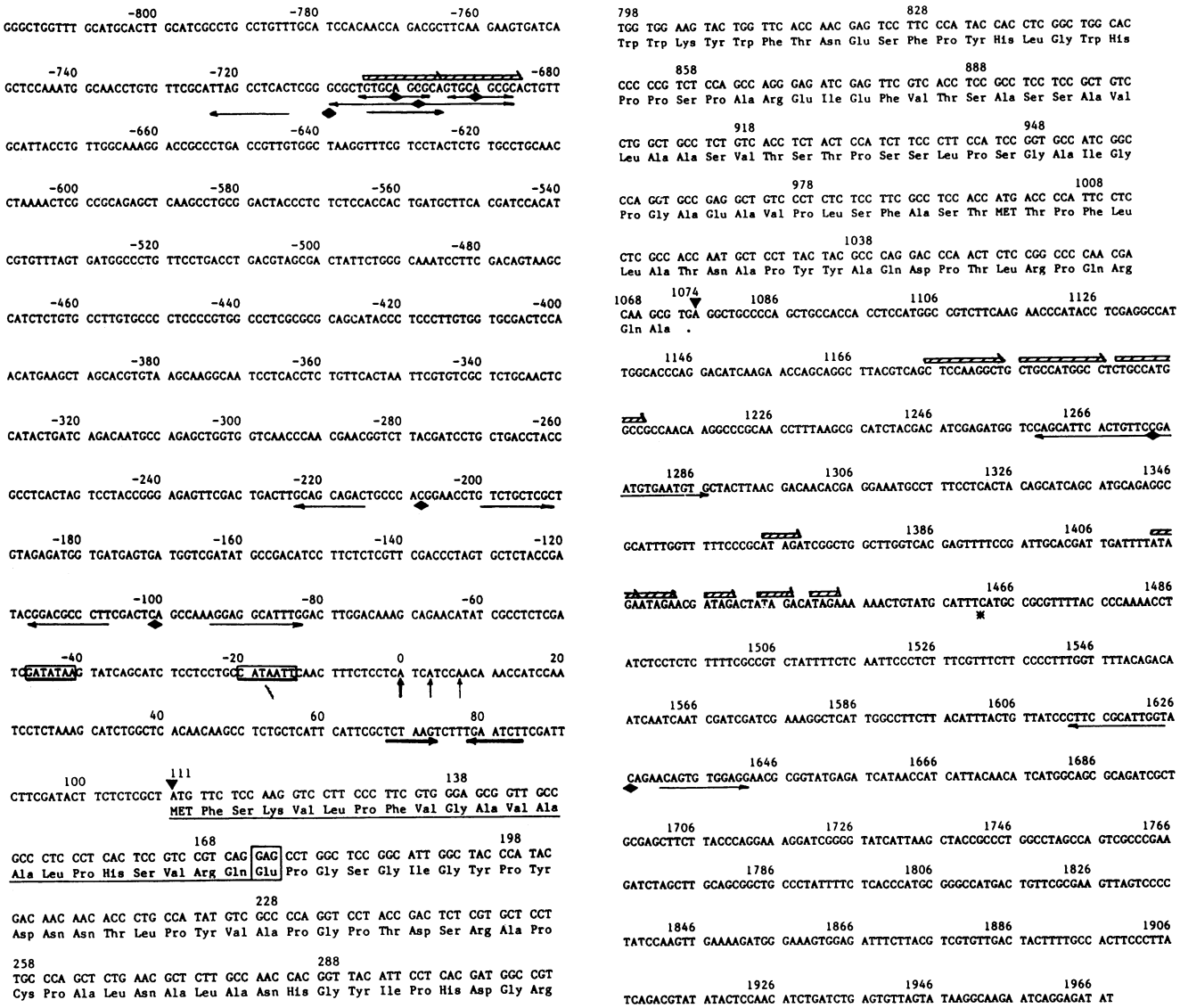

FIG. 1. DNA sequence of the chloroperoxidase gene, determined by the dideoxynucleotide chain termination method. A set of overlapping primers was generated utilizing DNA sequence data from a previous reaction set to choose a sequence to synthesize for use as the next primer. The sequence was analyzed using programs available on the BIONET national computer resource for molecular biology. Regions of dyad symmetry are underlined with arrows (←◆→). The diamond marks the center of the dyad. Direct repeats are overlined with ▭. Inverted repeats are underlined with arrows (→ ←). The transcription initiation sites (↑) were determined by extension of a synthetic oligonucleotide primer, using as a template poly(A)$^+$ mRNA from chloroperoxidase-producing *C. fumago* cells. The two sequences with TATA box homology are boxed (▭). The polyadenylation site of the transcript (*) was inferred from the existence of a poly(A) sequence following position 1462 in a cDNA clone (pGFx79 [4]). The borders of the protein-coding region are indicated with ▼. The amino-terminal glutamic acid residue of the mature chloroperoxidase is boxed (□). The signal peptide preceding the mature amino terminus is simply underlined.

mately 100 bp upstream from the start of transcription revealed two sites showing TATA box-like sequences. The two sites, CATAATT and ATATAA, are located at positions −20 to −14 and −46 to −41. Genes totally lacking a TATA box have been shown to be transcribed from multiple

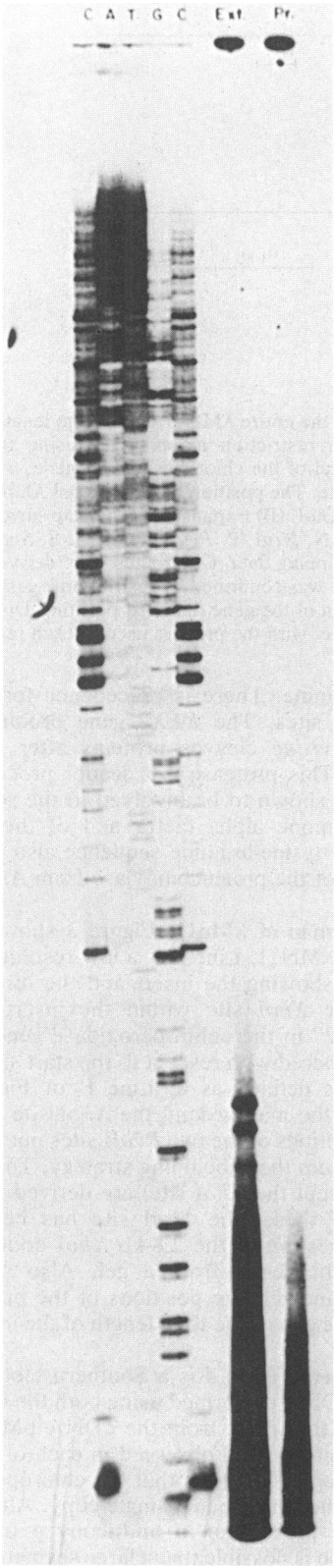

FIG. 2. Mapping the transcription initiation site of the chloroperoxidase gene. The transcription initiation site of the chloroperoxidase gene was determined by the primer extension technique. TRANSINIT oligonucleotide was 5′ labeled with ^{32}P to a specific activity of 3×10^9 cpm/nmol. A 43-pmol sample (1.4×10^8 cpm) was annealed for 3 min at 65°C to 10 μg of poly(A)$^+$ RNA from *C. fumago* cells that had been grown on fructose-containing medium for 3 to 4 days and were shown to be secreting chloroperoxidase. After annealing, the mixture was placed on ice, and reactants for the primer extension reaction were added immediately. Conditions for the reaction are described in reference 2. After the primer extension reaction, the products were ethanol precipitated, dried under vacuum, and redissolved in buffer of the same composition as that used to load the DNA sequencing reactions used as markers. To provide markers for lengths of the products, the TRANSINIT primer was used as the primer for a set of dideoxynucleotide chain-termination sequencing reactions, using a template including the suspected transcription initiation site (M13mp8/λMN11-D11). Lanes: C, A, T, G, and C, dideoxynucleotide sequencing reactions; Ext., primer extension reaction; Pr., labeled primer alone.

initiation sites (6). No CAT or GC box elements were present in the proximal upstream region. None of the five elements shown to date to be sufficient to provide a high basal level of transcription (11) was found in the upstream region of the gene.

Polyadenylation site. A poly(A) sequence, followed by vector sequence, was found in the cDNA of pGFx79 after the sequence GCATTTC (4), which comprises nucleotides 1456 to 1462 of the genomic sequence. Position 1462 is thus considered the putative polyadenylation site of the chloroperoxidase transcript. Nucleotide 1463 is an adenine residue, so this assignment is ambiguous by one nucleotide (see Fig. 1). The usual polyadenylation signal, AATAAA (13), was not found in the nearby upstream region. However, a similar sequence, AAAAAA, is located 13 to 18 nucleotides upstream of the polyadenylation site and might function as the polyadenylation signal for this gene.

Symmetry considerations. Two regions having particularly striking symmetrical structures are found in the genomic clone. The first region lies about 700 base pairs upstream from the start of transcription, involving nucleotides −724 to −685. The basic element in this structure is a small dyad repeat, GTGCAGCGCA. Nucleotides −705 to −685 constitute a direct repeat of the basic element. Nucleotides −709 to −685 represent a slightly enlarged region of dyad symmetry centered between the C and G bases at positions −698 and −697. The region from −704 to −695 is also the downstream arm of an imperfect dyad symmetry centered at the G residue at position −709. Nucleotides −723 to −714 form the upstream arm. The layering of symmetry elements in this region is reminiscent of the structure of the *GAL1-GAL10* promoter (16).

The second striking feature is the string of direct repeats of the sequence ATAGA found in the region from nucleotide 1365 to 1445. This region contains six perfect repeats and one imperfect version of the sequence. The end of this series of repeats includes the first A of the AAAAAA sequence noted above.

The protein coding region. Figure 1 presents the translation of the nucleotide sequence data to an amino acid sequence for chloroperoxidase. These data are consistent with peptide sequence information (9). The codon for the amino-terminal glutamic acid residue of mature chloroperoxidase starts at position 173. The first methionine codon upstream from this position is at nucleotide 110. If this methionine codon is taken as the start site of translation, a rather typical signal peptide coding sequence precedes the amino-terminal glutamic acid residue found in the mature enzyme. Comparison of this putative signal sequence with known signal peptides shows it to have the basic amino-terminal region, a hydrophobic core, and the helix-breaking serine residue typical of

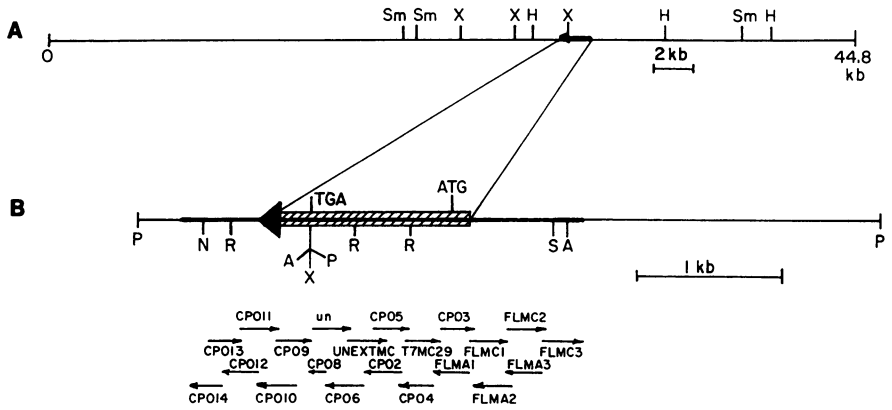

FIG. 3. Restriction map of the λMN11 clone. (A) Low-resolution restriction map of the entire λMN11 clone. The length of the insert is 16.3 kb. The direction of transcription was determined by Southern blotting of a restriction mapping gel using the TRANSINIT oligonucleotide (see the legend of Fig. 2), which hybridizes to a position near the 5′ end of the chloroperoxidase gene, as the probe. The Southern blot identified the restriction fragment containing the transcription initiation site. The position of the internal *Xho*I site was known from DNA sequence data. Abbreviations for restriction sites: H, *Hin*dIII; Sm, *Sma*I; X, *Xho*I. (B) Expansion of the map surrounding the *Xho*I site internal to the chloroperoxidase gene. Abbreviations for restriction sites: A, *Ava*I; N, *Nru*I; P, *Pvu*II; R, *Rsa*I; S, *Sac*I; X, *Xho*I. The *Xho*I site and the three *Pvu*II sites were derived from the subcloning and restriction mapping data. Other sites were derived from a search of the DNA sequencing using the SEQ program of BIONET. Of those sites, the *Nru*I site was confirmed by restriction digestion. Also shown are the positions of the start (ATG) and stop (TGA) codons within the transcribed region of the gene (hatched portion). Diagrammed below the map is the strategy employed in obtaining the sequence of the chloroperoxidase gene, with the primers used in each reaction set noted.

signal peptides (7). However, the glutamine residue at the cleavage site is atypical. The amino acid usually found at signal peptide processing sites is a residue having a small, neutral side chain (15). The residue preceding the cleavage site is an arginine. There is precedence for processing at basic residue sites. The *KEX2* gene product of *Saccharomyces cerevisiae* cleaves proteins after pairs of basic residues (5). This protease is a leader processing enzyme that has been shown to be involved in the secretion of the mating pheromone alpha factor and of the pre-K1 toxin peptide (8). A lysine-arginine sequence also is found at the cleavage site of the preglucoamylase from *Aspergillus awamori* (12).

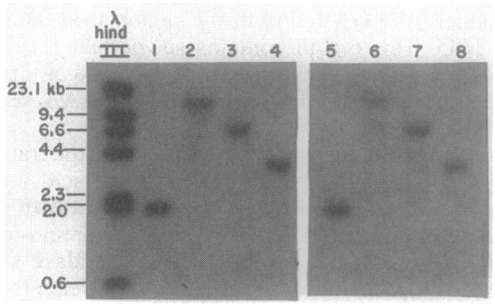

FIG. 4. Southern blots of *C. fumago* total cellular DNA using T7MC29 and insert from cDNA pMA340 as probes. *C. fumago* total cellular DNA was digested to completion with *Ava*I, *Bam*HI, *Eco*RI, or *Pvu*II. The resulting fragments were fractionated in duplicate sets by electrophoresis in a 0.7% agarose gel. ^{32}P-labeled λ *Hin*dIII fragments were used as molecular weight markers. The fragments were electroblotted to a Gene Screen Plus membrane and probed with either 5′-end-labeled oligonucleotide T7MC29 or nick-translated insert from the cDNA pMA340 clone (1). Probes were labeled with ^{32}P to a specific activity of 0.5×10^9 to 1×10^9 cpm/μg. For each hybridization, 1.5×10^6 cpm of probe was added. Hybridization with the oligonucleotide probe T7MC29 was performed at 50°C; hybridization with the pMA340 insert probe was performed at 65°C. Remaining details of the hybridization technique are presented in the "alternate method" in Technical Bulletin NEF-976, available from New England Nuclear Corp. (Boston, Mass.). Stringent wash steps were performed at the hybridization temperatures. ^{32}P-labeled λ *Hin*dIII molecular weight markers are shown in the left lane (the size of each fragment is noted at the left of the figure). Lanes 1 through 4, Restriction fragments of *C. fumago* DNA probed with oligonucleotide T7MC29 (1, *Ava*I; 2, *Bam*HI; 3, *Eco*RI; 4, *Pvu*II); lanes 5 through 8, restriction fragments of *C. fumago* DNA probed with the insert from the cDNA clone pMA340 (5, *Ava*I; 6, *Bam*HI; 7, *Eco*RI; 8, *Pvu*II).

Restriction map of λMn11. Figure 3 shows a restriction map of clone λMN11. Line A is a low-resolution map of the entire clone, showing the insert and the direction of transcription. The *Xho*I site within the insert is located at nucleotide 1127 in the chloroperoxidase gene. The nucleotides are numbered with respect to the start site of transcription, which is defined as 0. Line B of Fig. 3 shows an expansion of the map around the *Xho*I site internal to the gene. The positions of the two *Pvu*II sites outside of the gene are inferred from the subcloning strategy. The other restriction sites (except the *Xho*I site) are derived from the DNA sequence. Of these, the *Nru*I site has been proven by secondary digestion of the 2.8-kb *Xho*I doublet band after excision of that band from a gel. Also shown are the designations and relative positions of the primers used for sequencing the gene. The total length of the insert in λMN11 is 16.3 kb.

Copy number. Figure 4 is a Southern blot of *C. fumago* total cellular DNA performed using both the oligonucleotide T7MC29 and the insert from the cDNA pMA340 clone as probes. The single band observed in each of the restriction digestions strongly suggests that the chloroperoxidase gene is present in the genome in a single copy. Alternatively, the gene may be present on a multicopy extrachromosomal element. It also is possible that a large segment of DNA (>10 kb) containing the chloroperoxidase gene could be repeated without polymorphism. The restriction site data available cannot distinguish between these possibilities. When the complexity of the *C. fumago* genome has been determined, the quantitative hybridization experiments necessary to resolve this question can be done.

We thank Saw Kyin and Charles Mitchell of the Genetic Engineering Facility for their expeditious peptide sequencing and oligonucleotide synthesis. We also thank David Shapiro and John Blume for use of and help with the BIONET system and Stan Bower for Maxam-Gilbert sequencing of the primer extension products.

This work was supported by grants from the National Institutes of Health (GM 07768) and the National Science Foundation (DMB 85-03599).

LITERATURE CITED

1. **Axley, M. J., P. Kenigsberg, and L. P. Hager.** 1986. Fructose induces and glucose represses chloroperoxidase mRNA levels. J. Biol. Chem. **261**:15058–15061.
2. **Boel, E., J. Vuust, F. Norris, K. Norris, A. Wind, J. F. Rehfeld, and K. A. Marcker.** 1980. Molecular cloning of human gastrin cDNA: evidence for evolution of gastrin by gene duplication. Proc. Natl. Acad. Sci. USA **80**:2866–2869.
3. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
4. **Fang, G.-H., P. Kenigsberg, M. J. Axley, M. J. Nuell, and L. P. Hager.** 1986. Cloning and sequencing of chloroperoxidase cDNA. Nucleic Acids Res. **14**:8061–8071.
5. **Fuller, R. S., A. J. Brake, D. Julius, and J. Thorner.** 1985. The *KEX2* gene product required for processing of yeast prepro-α-factor is a calpain-like endopeptidase specific for cleaving at pairs of basic residues, p. 97–102. *In* M. J. Gething (ed.), Current communications in molecular biology, vol. 4. Protein transport and secretion. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Grosschedl, R., and M. L. Birnsteil.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA **77**:1432–1436.
7. **Hunt, E. C., and B. Pesold-Hunt.** 1985. The cleavable prepiece of the cytochrome *c* oxidase subunit IV carries all the information for importing an attached cytosolic protein into the yeast mitochondrial matrix, p. 200–205. *In* M. J. Gething (ed.), Current communications in molecular biology, vol. 4. Protein transport and secretion. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Julius, D., A. Brake, L. Blair, R. Kunisawa, and J. Thorner.** 1984. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-α-factor. Cell **37**:1075–1089.
9. **Kenigsberg, P., G.-H. Fang, and L. P. Hager.** 1987. Post-translational modifications of chloroperoxidase from *Caldariomyces fumago*. Arch. Biochem. Biophys. **254**:409–415.
10. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
11. **Meyers, R. W., D. Tully, and T. Maniatis.** 1986. Fine structure genetic analysis of a β-globin promoter. Science **232**:613–618.
12. **Nunberg, J. H., J. H. Meade, G. Cole, F. C. Lawyer, P. McCabe, V. Schweickart, R. Tal, V. P. Wittman, J. E. Flatgaard, and M. A. Innis.** 1984. Molecular cloning and characterization of the glucoamylase gene of *Aspergillus awamori*. Mol. Cell. Biol. **4**:2306–2315.
13. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
14. **Sanger, F., S. Nicklen, and R. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
15. **Von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
16. **West, R. W., Jr., R. R. Yocum, and M. Ptashne.** 1984. *Saccharomyces cerevisiae GAL1-GAL10* divergent promoter region: location and function of the upstream activating sequence UAS_G. Mol. Cell. Biol. **4**:2467–2478.