ISOLATION AND CHARACTERIZATION OF cDNAS AND GENOMIC DNA CODING FOR HUMAN MYELOPEROXIDASE

Kazuya HASHINAKA

The abbreviations uesd:

Me_2SO, dimethylsulfoxide

EDTA, ethylenediamine tetraacetic acid

SSPE, 0.15 M NaCl, 0.01 M sodium phosphate (pH 7.4), and EDTA.

SSC, 0.15 M NaCl and 0.015 M trisodium citrate (pH 7,0)

Pipes, 1,4-piperzinediethanesulfonic acid

SDS, sodium dodecylsulfate

(k)bp, (kilo)base pair(s)

(k)Da, (kilo)dalton(s)

ABSTRACT

A cDNA in pBR322 encoding the carboxyl-terminal fragment of the human myeloperoxidase heavy chain and three clones of full-length myeloperoxidase cDNA in λgt10 were isolated and characterized. Analysis of the nucleotide sequence of pMP1 cDNA clone in pBR322 showed that the clone contained 1,278 bp with an open reading frame of 474 bp and a 3' noncoding region of 804 bp. The amino acid sequece deduced from the nucleotide sequene consisted of 158 residues including a sequence of 14 amino acids known to be present in the heavy chain of the molecule. The cDNA also included a stop codon of TAG followed by a noncoding sequence that included a potential recognition site for polyadenylation and a poly(A) tail. RNA transfer blot analysis with a fragment of pMP1 indicated that myeloperoxidase mRNA was approximately 3.3 kb in length. In vitro translation of the mRNA selected by cDNA hybridization revealed prefential synthesis of a 74,000-Da polypeptide precursor that could be precipitated with anti-myeloperoxidase IgG. Next, 159 meloperoidase cDNA clones were selected from HL-60 cells cDNA library in λgt10 by hybridization with the cDNA fragment of pMP1.

Analysis of the nucleotide sequence of one of the cDNA clones, λMP-H17, indicated that the cDNA contained 3,207 bp with an open reading frame of 2,238 bp, a 5' noncoding region of 159 bp, a 3' noncoding region of 800 bp, and a poly(A) tail of 10 bp. cDNA of the two other clones, λMP-H7 and λMP-H14, each contained insertions with shorter sequences of 96 and 82 bp, on the open reading frame of λMP-H17 cDNA. A myeloperoxidase genomic clone

was isolated, and the structure of its 5' region was determined and compared with the structures of these cDNAs. Nucleotide sequence analysis of the 3' region of the cDNAs of several clones indicated that the mRNAs were polyadenylated at five different sites. Amino acid sequence determination of amino-terminal and carboxy-terminal portions of the myeloperoxidase light and heavy chains revealed that, during processing of a precursor polypeptide into the mature protein, the amino-terminal polypeptide, the small peptide between the light and heavy chains, and the carboxy-terminal amino acid were excised.

Pluripotent hematopoietic stem cells differentiate to erythrocytes, megakaryocytes, neutrophils, monocytes, eosinophils, basophils, mast cells, and B or T lymphocytes after through several stages of cell divison and maturation (Ogawa et al., 1983). Human neutrophilic granulocytes contain two types of chemically distinct granules, which appear at different stages of maturation. The azurophilic granule appears at the promyelocyte stage and contain myeloperoxidase in addition to numerous lysosomal enzymes, neutral proteases, glycoaminoglycans, cationic bactericidal proteins and lysozyme. The specific granule is formed during the myelocyte stage. It is defined by the absence of peroxidase, and contains lysozyme, lactoferrin and B_{12}-binding proteins (Bainton, 1981). Human promyelocytic leukemia cell line HL-60 cells contain four types of myeloperoxidase (Yamada et al. 1981).

Myeloperoxidase is a tetrameric form of $\alpha_2\beta_2$ subunits consisting of two polypeptides of M_r = 59,000 and 14,000, which carries two heme a type as prosthetic groups (Harrison et al., 1977, Andrews & Krinsky, 1981). Oxidation of chloride ion with myeloperoxidasase was related to the antimicrobial system in leukocytes (Klebanoff, 1975, Harrison & Schultz, 1976). HL-60 cells can be induced to differentiate into granulocytes by Me_2SO or retinoic acid, or into macrophages by TPA and active vitamine D_3 (Collins, 1978, Huberman & Callaham, 1979, Simpson et al., 1987). During differentiations of HL-60 cells by Me_2SO, retinoic acid and TPA, the amount of myeloperoxidase activity per cell and its synthesis decreased (Yamada & Kurahasi, 1984, Koeffler et al., 1985, Kasugai & Yamada, 1986).

Myeloperoxidase synthesis has been studied in HL-60 cells (Yamada, 1982, Yamada & Kurahashi, 1984, Hasilik et al., 1984, Olsson et al., 1984, Koeffler et al., 1985, Akin & Kinkade, 1986). In pulse-chase experiments, myeloperoxidase precursors of 89,000, 81,000, and 74,000 Da were labeled in HL-60 cells and chased into 59,000- and 15,000-Da subunits of the mature enzyme. Furthermore, a 74,000-Da myeloperoxidase polypeptide has been synthesized *in vitro* by translation of HL-60 cells poly(A)$^+$RNA.

As an initial step in elucidating the heterogeneity of myeloperoxidase and promyelocyte-specific expression of the gene, I cloned and characterized various forms of full-length cDNAs encoding myeloperoxidase.

EXPERIMENTAL PROCEDURES

Chemicals. All radiochemicals, and rabbit reticulocyte lysates, a nick translation kit, and a sequecing kit were purchased from Amersham; oligo(dT)$_{12-18}$ and terminal deoxynucleotidyltransferase were from Pharmacia P-L Biochemicals; AMV reverse transcriptase was from Life Science; human placenta RNase inhibitor and RNase H were from Wako Pure Chemical; DNA polymerase I (Klenow fragment) and pBR322 tailed at the Pst I site with dG were from Bethesda Research Laboratories or Toyobo; nuclease S1 was from Sankyo; T4 polynucleotide kinase was from Takara Shuzo and a synthetic oligodeoxynucleotide was made by Takara Shuzo; calf liver tRNA was from Boehringer-Mannheim; and BA 85/21 nitrocellulose filters were from Schleicher and Schuell; Escherichia coli DNA ligase and Eco RI methylase were from New England Biologicals; Eco RI-digested λgt10 DNA and Packergene were from Promega Biotec; carboxypeptidase Y was from Carbiotech; carboxypeptidase A (type I DFP) and carboxypeptidase B (DFP) were from Sigma Chemical Co.

Cell culture. HL-60 cells were grown at 37°C on RPMI-1640 medium supplemented with 10% heat-inactivated fetal calf serum, glutamine, and kanamycin (Yamada, 1982). For induction of differentiation into granulocytes, cells were seeded at 3 x 10^5 cells/ml and grown in the presence of 1.25% Me_2SO or 1 μM retinoic acid for 48 h (Yamada & Kurahashi, 1984).

Isolation and fractionation of poly(A)$^+$RNA. Poly(A)$^+$RNA from HL-60 cells was isolated by a modification of the guanidine thiocyanate-CsCl method as described previously (Yamada &

Kurahashi, 1984). The RNA was loaded into 10.6 ml of a 5-20% (w/v) sucrose gradient in 0.1 M LiCl, 1 mM EDTA, 0.1% SDS, and 0.01 M Tris-HCl (pH 7.5) and centrifuged at 38,000 rpm for 6 h in a Beckman SW 41T rotor at 4°C (Pawson et al., 1977). Fractions of 0.5 ml were collected from the tube and the RNA was precipitated with ethanol. Myeloperoxidase mRNA was identified by *in vitro* translation and immunoprecipitation as described below.

Construction of a cDNA library in pBR322. Double-stranded cDNA was synthesized as described by Maniatis *et al.* (1982). The reaction mixture (50 µl) for synthesis of the first DNA strand consisted of 6 µg of poly(A)$^+$RNA enriched for myeloperoxidase mRNA, 40 units of placental RNase inhibitor, 5 µg of oligo(dT)$_{12-18}$, 100 mM Tris-HCl (pH 8.3), 140 mM KCl, 10 mM $MgCl_2$, 1 mM each of 4 deoxynucleoside triphosphates, 37 µCi of [α-^{32}P]dCTP, 10 mM dithiothreitol, and 48 units of reverse transcriptase and was incubated at 42°C for 90 min. The reaction mixture (100 µl) for synthesis of the second strand consisted of 0.1 M Hepes-NaOH (pH 6.9), 10 mM $MgCl_2$, 2.5 mM dithiothreitol, 70 mM KCl, 0.5 mM each of 4 deoxynucleoside triphsphates, 35 units of Klenow fragment of DNA polymerase I, and single-stranded cDNA and was incubated at 15°C for 20 h. The cDNA was isolated from the mixture treated with 0.049 units of S1 nuclase for 30 min at 37°C. It was then fractionated on a Sepharose CL-4B column (1.8 ml) and precipitated with ethanol.

The cDNA was tailed with poly(dC) (approximately 20 nucleotides) by incubation with 0.5 unit of terminaldeoxynucleotidyltransferase for 30 min at 37°C. The dC-tailed cDNA was isolated and annealed with dG-tailed pBR322 DNA in

a molar ratio of 1:1 in a buffer containing 0.1 M NaCl, 10 mM Tris-HCl (pH 7.8), and 1 mM EDTA for 2 h at 57°C. Escherichia coli DH1 cells were then transformed with the annealed DNA by the method of Hanahan (1983). Transformants formed on SOB agar plates containing tetracycline (15 μg/ml) were identified and transferred to nitrocellulose filters in duplicate on agar plates and grown overnight.

Screening of the cDNA library by differential hybridization. Duplicate filters as described above were fixed and then hybridized with two cDNA probes by the method of Grunstein and Hogness (1975). One probe was synthesized with reverse transcriptase from a fraction of myeloperoxidase mRNA enriched in mRNA by sucrose gradient centrifugation of poly(A)$^+$RNA from HL-60 cells. The other probe was prepared in the same manner from mRNA from a corresponding sucrose gradient fraction of poly(A)$^+$RNA that had been depleted of myeloperoxidase mRNA. Depletion of these cells of the myeloperoxidase mRNA was carried out by treatment of HL-60 cells with Me_2SO for 2 days as described above. The specific radioactivity of the [α-^{32}P]dCTP used in the cDNA synthesis was 125 μCi/542 pmol. The fixed filters were incubated in prehybridization solution consisting of 5 x Denhardt's solution, 6 x SSPE, 0.1% SDS, and 100 μg/ml of heat denatured salmon sperm DNA at 65°C for 6 h. It was then incubated in hybridization solution containing one of the two probes in a prehybridization solution (8 x 10^5 to 2 x 10^6 cpm/filter) at 65°C for 85 h. The 1 x Denhardt's solution consisted of 0.1% Ficoll 400, 0.1% polyvinylpyrolidone K-30, and 0.1% bovine serum albumin. The filters were then washed 9 times with 2 x SSC-0.1% SDS for 10-

min periods at 68°C and twice with 1 x SSC-0.1% SDS for 1.5-h periods at 68°C and twice with 1 x SSC-0.1% SDS for 1.5-h periods at 68°C. The filters were exposed on X-ray film with an intensifying screen (Toshiba DMS-4) at -70°C. Clones showing strong hybridization signals with the enriched probe relative to the depleted probe were selected.

Amino acid sequencing of peptides from myeloperoxidase. Human leukocyte myeloperoxidase III was purified as previously described (Suzuki et al., 1986) and the two subunits were separated by SDS-polyacrylamide gel electrophorsis (Yamada et al., 1981). The subunits were eluted from the gel with 0.05 M NH_4HCO_3/0.05% SDS/0.01% 2-mercaptoethanol (Sreekrishna et al., 1980). The subunits were digested with lysylendopeptidase and the digests were separated by high-performance liquid chromatography (HPLC) on a Bakerbond wide-pore butyl column (0.46 x 25 cm) as described previously (Hirado et al., 1985). The amino acid sequences of several peptides were determined in an Applied Biosystem 470A protein sequencer (Hirado et al., 1985). The amino acid sequences of two peptides were used to synthesize oligonucleotide probes. For amino acids specified by more than one codon, all the possible codons were inserted into ambiguous positions.

Oligonuceleotide screening of the selected cDNA clones. The clones selected as described above were further screened with 17-base oligonucleotide probes containing 32 different sequences. Individual clones were grown overnight in SOB medium containing tetracycline. A pooled culture was made by inoculation of 0.5-ml volumes of each of 10 overnight cultures into 200 ml of fresh

medium and incubation at 37°C for 20 h (Parnes et al., 1981). Plasmid DNA was isolated from the culture and further purified by equilibrium centrifugation in CsCl-ethidium bromide solution. DNA was denatured by incubation in 0.3 M NaOH at room temperature for 1 h, neutralized, and adsorbed on a nitrocellulose filter. The filter was prehybridized at 37°C overnight and then hybridized with 5'-^{32}P-labeled oligonucleotide probe at 37°C for 20 h in the solution described above. The filter was washed twice with 6 x SSC at 30°C for 30-min periods and then twice with 3 M tetramethylammonium chloride/50 mM Tris-HCl (pH 8.0)/2 mM EDTA/0.1% SDS at 51°C for 30-min periods by the method of Wood et al. (1985). Filters were exposed to X-ray film and individual clones in each positive pool were screened in the same way.

cDNA hybridization-selected mRNA and in vitro translation.

cDNA hybridization selction of mRNA was performed as described by Ricciardi et al. (1979). pMP1 DNA (7.0 µg) was denatured in alkali, neutralized, and bound to a nitrocellulose disk of 0.55-cm diameter. The hybridization solution (100 µl) consisted of 65% $HCONH_2$, 18 mM Pipes-NaOH (pH 6.5), 0.2% SDS, 0.4 M NaCl, 43 µg of poly(A)$^+$RNA enriched in myeloperoxidase mRNA, and 6.5 µg of calf liver tRNA and was heated at 68°C for 10 min. Five disks were placed in the solution together and incubated at 50°C for 4 h. The disks were then washed 9 times with 0.5 ml of wash solution containing 10 mM Tris-HCl (pH 7.6)/0.15 M NaCl/1 mM EDTA /0.5% SDS prewarmed at 65°C. They were furthere washed twice with the wash solution without SDS prewarmed at 65°C. The bound RNA was eluted with 300 µl of water containing 5 µg of tRNA by heating at 100°C for 1.3 min and cooling rapidly in dry ice-ethanol. The RNA was

extracted with phenol/$CHCl_3$ and precipitated with ethanol. Translation of the RNA in reticulocyte lysates and analysis of the translation products and anti-myeloperoxidase serum precipitates by SDS-polyacrylamide gel electrophoresis and fluorography were performed as described previously (Yamada & Kurahashi, 1984).

RNA blot analysis. RNA transfer blots and RNA dot blots were analyzed using a ^{32}P-labeled Pst I fragment of 654 bp of cDNA by the method of Thomas (1980). Poly(A)$^+$RNA (1.4 µg/8 µl) isolated from HL-60 cells was denatured in 1 M glyoxal, 50% Me_2SO, and 10 mM sodium phosphate (pH 7.0) at 50°C for 1 h. RNA (0.7 µg) was separated by electrophoresis on a 1.1% agarose gel and then transferred to a nitrocellulose filter. The filter was prehybridized in 100 µl of a solution of 50% (v/v) $HCONH_2$, 1.5 x Denhart's solution, 6 x SSC, 250 µg/ml of denatured salmon sperm DNA, and 10% (w/v) dextran sulfate at 42°C for 21 h and then hybridized with ^{32}P-labeled, Pst I cDNA fragment (2 x 10^7 cpm/75 ng) in a prehybridization solution at 42°C for 63 h. The filter was then washed successively 9 times with 2 x SSC/0.1% SDS for 30 min each at 50°C, and twice with 0.1 x SSC/0.1% SDS for 30 min each at 50°C. The filter was then exposed to X-ray film. Various amounts of native or denatured poly(A)$^+$RNA in 20 x SSC were spotted on nitrocellulose and hybridized to the cDNA fragment in the same way as described above. The Pst I cDNA fragment was labeled with [α-^{32}P]dCTP by nick translation.

Construction and Screening of a cDNA Library in λgt10. Poly(A)$^+$RNA from HL-60 cells was isolated by a modification of the guanidine thiocyanate method (Kasugai & Yamada, 1986) and then was enriched in myeloperoxidase mRNA by 5-20% (w/v) sucrose gradient

centrifugation as described previously (Yamada et al., 1987), except that RNA samples were centrifugated at 24,000 rpm for 14 h at 4°C on a Beckman SW 41T rotor. Double-stranded cDNA was synthesized by the method of Gubler and Hoffman (1983), and a cDNA library was constructed in λgt10 from the cDNA as described by Huynh et al. (1985). Briefly, the first DNA strand was synthesized from the enriched poly(A)$^+$RNA (3 μg) with oligo(dT)$_{12-18}$ and reverse transcriptase in 20 μl of reaction mixture. The second strand was synthesized from the first strand with RNase H, DNA polymerase I, and E. coli DNA ligase. After treatment with T4 DNA polymerase, the cDNA was methylated with Eco RI methylase and then ligated with Eco RI linker. The cDNA was digested with Eco RI restriction endonuclase and fractionated by gel filteration on Bio-Gel A-50 m. The cDNA in the front half of the first peak was recovered by ethanol precipitation and ligated with Eco RI-digested λgt10 DNA. The ligated DNA was then packaged with Packergene and used to infect C600Hfl cells. The constructed cDNA library consisted of 1.7 x 10^6 independent recombinant phages.

The library was screened with a ^{32}P-labeled Pst I fragment (654 bp) of the cDNA from the myeloperoxidase cDNA clone pMP1 by plaque hybridization as described previously (Yamada et al., 1987). In all, 159 cDNA clones showing strong hybridization signals were selected, and several clones with cDNAs of more than 2.5 kbp were selected. The size of cDNA was determined by agarose gel electropresis. The Pst I fragment was labeled with [α-^{32}P]dCTP by the random oligoprimer method (Feinberg & Vogelstein, 1983).

Cloning of Myeloperoxidase Genomic DNA. A human genomic

library in Charon 4A (Lawn et al., 1978) was a gift from Dr. T. Maniatis. About 1.8 x 10^6 phages were screened by the plaque hybridization described by Lawn et al. (1978) with a ^{32}P-labeled cDNA (3207 bp) and a ^{32}P-labeled Eco RI-Xba I fragment (530 bp) from the 5' region of cDNA from λMP-H17 isolated as described above. Several genomic clones were selected. They were amplified and purified by the CsCl banding method (Maniatis et al., 1982).

DNA Sequencing. The insert DNAs of cDNA clones and genomic clones were subcloned in pUC19. The insert DNA was digested with various restriction endonucleases, and then the restriction fragments were subcloned into M13 mp10, mp11, mp18, or mp19 and the DNAs were isolated from the recombinant clones by the method of Messing (1983). The DNA was sequenced with [α-^{35}S]dCTP by the dideoxy chain termination method (Sanger et al., 1977).

Primer Extension Analysis. Enriched poly(A)$^+$RNA (1.8-6.9 μg) and a 17-base oligodeoxynucleotide (ATCCAGCTTCCAAGGAC, 3.3 pmol) were annealed in 2 μl of a solution of 125 mM Tris-HCl, pH 8.3, 25 mM $MgCl_2$, and 12.5 mM dithiothreitol at 60°C for 1 h and then cooled to room temperature. The solution was mixed with a solution (3.8 μl)containing 8.8 μCi of [α-^{35}S]dCTP (13.5 pmol), 6.25 nmol each of dATP, dGTP, and dTTP, 40 nmol of sodium pyrophosphate, and 39 units of reverse transcriptase and incubated at 42°C for 20 min. The reaction was chased with 0.12 μl of 25 mM dCTP for 15 min and then stopped by the addition of 4 μl of 96% formamide solution containing 20 nM EDTA, 0.1% bromophenol blue, and 0.1% xylene cyanol. The sample was boiled for 3 min, loaded on the top of an 8% polyacrylamide/8 M urea sequencing gel, and electrophoresed under the conditions used for sequencing. The

nucleotide length of the primer extension products was determined with reference to a sequence ladder of the 5' region of a genomic DNA formed by sequencing with the same primer as used for primer extension.

Amino Acid Sequencing of the Amino-Terminal and the Carboxy-Terminal Portions of the Light and Heavy Chains of Myeloperoxidase. Human leukocyte myeloperoxidase III was purified as described previously (Suzuki et al., 1986). The two subunits were separated by SDS-polyacrylamide gel electrophoresis and eluted from the gel as described previously (Yamada et al., 1987). The amino-terminal amino acid sequence was determined in an Applied Biosystem 470A protein sequencer (Hirado et al., 1985). For determination of the carboxy-terminal amino acid sequence, the subunit (5 nmol) was digested with a molar ratio of carboxypeptidase Y, A, or B to the substrate of 1/100 to 1/500 in 0.1% SDS solution containing 10 mM sodium phosphate (pH 6.5) for carboxypeptidase Y or 0.2 M N-ethlmorpholine acetate (pH 8.2) for carboxypeptidase A and B at 37°C, and aliquots containing 1 nmol were removed at intervals. The amino acids released were determined in a Hitachi 835S amino acid analyzer. The two subunits were also digested with lysylendopeptidase, the digests were separated by high-performance liquid chromatogrphy (Yamada et al., 1987), and the amino acid sequences of the subunits were determined.

RESULTS

Construction of cDNA Library in pBR322 Poly(A)$^+$RNA of HL-60 cells contains about 0.1% myeloperoxidase mRNA as estimated from amount of myeloperoxidase polypeptide synthesized *in vitro* (Yamada & Kurahashi, 1984). Poly(A)$^+$RNA was further fractionated by centrifugation on a 5-20% (w/v) sucrose gradient and the fracions were then translated in reticulocyte lysates. Myeloperoxidase mRNA was found in the fractions between 28 S and 18 S rRNA (data not shown) and amounted to 0.6%-1% of the mRNA recovered on the basis of the amount of myeloperoxidase polypeptide synthesized. Double stranded cDNA was prepared from mRNA with oligo(dT)$_{12-18}$ as primer using reverse transcriptase and the Klenow-fragment of DNA polymerase I and then treated with nuclease S1 to excise a loop region of the cDNA. The cDNA tailed with about 20 bases of deoxycytidine was inserted into *Pst* I cut poly(dG)-tailed pBR322 DNA and used to transform *E. coli* DH1. About 8,000 tetracycline resistant clones were collected and screened for a cDNA encoding myeloperoxidase.

Screening of cDNA Clones by Differential Hybridization and Synthetic Oligonucleotide Hybridization Myeloperoxidase mRNA is present in HL-60 cells, but not in HL-60 cells treated with Me_2SO (Yamada & Kurahashi, 1984). Two cDNA probes for differential screening were prepared from a sucrose gradient fraction of poly(A)$^+$RNAs of control and Me_2SO-treated HL-60 cells and uesd to screen two replicas of cDNA clones that had been arranged on nitrocellulose filters. The two filters were each hybridized with one of the two cDNA probes. Clones with a stronger hybridization

signal for the enriched probe than for the depleted probe were selected. In this way 410 clones were obtained from the cDNA library. A pair of filters hybridized to the two probes are shown in Figure 1A and B. Arrows in A and B indicate the positions of two clones giving a positive signal for the former probe and a negative signal for the latter. By hybidization with synthetic oligonucleotide these two clones were later shown to be cDNA clones for myeloperoxidase. In all 41 pools each containing 10 individual clones were made from the selected clones and cultured. Plasmid DNAs were isolated from these mixed cultures and adsorbed on nitrocellulose filter. For screening these DNAs, two oligonucleotide probes, A and B, with the amino acid sequences of the myeloperoxidase heavy and light chains, respectively, determined in this study were synthesized. Probe A is a 17-base oligonucleotide containing 32 different sequences and probe B is a 14-base oligonucleotide containing 32 different sequences. The two identical filters were hybridized with probe A and B, respectively. As shown in Figure 1C, only one of 41 pools of DNA gave a positive reaction on hybridization to probe A, and none reacted with probe B (data not shown). Ten clones of the pool were each cultured and plasmid DNAs isolated from the cultures were again sereened with probe A in the same way. Two clones gave a positive reaction on this hybridization (Figure 1D). Plasmid DNAs from these clones were digested with restriction enzyme Pst I. Agarose gel electrophoresis of the digest showed that the three Pst I fragments generated from the cDNA inserts in the two plasmid DNAs were the same, suggesting that the two clones were identical. We characterized one of these two cDNA clones, named

pMP1.

Sequence of the pMP1 cDNA Figure 2 shows the restriction enzyme map and sequencing strategy for the cDNA insert in pMP1 DNA. The cDNA insert was excised into three fragments by agarose gel electrophoresis. Moreover, digestions of pMP1 DNA with a number of restriction endonuclases yielded fragments appropriate for sequencing. These fragments were isolated and both of their strands were cloned either on M13 mp10, mp11, mp18 and mp19 and sequenced. As shown in Figure 2, the sequences of the end of neighbor fragments overlapped and covered the entire region of both strands of the cDNA insert. Figure 3 shows the nucleotide sequence of the cDNA. The cDNA consisted of 1,278 bp containing an open reading frame of 474 bp and a non-coding region of 804 bp. The open reading frame contained the sequence of oligonucleotide probe A used for screening. The amino acid sequence deduced from the nucleotide sequence consisted of 158 amino acid residues and in fact contained the sequence of 14 amimo acid residues determined for the lysylendopeptidase peptide of myeloperoxidase heavy chain as shown by an underline in the Figure 3, and a potential sequence, Asn-X-Ser, for asparagine N-glycosylation. The non-coding sequence contained a poly(A) addition signal sequence AATAAA located 15 bp upstream from the poly(A) addition site, and a poly(A) tail.

In Vitro Translation of pMP1 DNA Hybridization-Selected mRNA For further characterization of the cDNA, pMP1 DNA was immobilized on nitrocellulose discs and hybridized to mRNA enriched for myeloperoxidase mRNA from HL-60 cell poly(A)$^+$RNA and the bound RNA was then eluted from these discs. The RNA was translated in

rabbit reticulocytes lysates. The total translation products of the non-selected and selected mRNAs were revealed by SDS-polyacrylamide gel electrophoresis and fluorography (Figure 4). The non-selected mRNA was translated into many polypeptides ranging from 35,000 Da to 93,000 Da (lane 3), including myeloperoxidase polypeptides, which were shown by precipitation with anti-myeloperoxidase IgG (lanes 6 and 8). However, the selected mRNA was translated mainly into a polypeptide of 74,000 Da (lane 4), which corresponded to the polypeptide precipitated by anti-myeloperoxidase IgG (lane 8). The 42,000 Da polypeptide seen in lanes 4 and 8 was a product derived from myeloperoxidase polypeptide as observed previously (Yamada, 1982). These results indicate that RNA selected by the pMP1 cDNA is myeloperoxidase mRNA.

Hybridization of pMP1 cDNA to HL-60 Cell Poly(A)$^+$RNA

Poly(A)$^+$RNAs isolated from uninduced HL-60 cells or cells induced to differentiate into granulocytes by retinoic acid or Me_2SO were electrophoresed on an agarose gel and transferred to a nitrocellulose filter. The filter was hybridized with a nick translation labeled Pst I fragment of 654 bp from the cDNA of pMP1. Figure 5A shows an autoradiogram of the RNA blot hybridization. Only one band of RNA of 3,300 nucleotide length was seen with uninduced HL-60 cell poly(A)$^+$RNA (lane 1), and no band was detected with retinoic acid or Me_2SO-induced cell poly(A)$^+$RNA (lanes 2 and 3). These results indicate that myeloperoxidase mRNA is 3,300 nucleotides long and is lost in differentiated cells. For assay of myeloperoxidase mRNA, various amounts of denatured or native poly(A)$^+$RNAs from uninduced cells

or differentiation-induced cells were dotted and hybridized to the Pst I fragment (Figure 5B)(from 1 μg to 0.07 μg). Both denatured and native RNA from uninduced cells showed positive signals, as seen in lines a and d, respectively. However, no hybridization signal was detected with RNA blot (0.07 μg) of induced cells (lines, b, c, e, and f). The amount of myeloperoxidase mRNA of retinoic acid or Me_2SO-treted HL-60 cells were estimated to be a few percent of that uninduced HL-60 cells.

Cloning and Sequence Analysis of Various Full-Length cDNAs Encoding Myeloperoxidase. A cDNA library of HL-60 cells was constructed in λgt10. About 4.8×10^5 recombinant phages were screened by use of the cDNA encoding the carboxy terminus myeloperoxidase as a probe, and 254 positive clones were isolated. Four of these that contained a large cDNA were selected for sequence analysis. The clone that had the largest cDNA insert is referred to as λMP-H17, and the three other clones are referred to as λMP-H7, λMP-NY1, and λMP-H14.

The cDNA of λMP-H17 was sequenced. Figure 6 shows the restriction endonuclease map and the sequencing strategy for the cDNA. Both strands were sequenced, and the results are shown in Figure 7. The sequence revealed that the cDNA was 3,207 bp long and containd a 5' noncoding region of 159 bp, and open reading frame, starting at the second ATG codon of the cDNA and ending at a stop codon (TAG), coded for a 745 amino acid polypeptide. The amino acid sequence deduced from the nucleotide sequence is also shown in Figure 7. The 3' noncoding region contained a polyadenylation signal, AATAAA, and ended with a poly(A) tail of 10 bp.

The amino acid sequences of the amino-terminal portions of the light and heavy chains of myeloperoxidase and of several peptides produced by lysylendopeptidase digestions of the light and the heavy chains were determined. The carboxy-terminal sequence of the light chain was determined by digestions with carboxypeptidases Y and B, and that of the heavy chain was determined by digestions with carboxypeptidases Y and A. The results confirmed that the H17 cDNA encodes myeloperoxidase. Furthermore, it became evident that, for formation of the myeloperoxidase light chain (108 amino acid residues) and heavy chain (466 amino acid residues), a large amino-terminal polypeptide (164 amino acid residues), a small peptide (6 amino acid residues) between the light and the heavy chains, and a single amino acid at a carboxy terminus were removed from a primary polypeptide translated by the mRNA (Figure 7).

Structures of cDNAs with Alternative Exons. The four cDNA clones selected as described above were subcloned into the Eco RI site of pUC19 and characterized by comparison of the sizes of the fragments produced by digestions with various restriction enzymes. First, both ends of these cDNAs were sequenced. The sequences of the 5' end regions were all included in the 5' region of the λMP-H17 cDNA sequence. But their 5' ends were located at different sites in H17 cDNA, as shown in Figures 7 and 10. The 3' end regions determined were all identical with that of the H17 cDNA with a slight difference of several nucleotides for the polyadenylation site. From the results, the sizes of the cDNAs were expected to be in the order H17 cDNA, H7 cDNA, NY1 cDNA, and H14 cDNA.

Agarose gel electrophoresis analysis of the sizes of these cDNAs excised by Eco RI digestion revealed that H14 cDNA was a little larger than NY1 cDNA (data not shown), suggesting that there is an extra sequence in the H14 cDNA. For determination of the presence or absence of an extra sequence in other cDNAs, the DNAs of these plasmid clones were digested with various restriction enzymes specific for the cleavage of the cDNAs, and their digests were analyzed by agarose gel electrophoresis. The results are shown in Figure 8. The restriction fragments of the four plasmid DNAs produced by digestions with Eco RI and Xba I are shown in lanes 1-4. The digest of plasmid H17 DNA gave two bands of about 2.7 kb and 530 b (lane 1). The 2.7-kb band included two fragments from the vector DNA and the cDNA. The plasmid H7 digest gave two strong bands of about 2.8 and 2.7 kb and a faint band of 340 b (lane 2). The 2.8-kb fragment was not seen in the plasmid H17 digest and could not be expected from the sequence of H17 cDNA (Figure 7), suggesting the presence of an extra sequence in the fragment of H7 cDNA. The digests of pMP-NY1 and pMP-H14 each gave two bands of about 2.7 kb and 300 b (lane 3 and 4). The small fragment from the NY-1 digest was distinctly smaller than that from the H14 digest (lanes 3 and 4), suggesting the presence of an extra sequence in the fragment of H14 cDNA. The large fragments of the two cDNAs were the same as that of H17 cDNA.

The Nco I digests of the four plasmids each gave three strongly stained bands (lanes 5-8). The largest bands of 3.0-3.3 kb corresponded to a fragment containing a 5' region of cDNAs and a vector DNA portion. The second-largest bands of the four digests were all the same size (lanes 5-8). The third band of the

H7 digest of 1.2 kb was larger than that of the H17 digest (lanes 5 and 6), suggesting the presence of an extra sequence in the 1.2-kb fragment of H7 cDNA. The third bands of the NY1 and H14 digests were the same size as that of the H17 digest (lanes 5, 7, and 8).

The Pst I fragments of the digests of the four plasmids are shown in lanes 9-12. The 2.1-kb band was seen in the H17 digest (lane 9), but the H7 digest showed two bands of 1.6 kb and 500 b instead (lane 10), indicating that at least one additional Pst I site was present in extra sequence of H7 cDNA. The 650-b band was seen in all four plasmid digests. These results revealed that H7 and H14 cDNAs had an extra sequence not present in H17 and NY1 cDNAs and that it was located in different sites (Figure 7).

Next, the two extra sequences and their flanking regions were sequenced, and their extra locations were determined. For sequence determination of the H7 insert, plasmid H7 DNA was digested with Pvu II or with Sac I and Nae I, and then the Pvu II fragment of 376 bp and the Sac I-Nae I fragment of 503 bp were isolated from the digests and were subcloned into M13 DNA. For sequence determination of the H14 insert, plasmid H14 was digested with Eco RI and Xba I, and then the Eco RI-Xba I fragment of 327 bp was isolated from the digest and subcloned into M13 DNA. The sequence strategies and the length of the determined sequence of H7 and H14 cDNAs are shown in Figure 10. The sequences determined are shown in Figure 9. The sequence from H7 cDNA was 96 bp and was located on the 726th nucleotide of H17 cDNA. The sequence from H14 cDNA was 82 bp and was located on the 426th nucleotide. The 5' and 3' flanking sequences of the H7 and H14 inserts were

each identical with the sequences located on both sides of the 725th nucleotide and the 425th nucleotide, respectively, of H17 cDNA.

H7 cDNA contained the same reading frame as H17 cDNA, but it was longer and encoded a 777 amino acid polypeptide, whereas H17 cDNA encoded a 745 amino acid polypeptide. With the 82-b sequence insertion in H14 cDNA, the open reading frame started at the second ATG codon in the same way as in H17 cDNA, but was interrupted with a stop codon shortly after the start (Figure 7). A long open reading frame could start at the sixth ATG codon and was identical with the frame of H17 cDNA. Therefore, the cDNA could encode a 650 amino acid polypeptide. The polypeptide still included a leader peptide (69 amino acid residues) and the myeloperoxidase light and heavy chains.

Structure of Myeloperoxidase Genomic DNA. For clarification of the origins of the three cDNAs, myeloperoxidase genomic DNA was isolated from a human gene library and characterized. Two clones were selected. They both contained genomic DNAs of about 15 kb. They gave the same patterns of restriction fragments with various restriction enzymes, except that the 5' flanking region of the gene was longer in one clone than in the other. The former, named λMPO18, was sequenced. The sequence comprising the 5'flanking region and the 5' region of the genomic DNA determined (Figure 10).

Examination of the nucleotide sequences showed that those of the three cDNAs including the two extra sequences were all in the 5' region of the genomic DNA. Furthemore, the 5' region of the gene was split into four exons and three introns (Figure 11).

Parts of exon 2 and 4 were optional in these three cDNAs. Exon 2 was composed of two contiguous segments. The former half of the exon was present in all three cDNAs, while the latter half was found in H14 cDNA but not in the other cDNAs. Exon 4 was composed of three contiguous segments. The first segment and the third segment were found in all three cDNAs, but the second one was found in only H7 cDNA. These results indicated that these cDNAs could be generated by alternative splicing from a single transcript of a single gene.

Structure of the 3' Region of cDNAs with Different Polyadenylation Sites. Five classes of the 3' ends of cDNA clones are shown in Table 1. During sequencing of the 3' ends of the large cDNA clones described above, two classes of polyadenylation site were found 15- and 22-bp, respectively, downstream from the polyadenylation signal AATAAA located at the 3188th nucleotide (Table 1). Another polyadnylation signal, TATAAA, is known to be present 613-b upstream from this signal (Morishita et al., 1987a). Therefore, for examination of this polyadenylation signal cDNA clones were digested with Eco RI. The digests were electrophoresed on agarose gel and then hybridized to a ^{32}P-labeled Eco RI-Xba I fragment derived from the 5' region of H17 cDNA. The positive clones with inserts of less than 2.7 kbp were selected, and their 3' ends were sequenced. Three more classes of polyadenylation sites were found 10-, 14-, and 17-bp, respectively, downstream from the polyadenylation signal (Table 1).

DISCUSSION

In this work I isolated and characterized cDNAs encoding fragment of human myeloperoxidase. A clone, pMP1, was isolated from a HL-60 cell cDNA library constructed in plasmid pBR322. pMP1 cDNA was identified in the following ways as a cDNA clone coding for the carboxyl-terminal fragment of myeloperoxidase. The amino acid sequence deduced from the nucleotide sequence of pMP1 contained a sequence of 14 amino acid residues in the myeloperoxidases heavy chain uesd to prepare an oligonucleotide probe. In vitro translation of the mRNA selected by cDNA hybridization resulted in preferential formation of a myeloperoxidase polypeptide of 74,000 Da which was precipitated by anitibody against myeloperoxidase. An RNA transfer blot after agarose-gel electrophoresis of poly(A)$^+$RNA showed a band of 3.3 kb. This mRNA is long enough to code for myeloperoxidase polypeptide. Dot hybridization of poly(A)$^+$RNAs isolated from untreated HL-60 cells and Me_2SO- or retinoic acid-treated cells with the cDNA probe indicated that the amount of myeloperoxidase mRNA in the treated cells was only a few percent of that in the untreated cells. This was in good agreement with the finding that translatable myeloperoxidase mRNA assayed in a reticulocyte lysate is not present in HL-60 cells that have been induced to differentiate into granulocytes by Me_2SO or retinoic acid (Yamada & Kurahashi, 1984, Koeffler et al., 1985).

Next I isolated three types of full-length cDNAs encoding myeloperoxidase from cDNA library constructed in λgt10. Sequence analyses of these cDNAs indicated that λMP-H17 cDNA encodes a

myeloperoxidase polypeptide of 745 amino acids residues. This encoded polypeptide is the same as those encoded by the cDNAs reported by others (Morishita et al., 1987a; Johnson et al., 1987). The other cDNA clone, λMP-H7, contained another coding exon in the open reading frame of H17 cDNA and so encoded a polypeptide containing an additional 32 amino acid peptide in the middle of the myeloperoxidase light chain. At least three types of myeloperoxidase from human leukocytes and HL-60 cells are known. Myeloperoxidase III, among them, is a major type in HL-60 cells (Yamada et al., 1981). But as their primary structures are still not known, it is impossible to correlate them with the proteins predicted from the cDNAs. The amino acid sequence of a lysylendopeptidase peptide from the light chain of myeloperoxidase III suggests that myeloperoxidase III might be the protein encoded by H17 and H14 cDNAs, because in the H7 cDNA encoded protein the peptide should be interrupted at the arginine residue by addition of a unique sequence (Figure 7).

The sequence of H14 cDNA revealed that the open reading frame starts at the first ATG codon in exon 3 and can encode a 650 amino acid polypeptide including the myeloperoxidase light and heavy chains. This polypeptide was the same as that encoded by H17 cDNA, except that the preceding peptide of the former began at the 96th amino acid residue on that of the latter. Therefore, the primary translation products from the mRNAs corresponding to these cDNAs should differ in size. However, when poly(A)$^+$RNA from HL-60 cells was translated in a reticulocyte lysate, a myeloperoxidase polypeptide of approximately 74 kDa was detected as a single band on SDS-polyacrylamide gel electrophoresis (Yamada & Kurahashi,

1984). The first ATG codons of the open reading frames from H17 and H7 cDNAs were present in exon 1, and that from H14 cDNA was in exon 3. The latter ATG codon seems more plausible for initiation than the former because the sequence surrounding the initiation codon is the more homologous with consensus sequence for initiation in higher eukaryotes found by Kozak (1987). But it is still unkown whether the two putative initiator codons or only the initiator codon in exon 3 functions in translation of myeloperoxidase mRNA in protein synthesis.

Sequence analysis of a myeloperoxidase genomic DNA showed that these cDNAs all came from the genomic DNA. The gene organization of myeloperoxidase was reported by Morishita et al. (1987b) while this work was in progress. The structure of the 5' region was nearly the same as that found in this work. Futhermore, recently we and others found that a single myeloperoxidase gene is present in the human genome and located on chromosome 17 (Weil et al., 1987; Kudoh et al., 1987). Southern blot analysis of the Eco RI digest of human genomic DNA with myeloperoxidase cDNA as a probe showed the presence of a single Eco RI fragment of 20 kbp (Chang et al., 1986; Morishita et al., 1987a; Kudoh et al., 1987). The present results taken together indicate that these cDNAs are all derived from a single gene by alternative use of the exons. The three patterns of myeloperoxidase RNA splicing shown in this work were also found in various patterns of alternative RNA splicing in a wide variety of mRNAs from the other sources (Breitbart et al., 1987).

Primer extension experiments indicated that the major 5' end of myeloperoxidase mRNA is located 18 nucleotides upstream from

the 5' end of H17 cDNA (data not shown). A similar site for the 5' end of the mRNA was also detected by S1 nuclease protection experiments (Morishita et al., 1987b). Thus, myeloperoxidase gene transcription may be started mainly from the site. Examination of the 5' flanking region for up to about 500 nucleotides upstream from the initiation site showed no authentic sequences of CAAT and TATA responsible for initiation of transcription by eukaryotic RNA polymerase II (Breathnach & Chambon, 1981). However, an Alu sequence (Schmid & Jelinek, 1982), an enhancer core sequence (Weiher et al., 1983), and a TPA responsive element-like sequence (Angel et al., 1987) were found. Studies are required on whether some of these sequences are associated with down regulation of myeloperoxidase gene expression in HL-60 cells by 12-O-tetradecanoylphorbol 13-acetate or retinoic acid (Kasugai & Yamada, 1986; Koeffler et al., 1985; Yamada & Kurahashi, 1984). Several classes of multiple initiation sites for transcription of the mouse band 3 gene and several other genes were observed, and their promoter regions also lacked CAAT and TATA and consisted of multiple GC-rich sequences (Kopito et al., 1987, Maire et al., 1987).

Two species of myeloperoxidase mRNA from HL-60 cells detected by Northern blot analysis were approximately 3.3 or 2.6 kb in length (Chang et al., 1986, Yamada et al., 1987, Johnson et al., 1987, Weil et al., 1987). This size difference between the mRNAs could be interpreted mainly as due to alternative use of two polyadenylation signals, which were separated by approximately 600 nucleotides as shown in this work.

The present studies indicated that heterogeneous mRNAs coding

for myeloperoxidase are formed in various ways from a single gene. The amounts of individual species of these mRNAs were not estimated in this work. Therefore, it is hard to know how much these heterogeneous mRNAs are related with the syntheses of multiple forms of myeloperoxidase. Further studies on the chemical structures of myeloperoxidase are required for clarification of this.

ACNOWLEDGMENTS

I thank Dr. Michiyuki Yamada, Mrs. Chika Nishio, Dr. Sook-Jin Hur, Dr. Fumio Sakiyama, Dr. Susumu Tsunasawa, Dr. Kazuko Tsuneoka, and Dr. Takakiyo Saeki for help and valuable advce during this work. I also thank Dr. Earl W. Davie, Dr. Kiyoshi Kurahashi and Dr. Hiroshi Hatanaka for valuable advice and Dr. Tom Maniatis for sending a human gene library.

REFERENCES

Akin, D. T. & Kinkade, Jr. J. M. (1986) J. Biol. Chem. 261, 8370-8375.

Andrews, P. C., & Krinsky, N. I. (1981) J. Biol. Chem. 256, 4211-4218.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., & Karin, M. (1987) Cell 49, 729-739.

Bainton, D. F., Ullyot, J. L., and Farquar, M. G. (1971) J. Exp. Med. 134, 907-934.

Breathnach, R., & Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.

Breitbart, R. E., Andreadis, A., & Nadal-Ginard, B. (1987) Ann. Rev. Biochem. 56, 467-495.

Chang, K. S., Trujillo, J. M., Cook, R. G., & Stass, S. A. (1986) (1986) Blood 68, 1411-1414.

Collins, S. J. (1987) Blood 70, 1233-1244.

Feinberg, A. P., & Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.

Grunstein, M., and Hogness, D. S. (1975) Proc. Natl. Acad. Sci., U.S.A. 72, 3961-3965.

Gubler, U., & Hoffman, B. J. (1983) Gene 25, 263-269.

Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

Harrison, J. E., Pabalan, and Schultz, J. (1977) Biochim. Biophys. Acta 493, 247-259.

Hasilik, A., Pohlmann, R., Olsen, R. L., and Von Figure, K. (1984) EMBO J. 3, 2671-2676.

Henderson, W. R., & Klebanoff, S. J. (1983) J. Biol. Chem. 258, 13522-13527.

Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M., & Fujii, S. (1985) FEBS Lett. 186, 41-45.

Huberman, E., & Callaham, M. (1979) Proc. Natl Acad. Sci. U.S.A. 76, 1293-1297.

Huynth, T. V., Young, R. A., & Davis, R. W. (1985) in DNA cloning Vol. 1 (Glover, D.M., Ed) pp49-78, IRL Press, Oxford.

Johnson, K. R., Nauseef, W. M., Care, A., Wheelock, M. J, Miller, C., & Rovera, G. (1987) Nucleic Acids Res. 15, 2013-2028.

Kasugai, I., & Yamada, M. (1986) J. Biochem. 100, 381-388.

Koeffler, H. P., Ranyard, J., & Pertcheck, M. (1985) Bood 65, 484-491.

Kopito, R. R., Andersson, M. A., Andersson, M. A., & Lodish, H. F. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 7149-7153.

Kozak, M. (1987) J. Mol. Biol. 196, 947-950.

Kudoh, J., Minoshima, S., Hashinaka, K., Nishio, C., Yamada, M., Shimizu, Y., & Shimizu, N. (1987)(HGM9) Cytogenet. Cell Genet. 46, 641-642.

Lawn, R. M., Fritsh, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) Cell 15, 1157-1174.

Maire, P., Gauton, S, Hakim, V., Gregori, C., Mennecier, F., & Kahn, A. (1987) J. Mol. Biol. 197 425-438.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York.

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

Morishita, K., Kubota, N., Asano, S., Kaziro Y., & Nagata, S. (1987a) J. Biol. Chem. 262, 3844-3851.

Morishita, K., Tsuchiya, M., Asano, S., Kaziro, Y., & Nagata, S. (1987b) J. Biol. Chem. 262, 15208-15213.

Ogawa, M., Porter. P., & Nakahata, T., (1983) Blood 61, 823-829.

Olsson, I. Persson, A. M., and Stromberg, K. (1984) Biochem. J. 223, 911-920.

Parnes, J. R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E., and Seidman, J. G. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2253-2257.

Pawson, T., Harvey, R., and Smith, A. E. (1977) Natuer 268, 416-420.

Ricciardi, R. P., Miller, J. S., and Roberts, B. E. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4927-4931.

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

Schmid, C. W., & Jelinek, W. R. (1982) Science 216, 1065-1070.

Simpson, R. U., Hsu, T., Begley, D. A., Mitchell, B. S. , and Alizadeh, B. N., (1987). J. Biol. Chem. 262, 4104-4108.

Sreekrishna, K., Jones, C. E., Guetzow, K. A., Prasad, M. R., and Joshi, V. C. (1980) Anal. Biochem. 103, 55-57.

Suzuki, K., Yamada, M., Akashi, K. & Fujikura, T. (1986) Arch. Biochem. Biophys. 245, 167-173.

Thomas, P. S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.

Weiher, H., Konig, M., & Gruss, P. (1983) Science 219, 626-631.

Weil, S. C., Rosner, G. L., Reid, M. S., Chisholm, R. L., Farber, N. M., Spitznagel, J. K., & Swanson, M. S. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 2057-2061.

Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 1585-1588.

Yamada, M. (1982) J. Biol. Chem. 257, 5980-5982.

Yamada, M., Mori, M., & Sugimura, T. (1981) Biochemistry 20, 766-771.

Yamada, M., & Kurahashi, K. (1984) J. Biol. Chem. 259, 3021-3025.

Yamada, M., Hur, S.-J., Hashinaka, K., Tsuneoka, K., Saeki, T., Nishio, C., Sakiyama, F., & Tsunasawa, S. (1987) Arch. Biochem. Biophys. 255, 147-155.

Table 1: Multiple Polyadenylation Sites for Myeloperoxidase mRNA[a]

polyadenylation signal sequences and polyadenylation sites	clones
TYPE I	
---GCGAAGTA(2575)TATAAA—10 b—$(A)_{20}$	λMP-NY20
---GCGAAGTATATAAA—14 b—$(A)_{13}$	λMP-H123
---GCGAAGTATATAAA—17 b—$(A)_{21}$	λMP-H80, λGM706[b]
TYPE II	
---TTTATGAT(3188)AATAAA—15 b—$(A)_{25}$	λMP-NY1, pMP-1[c]
---TTTATGATAATAAA—22 b—$(A)_{10-17}$	λMP-H7, H10, H14, H17

a, Classification of the nucleotide sequences of the 3' regions of various cDNAs encoding myeloperoxidase.

b, Reported by Morishita et al. 1987a.

c, Reported by Yamada et al. 1987.

Legends to Figures

Fig. 1 Screening of the cDNA library by preferential cDNA hybridization and subsequent screening by oligonucleotide hybridization. A and B, example of preferential colony hybridization of two replicia filters with two cDNA probes enriched for myeloperoxidase, which was synthesized from sucrose gradient fractions of HL-60 cell poly(A)$^+$RNA; B, hybridization with the cDNA probe depleted of myeloperoxidase, which was synthesized from parallel fractions of Me_2SO-treated HL-60 cell poly(A)$^+$RNA. The arrows in A and B indicate two clones giving a strong hybridization signal with the enriched probe and no signal with the depleted probe. The two were found to be cDNA clones giving positive results on subsequent oligonucleotide screening, as shown in C and D. C, 41 pools, each made of 10 clones selected in the primary screening, were cultured. Plasmid DNAs were prepared from these cltures, and samples of 13 μg of the DNAs were dot-blotted on nitrocellulose filter and hybridized to 5'-^{32}P-labeled 17-base oligonucleotide probe A containing 32 different sequences. D, second round of dot hybridization of 10 individual cloned DNAs of the positive pool in C to the probe A, performed in the same way as in C. Nos. 1-10, plasmid DNAs from individual clones; No. 11, plasmid DNA from the positive pool in C.

Fig. 2 Restriction enzyme map of cDNA of pMP1 and sequencing strategy. The solid line shows the physical length of the cDNA in kilobase pairs. Arrows under the solid line indicate the direction of sequencing and arrow lengths correspond to the

extents of nucleotide sequences determined in each sequencing.

Fig. 3 Nucleotide sequence of cDNA and deduced amino acid sequence of the carboxy-terminal fragment of the myeloperoxidase heavy chain. Line above the base sequence, the region corresponding to oligonucleotide probe A; underline below the amino acid sequence, the sequence determined in the myeloperoxidase heavy chain; solid diamond, potential site for asparagine N-glycosylation; mark ***, stop codon; double underline below nucleotide sequence, poly(A) addition signal sequence.

Fig. 4 pMP1 DNA hybridization-selected mRNA and *in vitro* translation. Poly(A)$^+$RNA enriched for myeloperoxidasae mRNA was hybridized to discs containing pMP1 DNA and the hybridization-selected mRNA was translated in reticulocyte lysates as described under "Materials and Methods". Total translation products and their immunoprecipitates were subjected to SDS-polyacrylamide gel electrophoresis and fluorography. Lane 1, C^{14}-labeled molecular weight markers; Lane 2, total translation products without poly(A)$^+$RNA; Lane 3, total translation products of poly(A)$^+$RNA enriched for myeloperoxidase mRNA; Lane 4, total translation products of the pMP1 DNA hybridization-selected mRNA; Lanes 5 and 6, nonimmune and immune precipitates of the total translation products used in Lane 2, respectively; Lanes 7 and 8, nonimmune and immune precipitates of the total translation products used in Lane 4, respectively.

Fig. 5 Hybridization of cDNA Pst I fragment to RNA transfer

and RNA dot blots. A, poly(A)$^+$RNAs (0.7 μg) isolated from HL-60 cells (Lane 1), retinoic acid-treated cells (Lane 2), and Me_2SO-treated cells (Lane 3) were electrophoresed on 1.1% agarose gel, transferred to a nitrocellulose filter, and hybridized to a nick-translated Pst I fragment of 654 base pairs from the cDNA insert as described under "Materials and Methods". The hybridized blot was autoradiographed. Arrows on the left side of the autoradiogram show the positions of rat liver 18 S and 28 S RNA. B, the various amounts of the above three kinds of poly(A)$^+$RNAs were dot-blotted and hybridized to the Pst I fragment in the same way as in A. Line a, dot blot of denatured poly(A)$^+$RNA of HL-60 cells; Line b, that of retinoic acid treated cells; Line c, that of Me_2SO-treated cells, where Lane 1, 1.0 μg and Lane 3, 0.1 μg RNA. Lines d, e and f, the same, except that native RNAs were blotted with Lane 2, 0.13 μg and Lane 4, 0.07 μg RNA.

Fig. 6 Restriction enzyme map of cDNA of λMP-H17 and sequencing strategy. The solid line shows the physical length of the cDNA in kirobase pairs. The first nucleotide begins at an initiation site of transcription. Closed circles on the line indicate Pvu II sites. Arrows parallel to the solid line indicate the direction and extent of nucleotide of sequencing.

Fig. 7 Nucleotide sequence of cDNA of λMP-H17 and deduced amino acid sequence of the myeloperoxidase. The nucleotide sequence in parenthesis at the 5' end was determined from the genomic DNA. Arrows H7, NY1 and H14 show the 5' ends of the cDNAs of λMP-H7, λMP-NY1 and λMP-H14, respectively. Triangles H14 and

H7 show the positions of the insertion of an extra sequence for the cDNAs of λMP-H14 and λMP-H7. Arrow heads show the amino termini of the light and heavy chains of myeloperoxidase; curved arrows show the carboxy termini of the light and heavy chains for the enzyme. Underlines show the amino acid sequences determined for the myeloperoxidase light and heavy chains and lysylendopeptidase peptides from the light and heavy chains, and broken underlines show the amino acid resides unidentified; underlines Y, B, and A are the sequences determined with carboxypeptidase Y, B, and A, respectively; *** shows the stop codon; the double underlines below nucleotide sequences indicated polyadenylation signal sequences; *, the nucleotide C was replaced by nucleotide T in cDNA of pMP1 (Yamada et al., 1987).

Fig. 8 Agarose gel electrophoresis of restriction enzyme digests of various cDNAs. (A) cDNAs of four clones subcloned in the Eco RI site of pUC19 were digested with Eco RI and Xba I, Nco I, or Pst I. The digests were electrophresed on 1.0% agarose gel, and stained with ethidium bromide. Lanes 1, 5 and 9, pMP-H17; lanes 2, 6 and 10, pMP-H7; lanes 3, 7 and 11, pMP-NY1; lanes 4, 8 and 12, pMP-H14. Lanes 1 to 4, digests with Eco RI and Xba I; lanes 5 to 8, Nco I digests; lanes 9 to 12, Pst I digests. The positions of λ-Hin dIII fragements and of pMP-H17 digests are shown with reference to a standard size marker in kirobasepair. Arrow heads indicate bands with different mobilities. (B) Restrition enzyme maps of four cDNAs from plasimds H17, H7, NY1 and H14. The thick line shows cDNA and the thin line vector DNA. Triangles above lines show the insertion sites of extra sequences

of H7 and H14 cDNAs determined from sequencing data. E, Eco RI; N, Nco I; P, Pst I; X, Xba I.

Fig. 9 Nucleotide sequences of short inserts and their flanking region in H7 and H14 cDNAs and the deduced amino acid sequence. Lines above nucleotide sequence indicate the inserted sequences in H7 cDNA (A) and H14 cDNA (B). The numbers of 726 and 426 at the beginning of the lines indicate the positions of the first nucleotides of H7 and H14 inserts, respectively, on H17 cDNA. *** shows the stop codon.

Fig. 10 Sequence of the 5' flanking region and 5' region of myeloperoxidase genomic DNA. The genomic DNA of MPO18 was digested with various restriction enzymes and the fragments were subcloned in M13 and sequenced. Underlines A, E, T and TA show the Alu sequence, enhancer core sequence, phorbol ester responsive element-like sequence, and degenerated TATA box, respectively; boxed sequences indicate direct repeats flanking the Alu family; double underline shows the sequence for the synthetic oligodeoxynucleotide used in primer extension analysis. Arrow heads show major positions of the 5' ends of myeloperoxidase mRNA detected by primer extension analyses as described under Experimental procedure. Lines H14 and H17 above nucleotide sequence indicate short inserted sequences in H14 and H17 cDNAs, respectively.

Fig. 11 Structural organization of the myeloperoxidase gene 5' region and the three spliced mRNAs. Black boxes of the gene

from a genomic clone λMP018 show constitutive exons; dotted boxes show alternative sequences; solid lines show the 5' flanking region and introns. Three patterns of alternative splicing of mRNA are shown parallel to the gene organization. The splicing pathway is trailed by connecting boxes of exon sequences by a thin line. Striped boxes indicate the part encoding the leader peptide of a precursor; black boxes show the part encoding the light chain and white boxes show the 5' non-coding exon. Splicing was shown by cDNA sequencing. Arrows indicate the direction of sequencing and arrow lengths correspond to the extents of nucleotide sequences determined in each sequencing. Na, Nae I; P, Pst I; Pv, Pvu II; S, Sac I; X, Xba I. The Pvu II site on only exon 4 was shown for clarifiation.

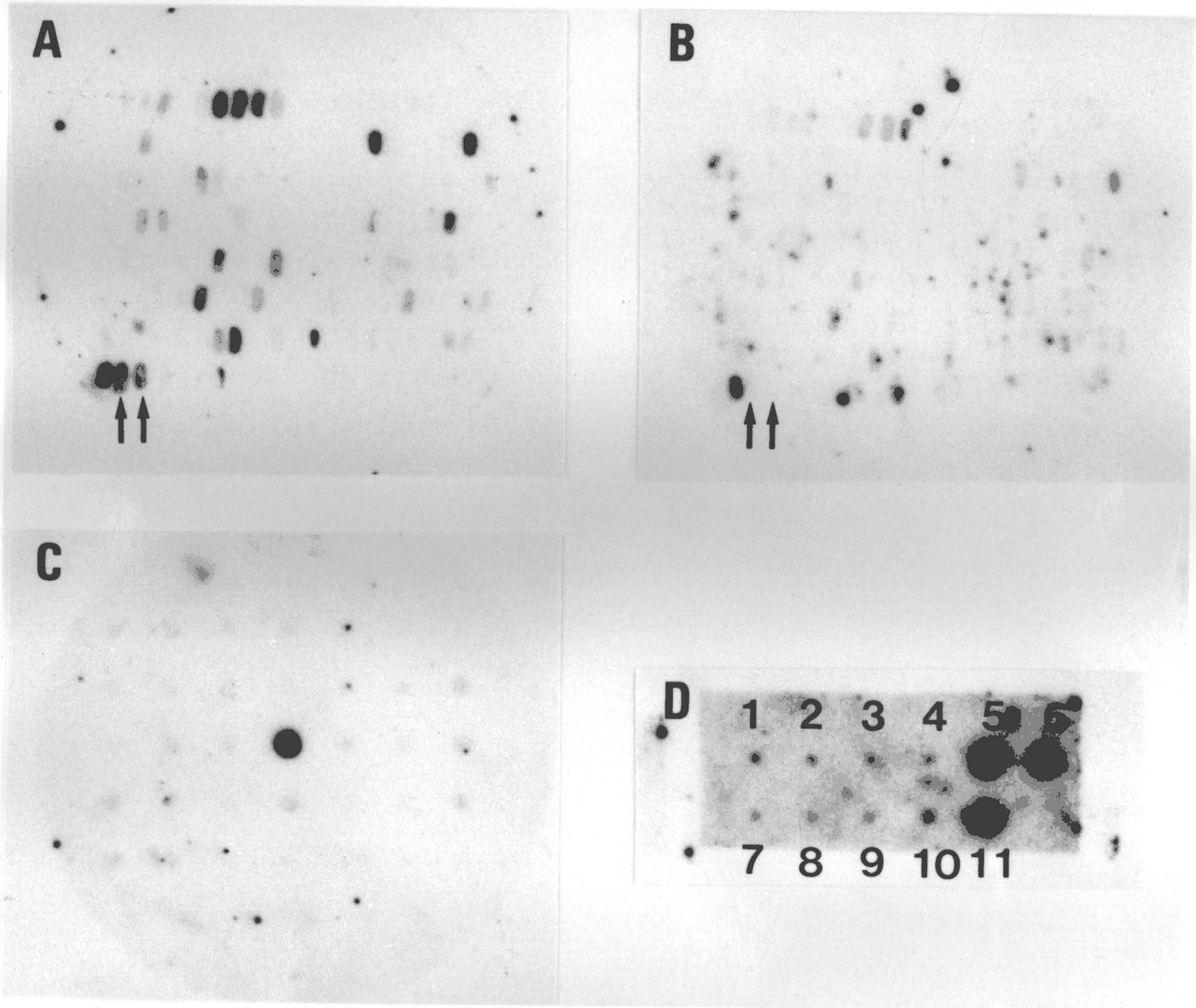

Fig. 1

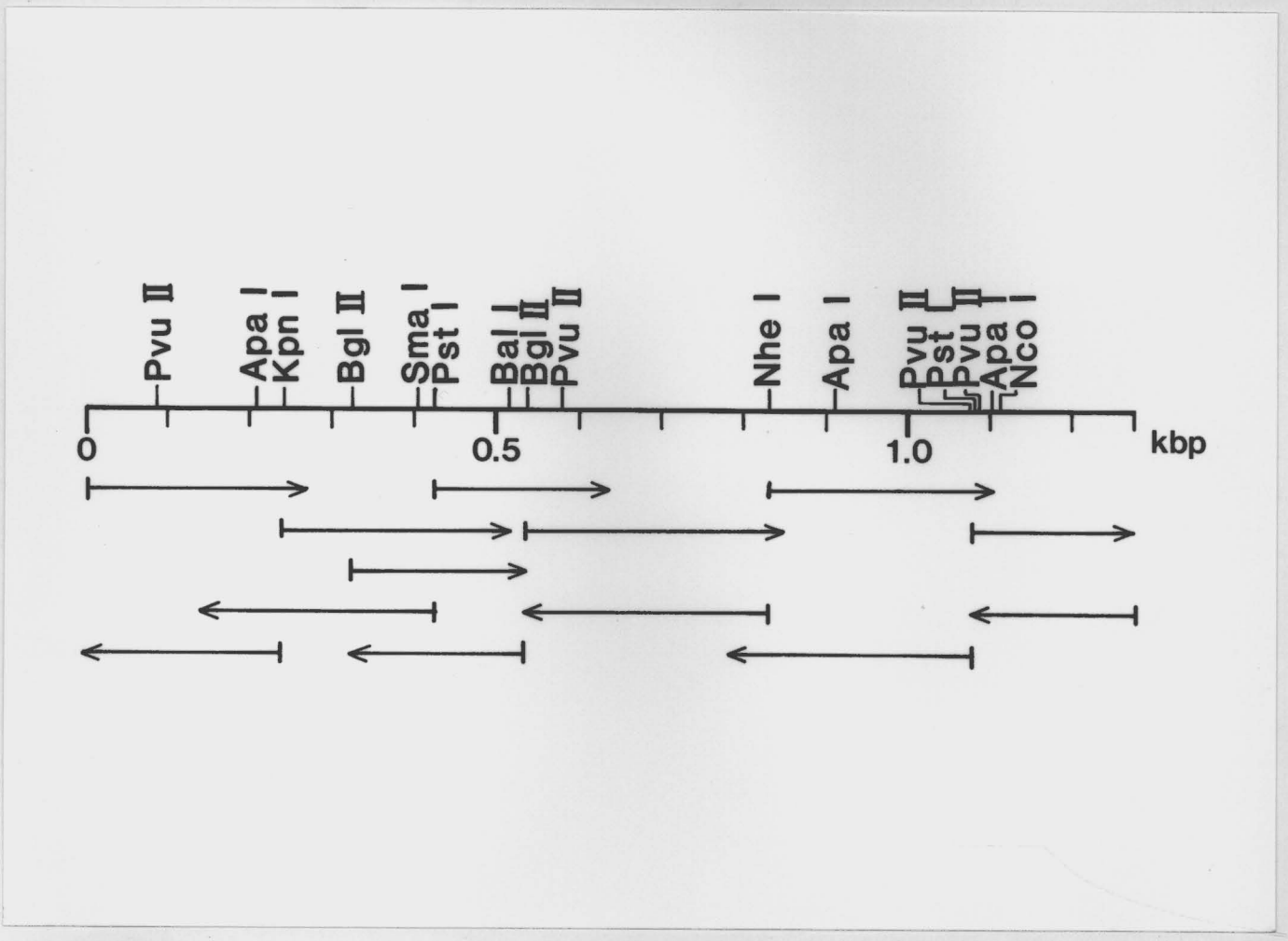
Pvu II
Apa I
Kpn I
Bgl II
Sma I
Pst I
Bal I
Bgl II
Pvu II
Nhe I
Apa I
Pvu II
Pst I
Pvu II
Apa I
Nco I
0
0.5
1.0
kbp

Fig. 2

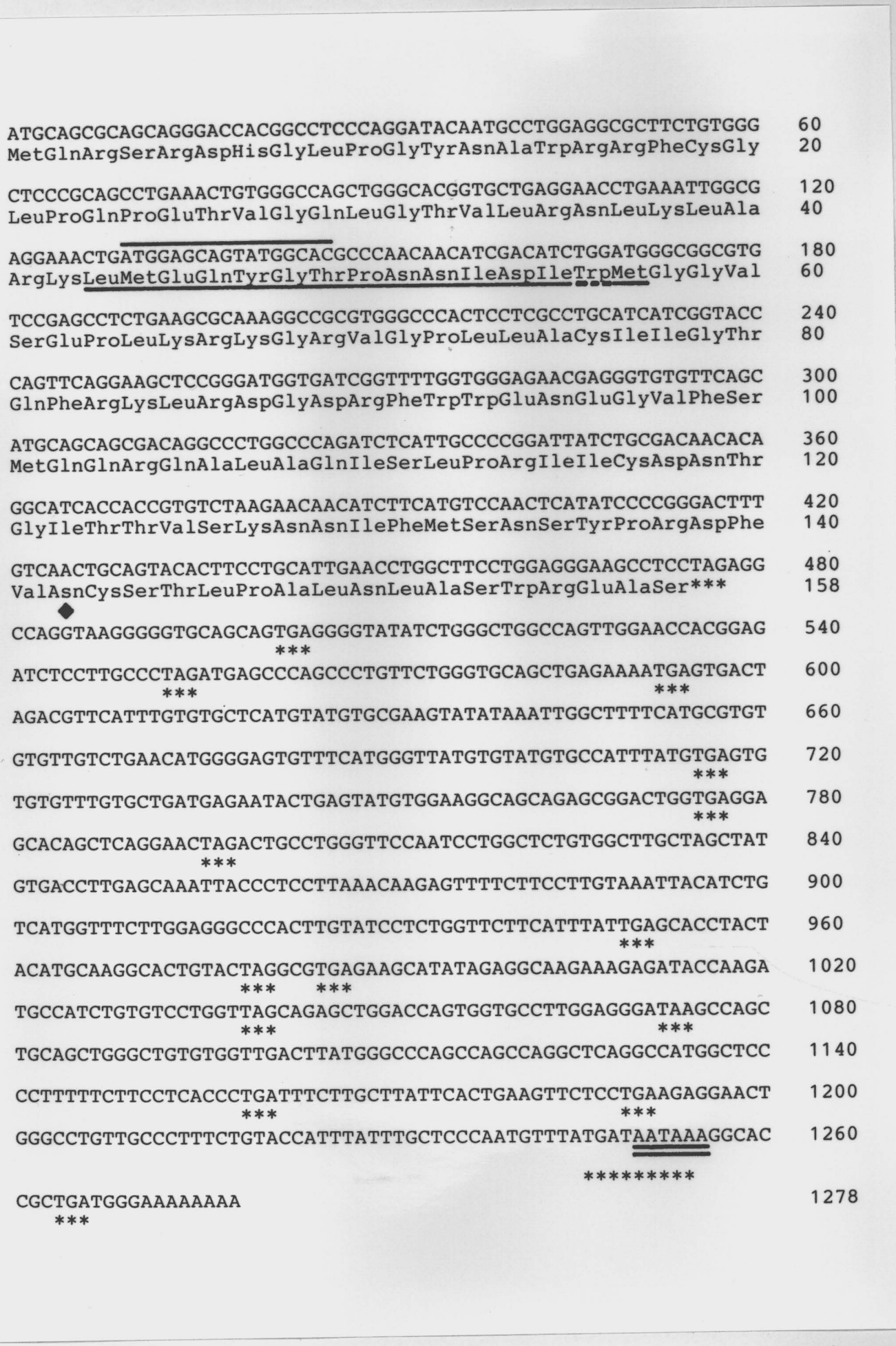
ATGCAGCGCAGCAGGGACCACGGCCTCCCAGGATACAATGCCTGGAGGCGCTTCTGTGGG 60
MetGlnArgSerArgAspHisGlyLeuProGlyTyrAsnAlaTrpArgArgPheCysGly 20
CTCCCGCAGCCTGAAACTGTGGGCCAGCTGGGCACGGTGCTGAGGAACCTGAAATTGGCG 120
LeuProGlnProGluThrValGlyGlnLeuGlyThrValLeuArgAsnLeuLysLeuAla 40
AGGAAACTGATGGAGCAGTATGGCACGCCCAACAACATCGACATCTGGATGGGCGGCGTG 180
ArgLysLeuMetGluGlnTyrGlyThrProAsnAsnIleAspIleTrpMetGlyGlyVal 60
TCCGAGCCTCTGAAGCGCAAAGGCCGCGTGGGCCCACTCCTCGCCTGCATCATCGGTACC 240
SerGluProLeuLysArgLysGlyArgValGlyProLeuLeuAlaCysIleIleGlyThr 80
CAGTTCAGGAAGCTCCGGGATGGTGATCGGTTTTGGTGGGAGAACGAGGGTGTGTTCAGC 300
GlnPheArgLysLeuArgAspGlyAspArgPheTrpTrpGluAsnGluGlyValPheSer 100
ATGCAGCAGCGACAGGCCCTGGCCCAGATCTCATTGCCCCGGATTATCTGCGACAACACA 360
MetGlnGlnArgGlnAlaLeuAlaGlnIleSerLeuProArgIleIleCysAspAsnThr 120
GGCATCACCACCGTGTCTAAGAACAACATCTTCATGTCCAACTCATATCCCCGGGACTTT 420
GlyIleThrThrValSerLysAsnAsnIlePheMetSerAsnSerTyrProArgAspPhe 140
GTCAACTGCAGTACACTTCCTGCATTGAACCTGGCTTCCTGGAGGGAAGCCTCCTAGAGG 480
ValAsnCysSerThrLeuProAlaLeuAsnLeuAlaSerTrpArgGluAlaSer*** 158
CCAGGTAAGGGGGTGCAGCAGTGAGGGGTATATCTGGGCTGGCCAGTTGGAACCACGGAG 540
ATCTCCTTGCCCTAGATGAGCCCAGCCCTGTTCTGGGTGCAGCTGAGAAAATGAGTGACT 600
AGACGTTCATTTGTGTGCTCATGTATGTGCGAAGTATATAAATTGGCTTTTCATGCGTGT 660
GTGTTGTCTGAACATGGGGAGTGTTTCATGGGTTATGTGTATGTGCCATTTATGTGAGTG 720
TGTGTTTGTGCTGATGAGAATACTGAGTATGTGGAAGGCAGCAGAGCGGACTGGTGAGGA 780
GCACAGCTCAGGAACTAGACTGCCTGGGTTCCAATCCTGGCTCTGTGGCTTGCTAGCTAT 840
GTGACCTTGAGCAAATTACCCTCCTTAAACAAGAGTTTTCTTCCTTGTAAATTACATCTG 900
TCATGGTTTCTTGGAGGGCCCACTTGTATCCTCTGGTTCTTCATTTATTGAGCACCTACT 960
ACATGCAAGGCACTGTACTAGGCGTGAGAAGCATATAGAGGCAAGAAAGAGATACCAAGA 1020
TGCCATCTGTGTCCTGGTTAGCAGAGCTGGACCAGTGGTGCCTTGGAGGGATAAGCCAGC 1080
TGCAGCTGGGCTGTGTGGTTGACTTATGGGCCCAGCCAGCCAGGCTCAGGCCATGGCTCC 1140
CCTTTTTCTTCCTCACCCTGATTTCTTGCTTATTCACTGAAGTTCTCCTGAAGAGGAACT 1200
GGGCCTGTTGCCCTTTCTGTACCATTTATTTGCTCCCAATGTTTATGATAATAAAGGCAC 1260
CGCTGATGGGAAAAAAAA 1278

Fig. 3

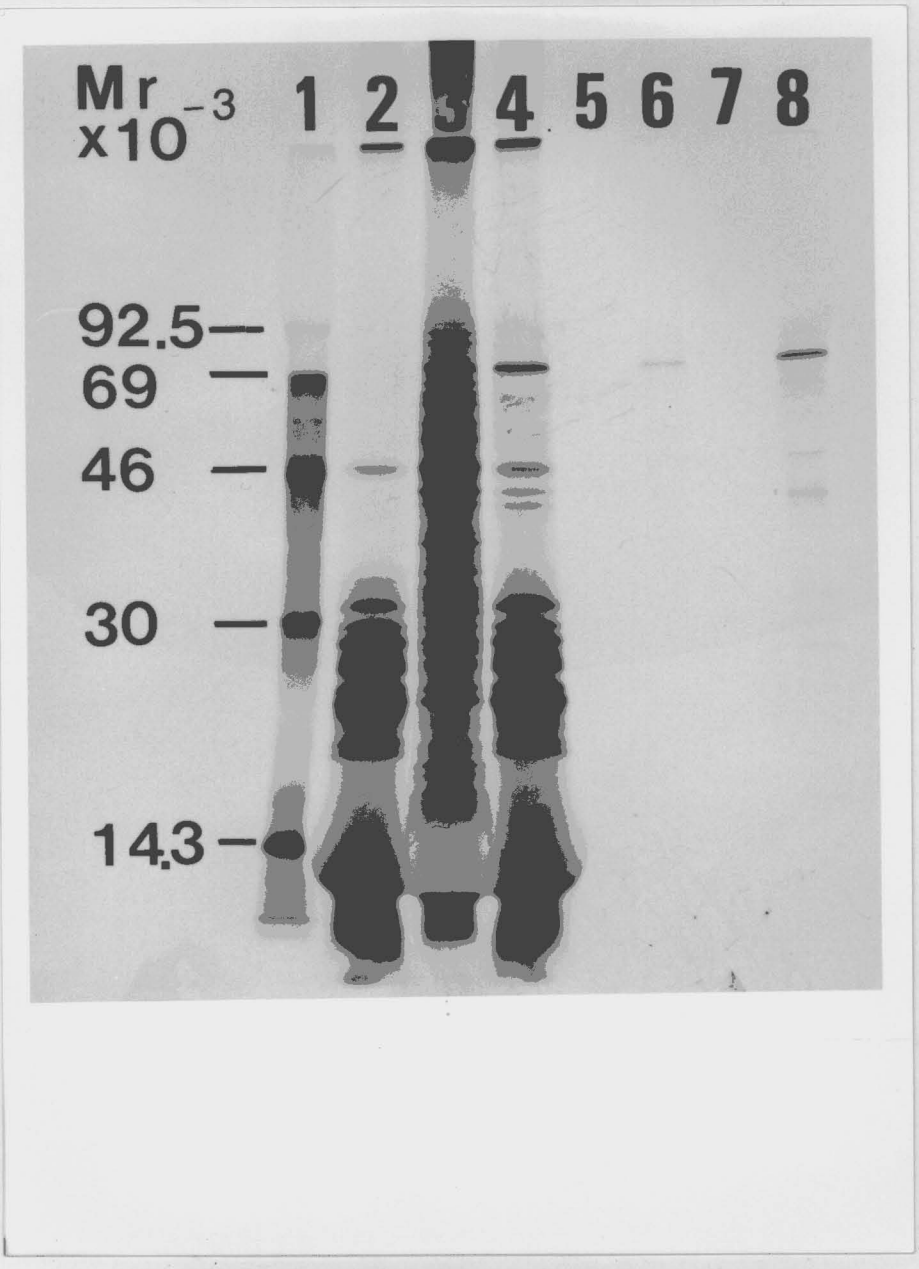

Fig. 4

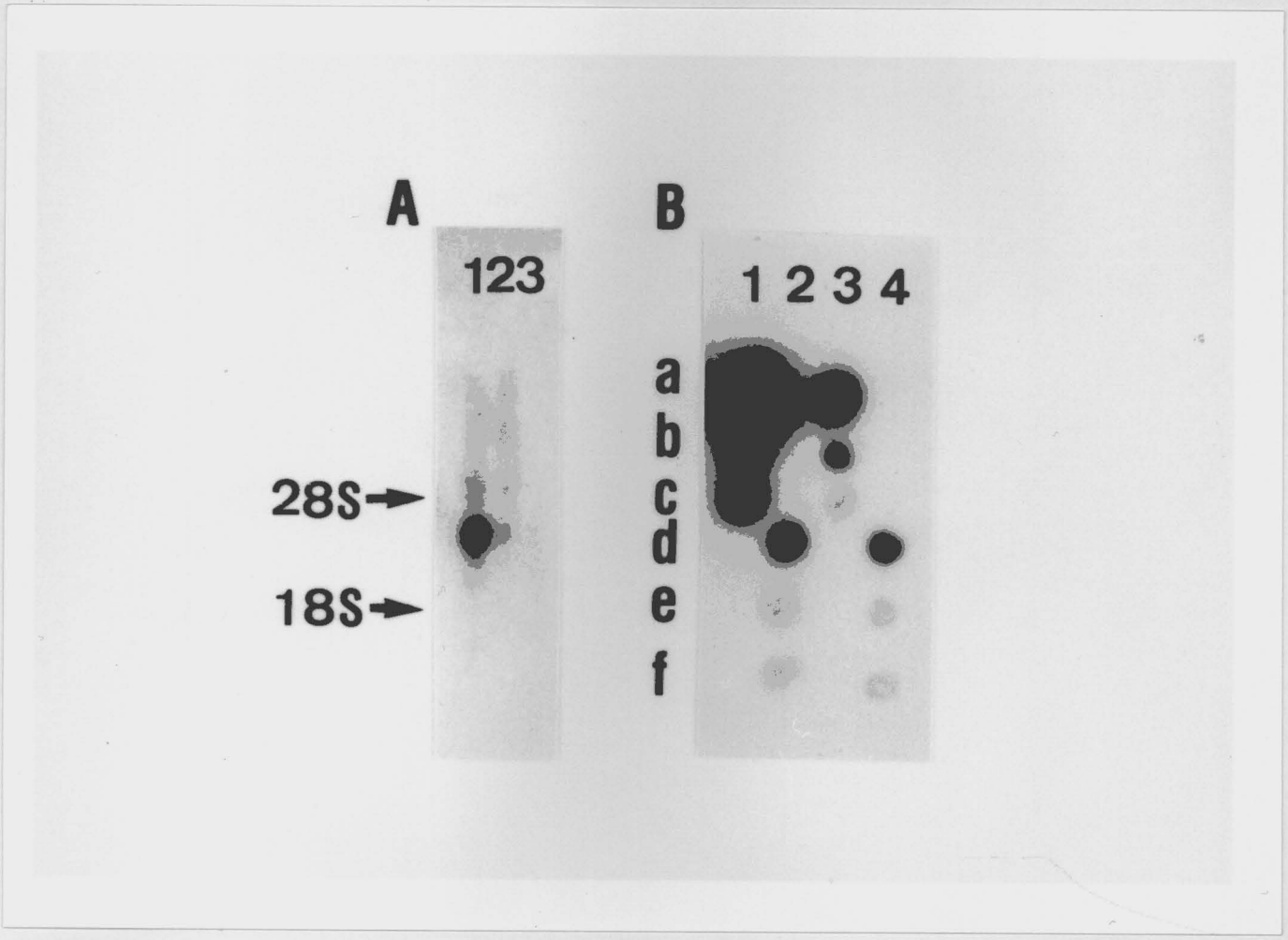

Fig. 5

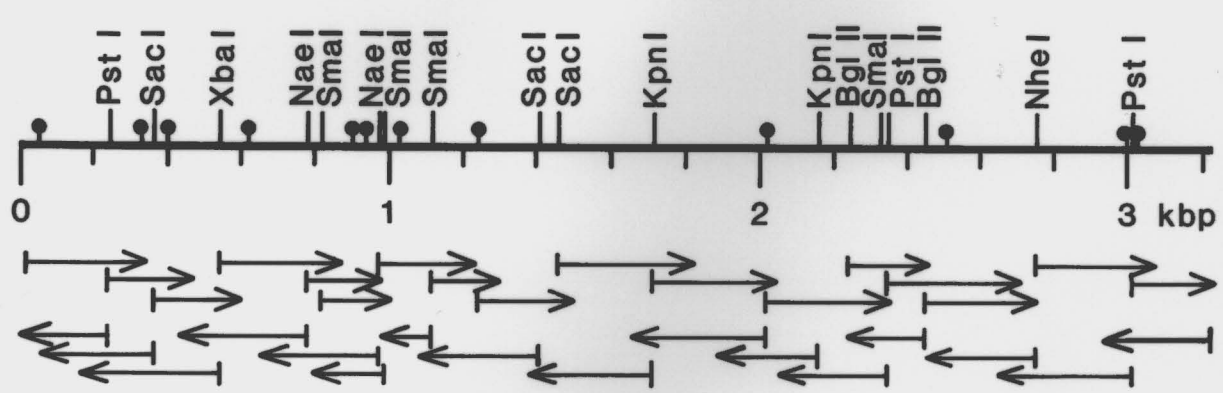
Pst I
Sac I
Xba I
Nae I
Sma I
Nae I
Sma I
Sma I
Sac I
Sac I
Kpn I
Kpn I
Bgl II
Sma I
Pst I
Bgl II
Nhe I
Pst I
0
1
2
3 kbp

Fig. 6

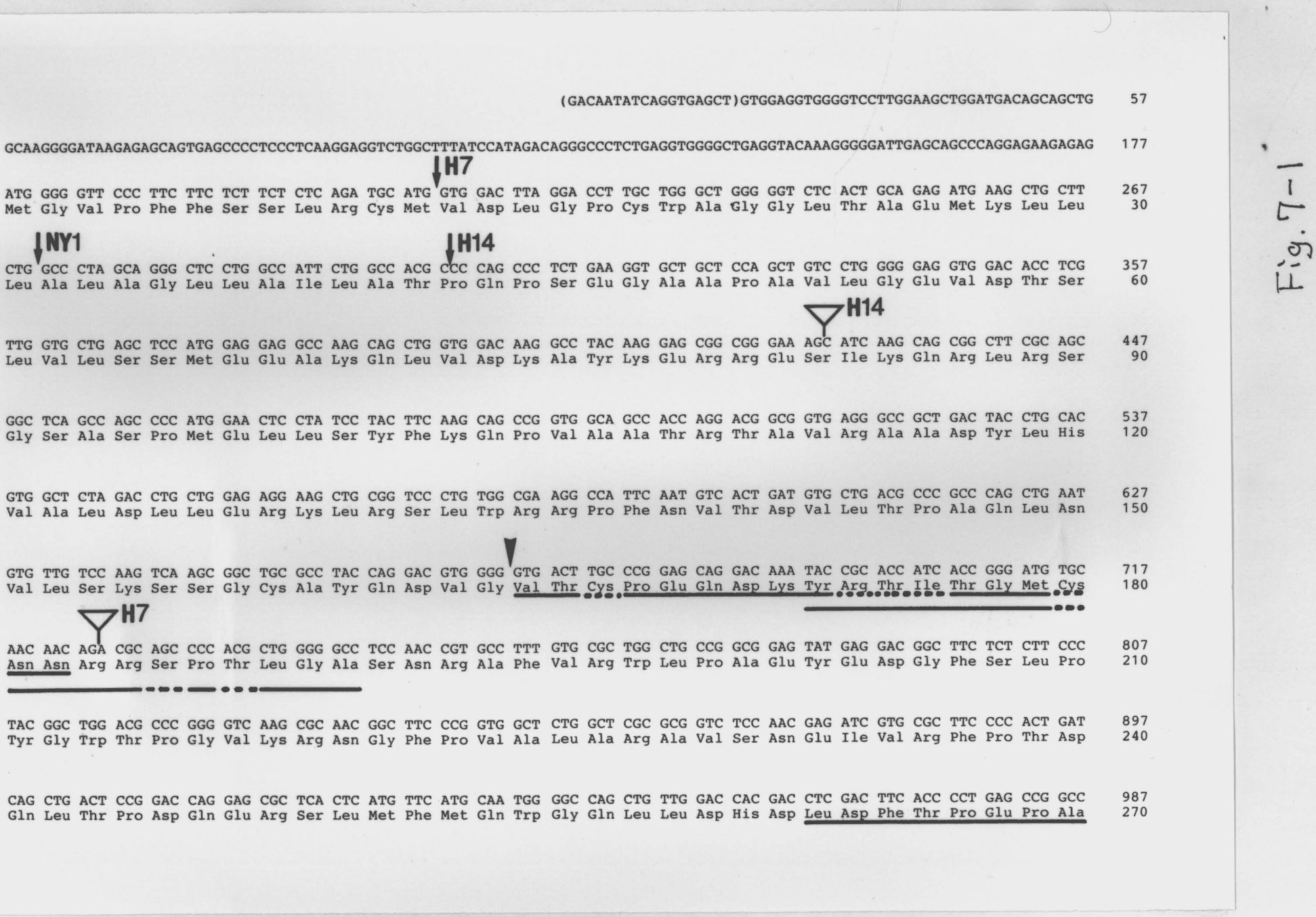
(GACAATATCAGGTGAGCT)GTGGAGGTGGGGTCCTTGGAAGCTGGATGACAGCAGCTG 57
GCAAGGGGATAAGAGAGCAGTGAGCCCCTCCCTCAAGGAGGTCTGGCTTTATCCATAGACAGGGCCCTCTGAGGTGGGGCTGAGGTACAAAGGGGGATTGAGCAGCCCAGGAGAAGAGAG 177
H7
ATG GGG GTT CCC TTC TTC TCT TCT CTC AGA TGC ATG GTG GAC TTA GGA CCT TGC TGG GCT GGG GGT CTC ACT GCA GAG ATG AAG CTG CTT 267
Met Gly Val Pro Phe Phe Ser Ser Leu Arg Cys Met Val Asp Leu Gly Pro Cys Trp Ala Gly Gly Leu Thr Ala Glu Met Lys Leu Leu 30
NY1
H14
CTG GCC CTA GCA GGG CTC CTG GCC ATT CTG GCC ACG CCC CAG CCC TCT GAA GGT GCT GCT CCA GCT GTC CTG GGG GAG GTG GAC ACC TCG 357
Leu Ala Leu Ala Gly Leu Leu Ala Ile Leu Ala Thr Pro Gln Pro Ser Glu Gly Ala Ala Pro Ala Val Leu Gly Glu Val Asp Thr Ser 60
H14
TTG GTG CTG AGC TCC ATG GAG GAG GCC AAG CAG CTG GTG GAC AAG GCC TAC AAG GAG CGG CGG GAA AGC ATC AAG CAG CGG CTT CGC AGC 447
Leu Val Leu Ser Ser Met Glu Glu Ala Lys Gln Leu Val Asp Lys Ala Tyr Lys Glu Arg Arg Glu Ser Ile Lys Gln Arg Leu Arg Ser 90
GGC TCA GCC AGC CCC ATG GAA CTC CTA TCC TAC TTC AAG CAG CCG GTG GCA GCC ACC AGG ACG GCG GTG AGG GCC GCT GAC TAC CTG CAC 537
Gly Ser Ala Ser Pro Met Glu Leu Leu Ser Tyr Phe Lys Gln Pro Val Ala Ala Thr Arg Thr Ala Val Arg Ala Ala Asp Tyr Leu His 120
GTG GCT CTA GAC CTG CTG GAG AGG AAG CTG CGG TCC CTG TGG CGA AGG CCA TTC AAT GTC ACT GAT GTG CTG ACG CCC GCC CAG CTG AAT 627
Val Ala Leu Asp Leu Leu Glu Arg Lys Leu Arg Ser Leu Trp Arg Arg Pro Phe Asn Val Thr Asp Val Leu Thr Pro Ala Gln Leu Asn 150
GTG TTG TCC AAG TCA AGC GGC TGC GCC TAC CAG GAC GTG GGG GTG ACT TGC CCG GAG CAG GAC AAA TAC CGC ACC ATC ACC GGG ATG TGC 717
Val Leu Ser Lys Ser Ser Gly Cys Ala Tyr Gln Asp Val Gly Val Thr Cys Pro Glu Gln Asp Lys Tyr Arg Thr Ile Thr Gly Met Cys 180
H7
AAC AAC AGA CGC AGC CCC ACG CTG GGG GCC TCC AAC CGT GCC TTT GTG CGC TGG CTG CCG GCG GAG TAT GAG GAC GGC TTC TCT CTT CCC 807
Asn Asn Arg Arg Ser Pro Thr Leu Gly Ala Ser Asn Arg Ala Phe Val Arg Trp Leu Pro Ala Glu Tyr Glu Asp Gly Phe Ser Leu Pro 210
TAC GGC TGG ACG CCC GGG GTC AAG CGC AAC GGC TTC CCG GTG GCT CTG GCT CGC GCG GTC TCC AAC GAG ATC GTG CGC TTC CCC ACT GAT 897
Tyr Gly Trp Thr Pro Gly Val Lys Arg Asn Gly Phe Pro Val Ala Leu Ala Arg Ala Val Ser Asn Glu Ile Val Arg Phe Pro Thr Asp 240
CAG CTG ACT CCG GAC CAG GAG CGC TCA CTC ATG TTC ATG CAA TGG GGC CAG CTG TTG GAC CAC GAC CTC GAC TTC ACC CCT GAG CCG GCC 987
Gln Leu Thr Pro Asp Gln Glu Arg Ser Leu Met Phe Met Gln Trp Gly Gln Leu Leu Asp His Asp Leu Asp Phe Thr Pro Glu Pro Ala 270

Fig. 7-1

Fig. 7-1


GCC CGG GCC TCC TTC GTC ACT GGC GTC AAC TGC GAG ACC AGC TGC GTT CAG CAG CCG CCC TGC TTC CCG CTC AAG ATC CCG CCC AAT GAC 1077
Ala Arg Ala Ser Phe Val Thr Gly Val Asn Cys Glu Thr Ser Cys Val Gln Gln Pro Pro Cys Phe Pro Leu Lys Ile Pro Pro Asn Asp 300
Y
B
CCC CGC ATC AAG AAC CAA GCC GAC TGC ATC CCG TTC TTC CGC TCC TGC CCG GCT TGC CCC GGG AGC AAC ATC ACC ATC CGC AAC CAG ATC 1167
Pro Arg Ile Lys Asn Gln Ala Asp Cys Ile Pro Phe Phe Arg Ser Cys Pro Ala Cys Pro Gly Ser Asn Ile Thr Ile Arg Asn Gln Ile 330
AAC GCG CTC ACT TCC TTC GTG GAC GCC AGC ATG GTG TAC GGC AGC GAG GAG CCC CTG GCC AGG AAC CTG CGC AAC ATG TCC AAC CAG CTG 1257
Asn Ala Leu Thr Ser Phe Val Asp Ala Ser Met Val Tyr Gly Ser Glu Glu Pro Leu Ala Arg Asn Leu Arg Asn Met Ser Asn Gln Leu 360
GGG CTG CTG GCC GTC AAC CAG CGC TTC CAA GAC AAC GGC CGG GCC CTG CTG CCC TTT GAC AAC CTG CAC GAT GAC CCC TGT CTC CTC ACC 1347
Gly Leu Leu Ala Val Asn Gln Arg Phe Gln Asp Asn Gly Arg Ala Leu Leu Pro Phe Asp Asn Leu His Asp Asp Pro Cys Leu Leu Thr 390
AAC CGC TCA GCG CGC ATC CCC TGC TTC CTG GCA GGG GAC ACC CGT TCC AGT GAG ATG CCC GAG CTC ACC TCC ATG CAC ACC CTC TTA CTT 1437
Asn Arg Ser Ala Arg Ile Pro Cys Phe Leu Ala Gly Asp Thr Arg Ser Ser Glu Met Pro Glu Leu Thr Ser Met His Thr Leu Leu Leu 420
CGG GAG CAC AAC CGG CTG GCC ACA GAG CTC AAG AGC CTG AAC CCT AGG TGG GAT GGG GAG AGG CTC TAC CAG GAA GCC CGG AAG ATC GTG 1527
Arg Glu His Asn Arg Leu Ala Thr Glu Leu Lys Ser Leu Asn Pro Arg Trp Asp Gly Glu Arg Leu Tyr Gln Glu Ala Arg Lys Ile Val 450
GGG GCC ATG GTC CAG ATC ATC ACT TAC CGG GAC TAC CTG CCC CTG GTG CTG GGG CCA ACG GCC ATG AGG AAG TAC CTG CCC ACG TAC CGT 1617
Gly Ala Met Val Gln Ile Ile Thr Tyr Arg Asp Tyr Leu Pro Leu Val Leu Gly Pro Thr Ala Met Arg Lys Tyr Leu Pro Thr Tyr Arg 480
TCC TAC AAT GAC TCA GTG GAC CCA CGC ATC GCC AAC GTC TTC ACC AAT GCC TTC CGC TAC GGC CAC ACC CTC ATC CAA CCC TTC ATG TTC 1707
Ser Tyr Asn Asp Ser Val Asp Pro Arg Ile Ala Asn Val Phe Thr Asn Ala Phe Arg Tyr Gly His Thr Leu Ile Gln Pro Phe Met Phe 510
CGC CTG GAC AAT CGG TAC CAG CCC ATG GAA CCC AAC CCC CGT GTC CCC CTC AGC AGG GTC TTT TTT GCC TCC TGG AGG GTC GTG CTG GAA 1797
Arg Leu Asp Asn Arg Tyr Gln Pro Met Glu Pro Asn Pro Arg Val Pro Leu Ser Arg Val Phe Phe Ala Ser Trp Arg Val Val Leu Glu 540
GGT GGC ATT GAC CCC ATC CTC CGG GGC CTC ATG GCC ACC CCT GCC AAG CTG AAT CGT CAG AAC CAA ATT GCA GTG GAT GAG ATC CGG GAG 1887
Gly Gly Ile Asp Pro Ile Leu Arg Gly Leu Met Ala Thr Pro Ala Lys Leu Asn Arg Gln Asn Gln Ile Ala Val Asp Glu Ile Arg Glu 570

Fig.7-2

Fig.7-2

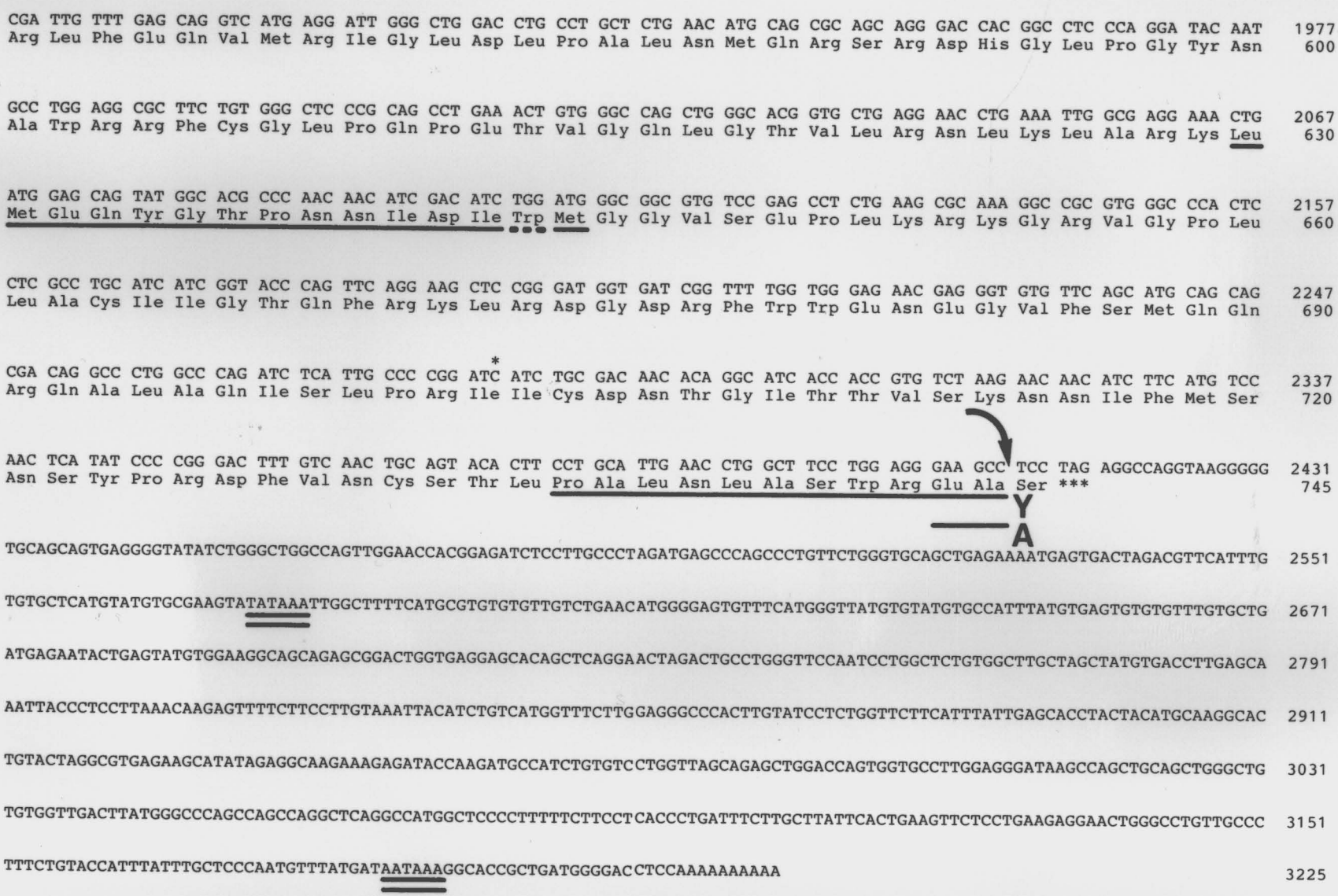
CGA TTG TTT GAG CAG GTC ATG AGG ATT GGG CTG GAC CTG CCT GCT CTG AAC ATG CAG CGC AGC AGG GAC CAC GGC CTC CCA GGA TAC AAT 1977
Arg Leu Phe Glu Gln Val Met Arg Ile Gly Leu Asp Leu Pro Ala Leu Asn Met Gln Arg Ser Arg Asp His Gly Leu Pro Gly Tyr Asn 600
GCC TGG AGG CGC TTC TGT GGG CTC CCG CAG CCT GAA ACT GTG GGC CAG CTG GGC ACG GTG CTG AGG AAC CTG AAA TTG GCG AGG AAA CTG 2067
Ala Trp Arg Arg Phe Cys Gly Leu Pro Gln Pro Glu Thr Val Gly Gln Leu Gly Thr Val Leu Arg Asn Leu Lys Leu Ala Arg Lys Leu 630
ATG GAG CAG TAT GGC ACG CCC AAC AAC ATC GAC ATC TGG ATG GGC GGC GTG TCC GAG CCT CTG AAG CGC AAA GGC CGC GTG GGC CCA CTC 2157
Met Glu Gln Tyr Gly Thr Pro Asn Asn Ile Asp Ile Trp Met Gly Gly Val Ser Glu Pro Leu Lys Arg Lys Gly Arg Val Gly Pro Leu 660
CTC GCC TGC ATC ATC GGT ACC CAG TTC AGG AAG CTC CGG GAT GGT GAT CGG TTT TGG TGG GAG AAC GAG GGT GTG TTC AGC ATG CAG CAG 2247
Leu Ala Cys Ile Ile Gly Thr Gln Phe Arg Lys Leu Arg Asp Gly Asp Arg Phe Trp Trp Glu Asn Glu Gly Val Phe Ser Met Gln Gln 690
CGA CAG GCC CTG GCC CAG ATC TCA TTG CCC CGG ATC ATC TGC GAC AAC ACA GGC ATC ACC ACC GTG TCT AAG AAC AAC ATC TTC ATG TCC 2337
Arg Gln Ala Leu Ala Gln Ile Ser Leu Pro Arg Ile Ile Cys Asp Asn Thr Gly Ile Thr Thr Val Ser Lys Asn Asn Ile Phe Met Ser 720
AAC TCA TAT CCC CGG GAC TTT GTC AAC TGC AGT ACA CTT CCT GCA TTG AAC CTG GCT TCC TGG AGG GAA GCC TCC TAG AGGCCAGGTAAGGGGG 2431
Asn Ser Tyr Pro Arg Asp Phe Val Asn Cys Ser Thr Leu Pro Ala Leu Asn Leu Ala Ser Trp Arg Glu Ala Ser *** 745
Y
A
TGCAGCAGTGAGGGGTATATCTGGGCTGGCCAGTTGGAACCACGGAGATCTCCTTGCCCTAGATGAGCCCAGCCCTGTTCTGGGTGCAGCTGAGAAAATGAGTGACTAGACGTTCATTTG 2551
TGTGCTCATGTATGTGCGAAGTATATAAATTGGCTTTTCATGCGTGTGTGTTGTCTGAACATGGGGAGTGTTTCATGGGTTATGTGTATGTGCCATTTATGTGAGTGTGTGTTTGTGCTG 2671
ATGAGAATACTGAGTATGTGGAAGGCAGCAGAGCGGACTGGTGAGGAGCACAGCTCAGGAACTAGACTGCCTGGGTTCCAATCCTGGCTCTGTGGCTTGCTAGCTATGTGACCTTGAGCA 2791
AATTACCCTCCTTAAACAAGAGTTTTCTTCCTTGTAAATTACATCTGTCATGGTTTCTTGGAGGGCCCACTTGTATCCTCTGGTTCTTCATTTATTGAGCACCTACTACATGCAAGGCAC 2911
TGTACTAGGCGTGAGAAGCATATAGAGGCAAGAAAGAGATACCAAGATGCCATCTGTGTCCTGGTTAGCAGAGCTGGACCAGTGGTGCCTTGGAGGGATAAGCCAGCTGCAGCTGGGCTG 3031
TGTGGTTGACTTATGGGCCCAGCCAGCCAGGCTCAGGCCATGGCTCCCCTTTTTCTTCCTCACCCTGATTTCTTGCTTATTCACTGAAGTTCTCCTGAAGAGGAACTGGGCCTGTTGCCC 3151
TTTCTGTACCATTTATTTGCTCCCAATGTTTATGATAATAAAGGCACCGCTGATGGGGACCTCCAAAAAAAAAAA 3225

Fig.7-3

Fig.7-3

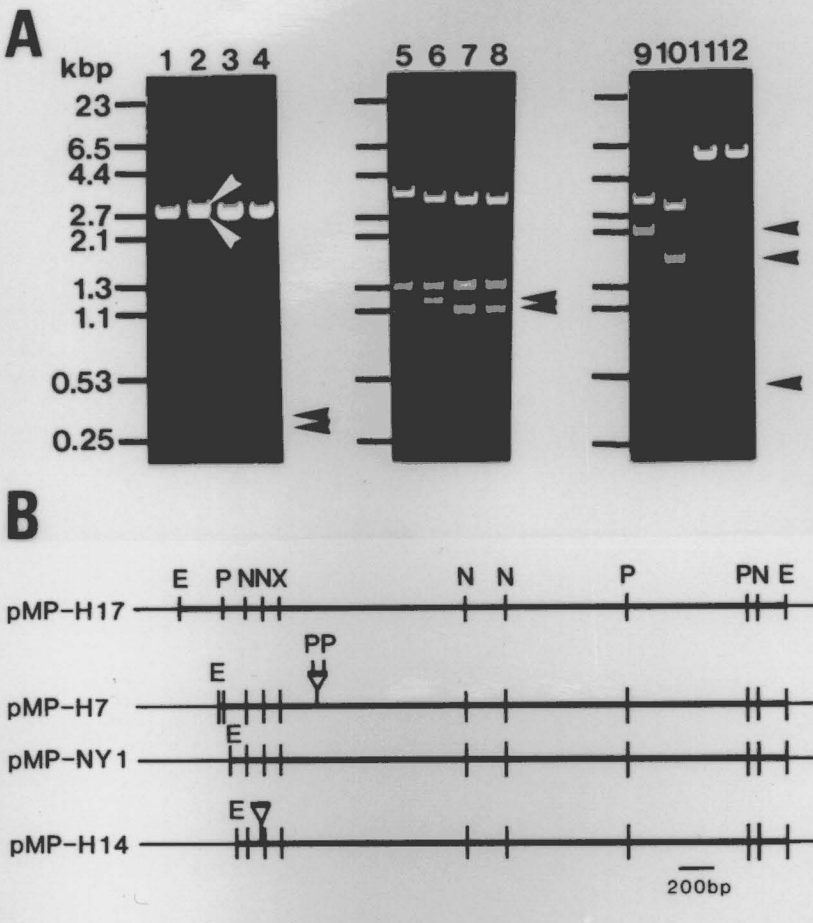
A
kbp
1 2 3 4
5 6 7 8
9 10 11 12
23
6.5
4.4
2.7
2.1
1.3
1.1
0.53
0.25
B
E P N N X
N N
P
P N E
pMP-H17
PP
E
pMP-H7
E
pMP-NY1
E
pMP-H14
200bp

Fig. 8

A λMP-H7

```
Pvu II
CAG CTG AAT GTG TTG TCC AAG TCA AGC GGC TGC GCC TAC CAG GAC   45
Gln Leu Asn Val Leu Ser Lys Ser Ser Gly Cys Ala Tyr Gln Asp   15

GTG GGG GTG ACT TGC CCG GAG CAG GAC AAA TAC CGC ACC ATC ACC   90
Val Gly Val Thr Cys Pro Glu Gln Asp Lys Tyr Arg Thr Ile Thr   30

                    726                              Pst I
GGG ATG TGC AAC AAC AGG TGC GGC TGG CTG GGG GTG GCT GCA GGA  135
Gly Met Cys Asn Asn Arg Cys Gly Trp Leu Gly Val Ala Ala Gly   45

ACC GGG CTC AGA GAG GCG TCC CGG ACG CCA CAA GCC TCC CGG TGT  180
Thr Gly Leu Arg Glu Ala Ser Arg Thr Pro Gln Ala Ser Arg Cys   60

                        Pst I
CAG CGC CCT GTC CTC CCC TGC AGA CGC AGC CCC ACG CTG GGG GCC  225
Gln Arg Pro Val Leu Pro Cys Arg Arg Ser Pro Thr Leu Gly Ala   75

                                  Nae I
TCC AAC CGT GCC TTT GTG CGC TGG CTG CCG GC                   257
Ser Asn Arg Ala Phe Val Arg Trp Leu Pro                       85
```

B λMP-H14

```
   Sac I
G AGC TCC ATG GAG GAG GCC AAG CAG CTG GTG GAC AAG GCC TAC    43
  Ser Ser Met Glu Glu Ala Lys Gln Leu Val Asp Lys Ala Tyr    14

                426
AAG GAG CGG CGG GAA AGG TGG GGC ACG AGG CAC TGC CCA GCT CTG   88
Lys Glu Arg Arg Glu Arg Trp Gly Thr Arg His Cys Pro Ala Leu   29

GGC AAG GAT GTC CCA GGC CTT TCA GAG AAG CAG CAG GCA GCA GGG  133
Gly Lys Asp Val Pro Gly Leu Ser Glu Lys Gln Gln Ala Ala Gly   44

AGC TTG AAG CAT CAA GCA GCG GCT TCG CAG CGG CTC AGC CAG CCC  178
Ser Leu Lys His Gln Ala Ala Ala Ser Gln Arg Leu Ser Gln Pro   59

CAT GGA ACT CCT ATC CTA CTT CAA GCA GCC GGT GGC AGC CAC CAG  223
His Gly Thr Pro Ile Leu Leu Gln Ala Ala Gly Gly Ser His Gln   74

                                        Xba I
GAC GGC GGT GAG GGC CGC TGA CTACCTGCACGTGGCTCTAGA            265
Asp Gly Gly Glu Gly Arg ***                                   80
```

Fig. 9

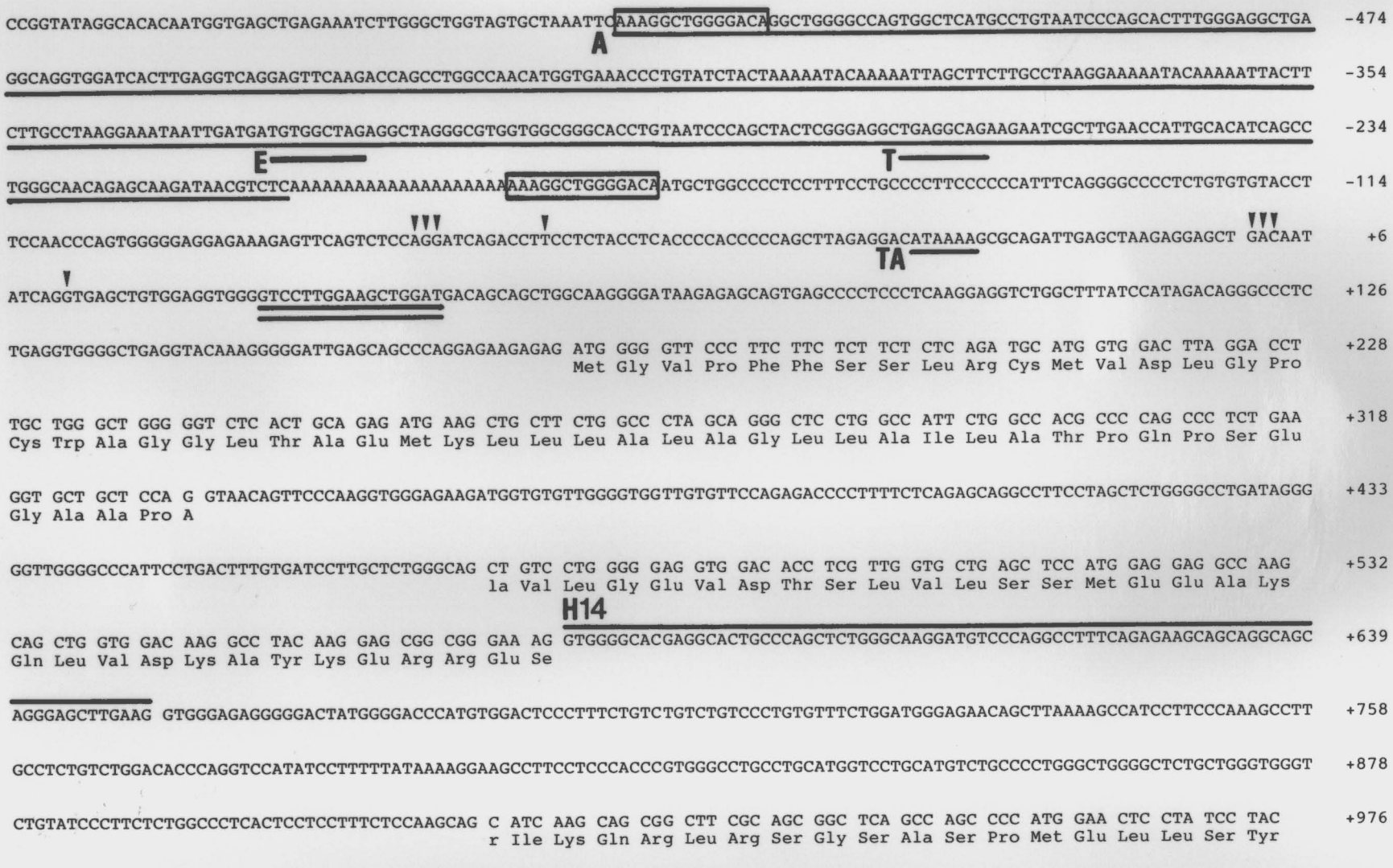

Fig. 10-1

Fig. 10-1

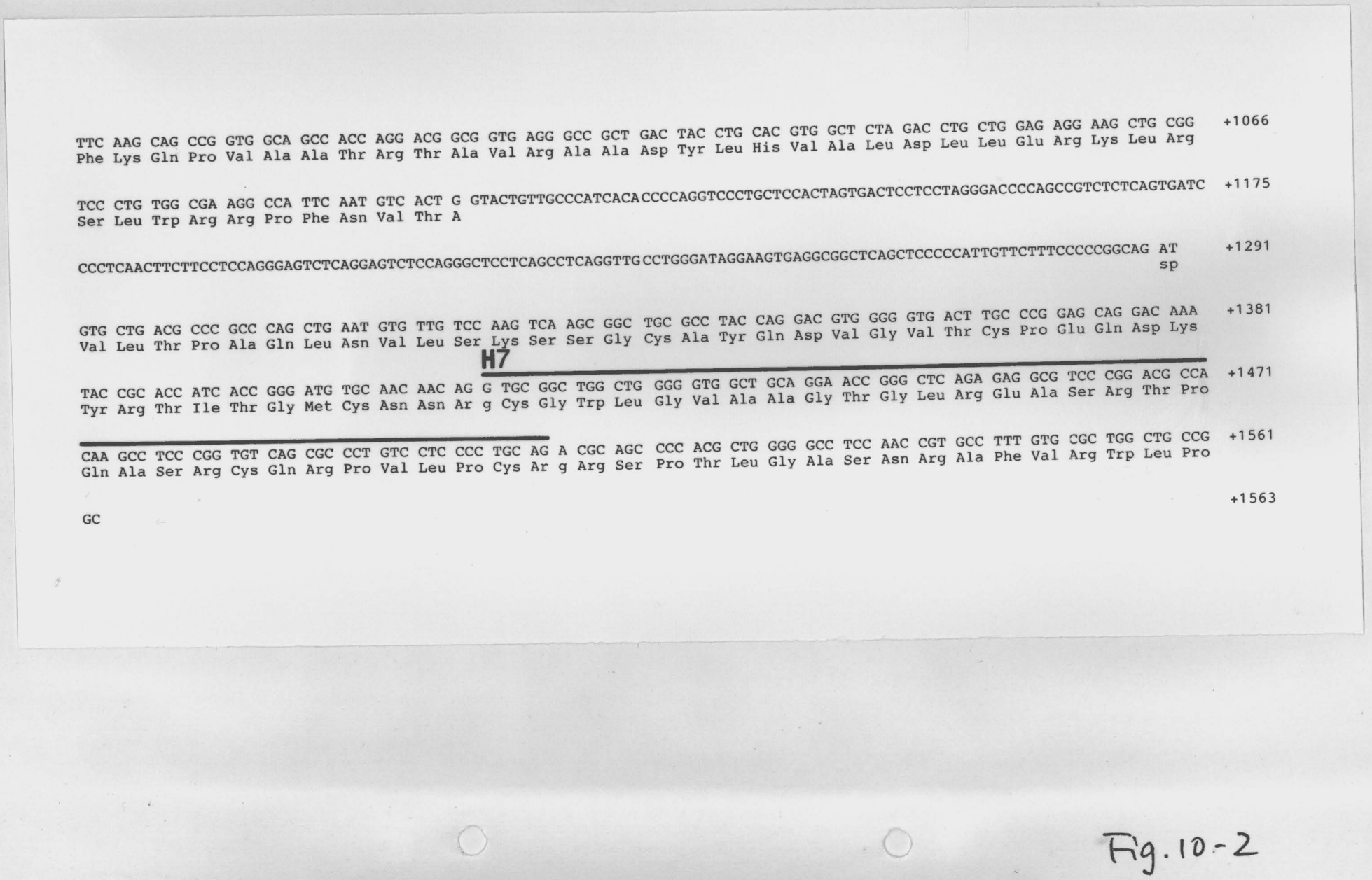

Fig. 10-2

Fig. 10-2

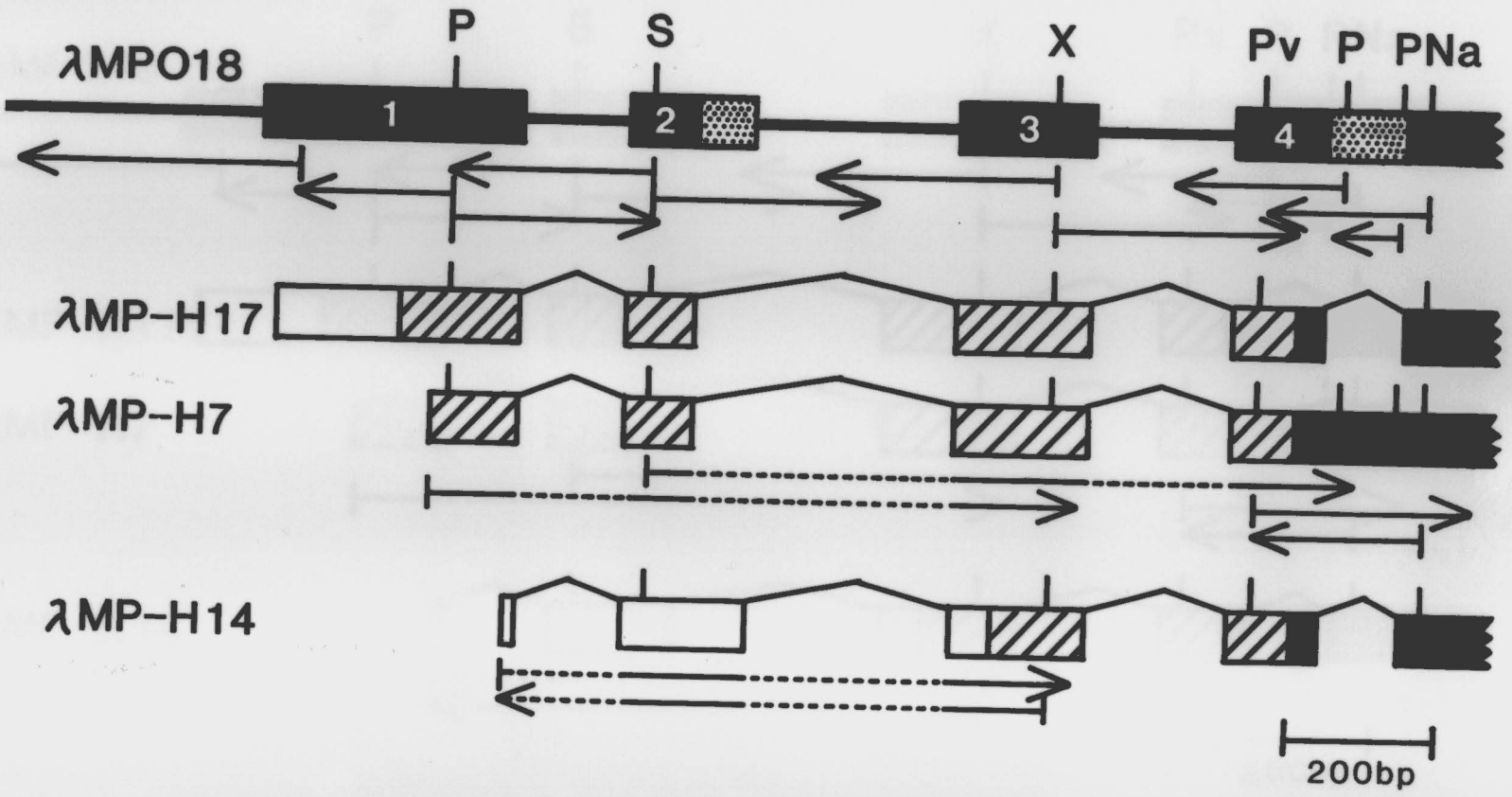

Fig. 11

Fig. 11