Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 2866–2869, May 1983
Biochemistry

Molecular cloning of human gastrin cDNA: Evidence for evolution of gastrin by gene duplication

(pancreatic gastrinoma/oligodeoxynucleotide priming/DNA sequence analysis/sequence homology)

ESPER BOEL*†, JENS VUUST*, FANNY NORRIS†, KJELD NORRIS†, ANETTE WIND*, JENS F. REHFELD‡, AND KJELD A. MARCKER*

*Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Aarhus C, Denmark; †Laboratory of Genetics, NOVO Research Institute, DK-2880 Bagsvaerd, Denmark; and ‡University Department of Clinical Chemistry, Rigshospitalet, DK-2100 Copenhagen Ø, Denmark

Communicated by Diter von Wettstein, February 7, 1983

ABSTRACT **An oligo(dT)-primed cDNA copy of the mRNA coding for the human gastrin precursor was constructed from poly(A)-containing RNA from a human pancreatic, gastrin-producing tumor (a gastrinoma). The cDNA was inserted into the *Pst* I endonuclease site of plasmid pBR322 by the use of the poly(dC) and poly(dG) tailing procedure. Clones containing gastrin sequences were selected by hybridization to a purified single-stranded ^{32}P-labeled gastrin cDNA probe. This probe was constructed with gastrinoma mRNA as template. As primer for the cDNA synthesis, we used a synthetic oligonucleotide mixture, d($\frac{A}{G}$-A-A-$\frac{A}{G}$-T-C-C-A-T-C-C-A), corresponding to the gastrin-specific amino acid sequence Trp-Met-Asp-Phe. In this way we determined the nucleotide sequence of the entire coding region (303 nucleotides), the entire 3′ untranslated region (102 nucleotides), and 8 nucleotides of the 5′ untranslated region. A striking homology between parts of the coding region suggests that evolution of the gastrin gene has involved a gene duplication.**

The polypeptide hormone gastrin is the major regulator of gastric acid secretion and growth of the gastrointestinal mucosa. Both effects are amply illustrated in patients with gastrin-producing tumors (gastrinomas), who suffer from multiple duodenal ulcers due to hypersecretion of acid from a hyperplastic mucosa.

Gastrin was originally isolated and sequenced from porcine and human antral mucosa in the principal hormonal form, the heptadecapeptide amide (1, 2). Later, longer and shorter forms with and without sulfated tyrosine residues have been discovered (3–6). Available, although indirect, evidence suggests that these forms reflect different steps in the posttranslational processing (7, 8). Recently, gastrins have also been found in extraantral endocrine cells and neurons, in which, however, the processing differs from that of the antral cells (9–12).

In order to elucidate how the synthesis of gastrin is regulated in the different normal and transformed cells, it is necessary to analyze the chromosomal gastrin gene(s). A decisive step towards this goal is the construction and cloning of gastrin cDNA. The cDNA complementary to the *porcine* gastrin mRNA has recently been cloned (13). We report now the cloning of *human* cDNA derived by transcription of mRNA isolated from a gastrinoma. A striking homology appeared between the mRNA coding for amino acids 29–54 and amino acids 62–87 of the precursor, indicating that duplication may have occurred during evolution of the gastrin gene.

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

MATERIALS AND METHODS

Enzymes and Reagents. Polynucleotide kinase (ATP:5′-dephosphopolynucleotide 5′-phosphotransferase, EC 2.7.1.78; from *Escherichia coli* strain B infected with phage T4), DNA polymerase I "Klenow fragment" (EC 2.7.7.7; from *E. coli*), and $(dT)_{12-18}$ were purchased from P-L Biochemicals. Reverse transcriptase (RNA-dependent DNA polymerase, EC 2.7.7.49; from avian myeloblastosis virus) was purchased from J. W. Beard (Life Sciences). Terminal deoxynucleotidyl transferase (EC 2.7.7.31) was from Bethesda Research Laboratories. Restriction endonucleases were from New England BioLabs. RNasin (human placenta ribonuclease inhibitor) was from Biotec (Madison, WI). Nuclease S1 (EC 3.1.30.1; from *Aspergillus oryzae*) was purchased from Sigma. Oligo(dT)-cellulose was type T-2 from Collaborative Research (Waltham, MA). DEAE-cellulose was DE52 from Whatman. [γ-^{32}P]ATP (7,500 Ci/mmol; 1 Ci = 3.7×10^{10} Bq) and [α-^{32}P]-dCTP (3,200 Ci/mmol) were obtained from New England Nuclear. [α-^{32}P]dATP (2,600 Ci/mmol) was from ICN.

Gastrinoma Tissue. Tissue was obtained from a patient with a classical Zollinger–Ellison (= gastrinoma) syndrome. The diagnosis was confirmed and the tumor was localized to the tail of the pancreas by using selective catheterization preoperatively and a specific gastrin radioimmunoassay (5).

Preparation of RNA. The tumor tissue was frozen in liquid nitrogen immediately after resection, and 0.7 g of tissue was subsequently pulverized under liquid nitrogen in a mortar, and suspended in 20 ml of a buffer containing 5 M guanidinium rhodanide, 50 mM Tris·HCl at pH 7.5, 10 mM EDTA, 5% (vol/vol) 2-mercaptoethanol, 4% (wt/vol) *N*-lauroylsarcosine, and 15% (wt/vol) CsCl. After homogenization with several strokes in a loose-fitting Dounce homogenizer, insoluble material was pelleted at 10,000 × *g* and 4°C for 20 min. Total RNA was isolated from supernatant fraction as described (14) by pelleting through a 5.7 M CsCl cushion. Poly(A)-containing RNA was obtained by chromatography on oligo(dT)-cellulose (15).

Oligodeoxyribonucleotide Synthesis. A mixture of four dodecamers, d($\frac{A}{G}$-A-A-$\frac{A}{G}$-T-C-C-A-T-C-C-A), one of which is complementary to gastrin mRNA in the region coding for the COOH-terminal tetrapeptide, was synthesized by the triester method on a 1% crosslinked polystyrene support (16). The polymer was packed in a short column, and solvents and reagents were delivered semi-automatically by means of a HPLC pump and a control module.

Labeling of Oligodeoxyribonucleotides. The oligonucleotide was labeled at the 5′ end with T4 polynucleotide kinase in a 30-μl reaction mixture containing 60 mM Tris·HCl at pH 7.5, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 250 μCi of [γ-^{32}P]ATP (30 pmol), 30 pmol of oligonucleotide, and 5 units of kinase. The reaction was carried out at 37°C for 40 min, and the kinase

was inactivated by heating to 70°C for 3 min. The specific activity of the labeled oligonucleotide was estimated by chromatography of an aliquot on a small pipette-tip DEAE-cellulose column to be 4.3×10^6 cpm/pmol.

Oligodeoxynucleotide-Primed cDNA Synthesis. In a typical reaction 10 μg of gastrinoma RNA and 60 pmol of oligodeoxynucleotide primer in 10 mM EDTA (pH 7) and a total volume of 20 μl was incubated at 70°C for 3 min. The solution was made 150 mM in KCl and placed in ice for 30 min. Then the volume was adjusted to 50 μl with additions to obtain the following final concentrations: 50 mM Tris·HCl (pH 8.3), 140 mM KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM each of the four deoxynucleoside triphosphates, RNasin at 1,000 units/ml, and reverse transcriptase at 1,000 units/ml. The mixture was incubated at 37°C for 1 hr. When general labeling of the cDNA was desired for hybridization purpose, 50–100 μCi of [α-^{32}P]dATP was included in the reaction mixture; a high specific activity was achieved by using initially a 5 μM dATP concentration and raising it to 500 μM after 15 min of incubation at 37°C. When end-labeling of the cDNA was required for sequence analysis, end-labeled oligodeoxynucleotide primer was used and only nonradioactive dNTPs were added. The synthesized cDNA was precipitated once with ethanol and dried under reduced pressure, dissolved in 8 M urea, and heated to 70°C for 3 min before direct loading on a 20 × 20 cm, 1 mm thick 5% polyacrylamide/7 M urea gel.

Electrophoresis was at 23 W for 3 hr and then the gel was autoradiographed, using Agfa–Gevaert Curix RP2 x-ray films for 30 min with a DuPont Cronex Lightning Plus AH intensifier screen at −70°C. The dominant cDNA band was electroeluted, extracted once with phenol and once with $CHCl_3$, and precipitated with ethanol for further use, either in DNA sequence analysis or as a hybridization probe.

Construction and Cloning of Double-Stranded cDNA. Oligo(dT)-primed cDNA was synthesized by using reverse transcriptase as described by Buell *et al.* (17). After hydrolysis of RNA with 0.3 M NaOH, the cDNA was extracted with phenol and purified by chromatography on Sephadex SP C-50 (18). The second strand was copied from the first strand in reaction mixtures containing about 5 μg of cDNA, 100 mM Hepes/KOH at pH 6.9, each dNTP at 500 μM, 70 mM KCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol, and 6 units of the Klenow fragment of DNA polymerase I. Incubation was for 2 hr at 14°C. After phenol extraction and chromatography on Sephadex SP C-50, the double-stranded cDNA was made blunt-ended by digestion with 100 units of S1 nuclease in the presence of 3 mM $ZnCl_2$. Homopolymeric tailing with dCTP, about 20 residues at each 3′ end, was carried out by using terminal transferase as described by Roychoudhury *et al.* (19). Similarly, pBR322 linearized with *Pst* I was tailed with dGTP. Annealing of tailed cDNA and plasmid was performed as described by Peacock *et al.* (20). Transformation of *E. coli* K803 (21) was according to the procedure of Dagert and Ehrlich (22).

Colony Hybridizations. Bacterial colonies containing recombinant plasmids were transferred to Millipore filters, lysed, and immobilized according to Grunstein and Hogness (23). Hybridization with ^{32}P-labeled single-stranded cDNA was carried out at 68°C overnight in 0.03 M sodium citrate (pH 7)/0.3 M NaCl with 0.2% each of bovine serum albumin, Ficoll, and polyvinylpyrrolidone, sonicated denatured salmon sperm DNA at 50 μg/ml, and 0.1% $NaDodSO_4$. The filters were washed in 0.015 M sodium citrate (pH 7)/0.15 M NaCl, containing 0.1% $NaDodSO_4$, at 68°C.

Recombinant Plasmid DNA Isolation. DNA was isolated by the method of Birnboim and Doly (24).

DNA Sequence Analysis. The sequences of 5′-end-labeled cDNA or 3′-end-labeled cloned DNA restriction fragments labeled with deoxynucleoside [α-^{32}P]triphosphates and DNA polymerase I were determined by the chemical cleavage procedure as described by Maxam and Gilbert (25).

RESULTS

Identification and Cloning of Human Gastrin cDNA. Poly(A)-containing RNA from the human gastrinoma was transcribed with reverse transcriptase, primed by the 5′-labeled dodecamer oligodeoxynucleotide mixture d($\overset{A}{G}$-A-A-$\overset{A}{G}$-T-C-C-A-T-C-C-A). The sequence of one of these oligonucleotides is complementary to the gastrin mRNA in the region coding for the COOH-terminal tetrapeptide Trp-Met-Asp-Phe. Analysis of the resulting cDNA on a denaturing 5% acrylamide gel revealed a major cDNA component with a length of about 335 bases (Fig. 1, lane A). The single-stranded cDNA in this band was electroeluted and its sequence was determined by the Maxam and Gilbert procedure (25). Its sequence was in complete agreement with that of human gastrin from the priming site towards the NH_2 terminus. Clones having cDNA inserts corresponding to the entire coding sequence of the gastrin precursor mRNA were prepared as described in *Materials and Methods*. The cloning efficiency was about 10^4 ampicillin-sensitive recombinants per pmol of double-stranded cDNA. Colony hybridization of 590 recombinants was carried out, using a ^{32}P-labeled single-stranded gastrin cDNA probe, eluted from a gel such as the one shown in Fig. 1, lane B. About 5% (28 colonies) gave positive signals, and 6 of these were selected for further analysis.

Sequence Analysis Strategy. For determining the sequences of the cloned cDNAs, we used the chemical cleavage procedure of Maxam and Gilbert (25). The sequences of inserts in

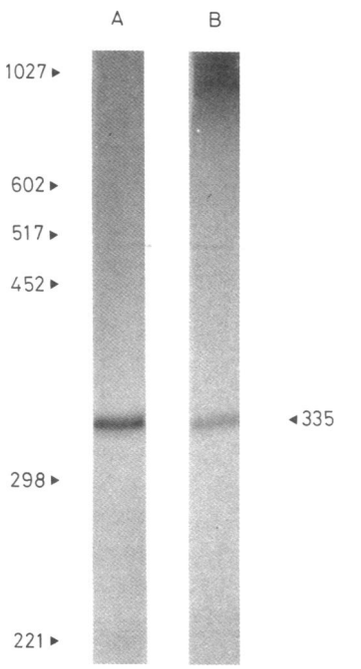

FIG. 1. Autoradiographic representations of 5% polyacrylamide/urea gels of d($\overset{A}{G}$-A-A-$\overset{A}{G}$-T-C-C-A-T-C-C-A)-primed gastrin cDNA. Lane A, cDNA transcribed from 5 μg of gastrinoma mRNA, using 5′-end-labeled primer and unlabeled dNTPs. Lane B, cDNA transcribed from 2 μg of gastrinoma mRNA, using unlabeled primer and [α-^{32}P]dATP as label. Plasmid pBR327 was digested with endonucleases *Hin*dIII and *Hin*fI and run in parallel; the nucleotide lengths and positions of these markers are indicated on the left.

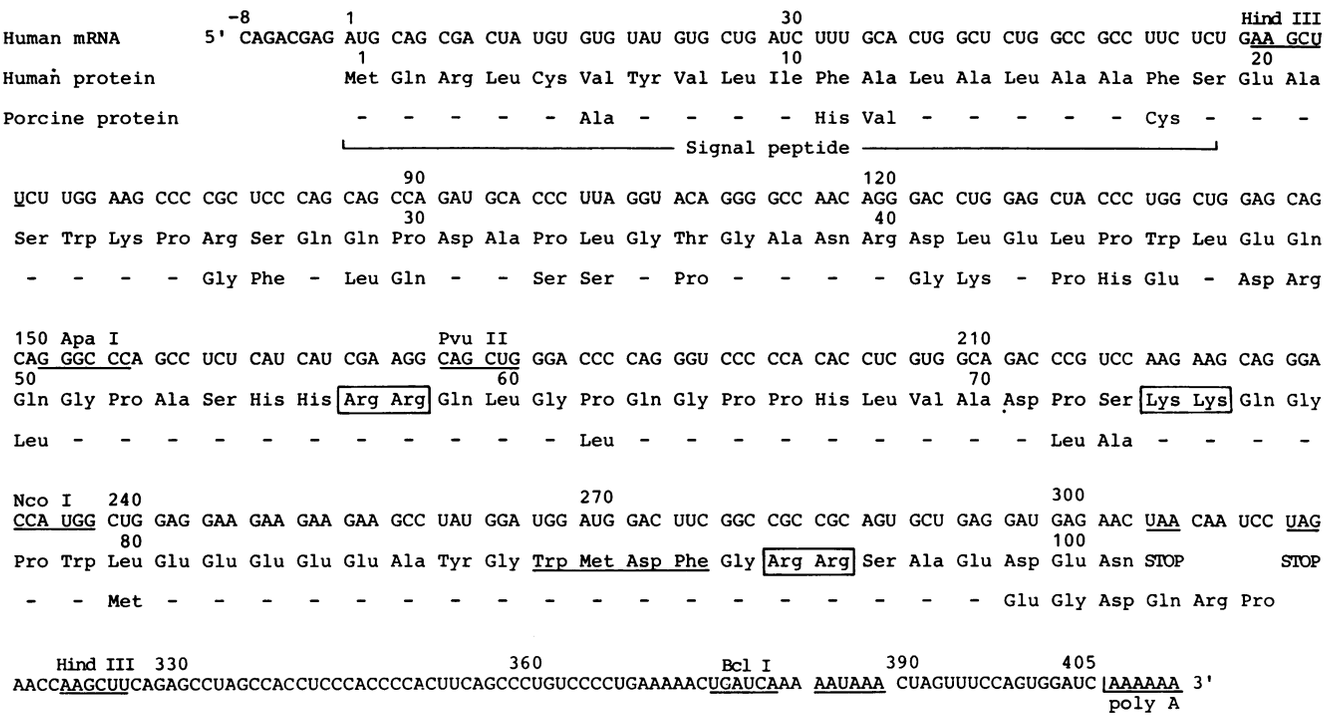

FIG. 2. Sequence of the human gastrin precursor mRNA, deduced from the sequence of the cDNA insert in clone pHG529. The positions are numbered from the A of the initiating methionine codon to the last nucleotide (405) before the start of the poly(A). Negative numbers are used in the 5′ untranslated region. Major restriction enzyme recognition sites are shown and underlined. The two stop codons in the correct reading frame and the polyadenylylation signal are underlined. The amino acid sequence deduced from the mRNA is shown, along with that of the porcine gastrin precursor. Amino acid identity is indicated by hyphens in the porcine sequence (13). Amino acids are numbered from the initiating methionine. In the protein sequence, the presumed signal peptide as well as the COOH-terminal tetrapeptide of gastrin are underlined. Possible dibasic processing sites for trypsin-like proteases are boxed.

the recombinant plasmids that showed hybridization to the gastrin cDNA probe were determined from the unique *Nco* I site and from the two *Hin*dIII sites found in the cDNA (Fig. 2). In this way the sequences of four different clones were determined to rule out errors from either cDNA synthesis or misreading of sequencing gels. One clone, pHG529, was found to carry a 463-base-pair insert, excluding the poly(dC) tails. It included 8 nucleotides of the 5′ untranslated region, the entire coding region (303 nucleotides), a 3′ untranslated region of 102 nucleotides, and a poly(A) stretch of 50 residues.

DISCUSSION

The present study demonstrates, as also found by others (26), that a specific oligodeoxynucleotide primer is a powerful tool for detection and transcription of a specific mRNA species. The results of the cDNA cloning experiments indicate that, in the gastrinoma tissue used, approximately 5% of the mRNA codes for gastrin.

The coding region of the human gastrin mRNA is 303 nucleotides long. The primary translation product consists of 101 amino acid residues and has a molecular weight of 11,381. The derived amino acid sequence for amino acids 56–92 is in agreement with the longest known amino acid sequence of human gastrin (27). Two pairs of basic residues, residues 57–58 (Arg-Arg) and 74–75 (Lys-Lys) represent processing sites for trypsin-like enzymes. The cleavage at these sites will yield two of the known forms of the hormone, gastrin-34 (amino acids 59–92) and gastrin-17 (amino acids 76–92). The coding region also contains a putative leader, or signal, sequence (28) of 19 amino acid residues in addition to a further sequence of 37 amino acids between the leader sequence and the first Arg-Arg processing site. The Gly-Arg-Arg sequence (residues 93–95) is probably important for the enzymatic amidation of the COOH-terminal phenylalanine of gastrin (13). The 3′ untranslated region of the gastrin mRNA contains a putative poly(A) addition signal (A-A-U-A-A-A) (29), 17 nucleotides from the poly(A) tract.

Comparison of the human gastrin precursor with the corresponding porcine sequence reveals 26 amino acid replacements. Moreover, the COOH terminus of the porcine polypeptide contains an additional Gln-Arg-Pro sequence. Four replacements of amino acids (out of 34) have occurred in the gastrin-34 region, which is therefore clearly the most conserved part of the molecule. Four amino acid replacements (out of 19) have occurred in the putative signal peptide. However, 15 replacements have taken place among the 37 amino acids of the region between the signal peptide and gastrin-34. This particular region is therefore the least conserved part of the polypeptide, although its length is the same in the human and porcine gastrin precursor. This region may possibly play some spacer role in transport or processing of gastrin. Moreover, the COOH-terminal part of the 37 amino acids is included in the largest circulating form of human gastrin, component I (4).

The human gastrin precursor contains the sequence Pro-Trp-Leu-Glu at positions 45–48 and 78–81. When these two sequences are aligned (Fig. 3), a close homology is observed between the regions corresponding to amino acids 41–54 and 74–87, respectively. In fact, 60% of the nucleotides in these sequences are identical. A less pronounced homology occurs between the regions corresponding to amino acids 29–33 and 62–66. In this case, there is 47% nucleotide sequence identity. Similar, although less obvious, homologies may be found in the nucleotide sequence of the porcine gastrin mRNA, reported by Yoo *et al.* (13). These observations suggest that the gastrin gene could have evolved through a gene duplication from an ances-

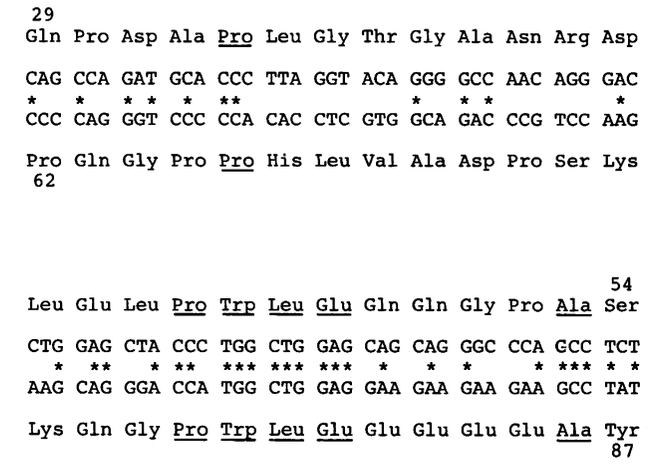

FIG. 3. Comparison of the regions of the human gastrin mRNA coding for amino acids 29–54 and amino acids 62–87. Identical nucleotides are indicated by asterisks, and identical amino acids are underlined.

tral DNA sequence coding for a stretch of approximately 38 amino acids.

Recently, it was observed (30) that gastrin-34 is homologous to the transforming protein ("middle T") of polyoma virus. We have compared the nucleotide sequence coding for gastrin-34 with that of polyoma middle T protein containing a modified tyrosine residue (31). This comparison reveals a 48% nucleotide sequence identity over the entire stretch of 34 amino acids, in good agreement with the observation of Baldwin (30). Because the two proteins seem to share the ability to stimulate cellular growth, and because it has been suggested that the homology may have arisen by incorporation of animal DNA sequences into simian virus 40-like genome, further studies on the significance of the homology between the gastrin and middle T DNA might be of great value.

We thank Lars Berglund, Iben Hjort, Marianne Nielsen, and Ole Nyman for skillful technical assistance. We also thank Linda Mygil for excellent secretarial assistance. This work was supported by grants from NOVO Industri A/S, from the Olga and Esper Boel Foundation, and from the Danish Medical and Science Research Councils.

1. Gregory, R. A. & Tracy, H. J. (1964) *Gut* **5**, 103–117.
2. Bentley, P. H., Kenner, G. W. & Sheppard, R. C. (1966) *Nature (London)* **209**, 583–585.
3. Yalow, R. S. & Berson, S. A. (1970) *Gastroenterology* **58**, 609–615.
4. Rehfeld, J. F. (1972) *Biochim. Biophys. Acta* **285**, 364–372.
5. Rehfeld, J. F. & Stadil, F. (1973) *Gut* **14**, 369–373.
6. Rehfeld, J. F. & Larsson, L.-I. (1979) *Acta Physiol. Scand.* **105**, 117–119.
7. Rehfeld, J. F. & Uvnäs-Wallensten, K. (1978) *J. Physiol. (London)* **283**, 379–396.
8. Dockray, G. J., Vaillant, C. & Hopkins, C. R. (1978) *Nature (London)* **273**, 770–772.
9. Uvnäs-Wallensten, K., Rehfeld, J. F., Larsson, L.-I. & Uvnäs, B. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5705–5710.
10. Rehfeld, J. F. (1978) *Nature (London)* **271**, 771–773.
11. Larsson, L.-I. & Rehfeld, J. F. (1979) *Nature (London)* **277**, 575–578.
12. Rehfeld, J. F. & Larsson, L.-I. (1981) *J. Biol. Chem.* **256**, 10426–10429.
13. Yoo, O. J., Powell, C. T. & Agarwal, K. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1049–1053.
14. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
15. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
16. Ito, H., Ike, Y., Ikata, S. & Itakura, K. (1982) *Nucleic Acids Res.* **10**, 1755–1769.
17. Buell, G. N., Wickens, M. P., Payvar, F. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2471–2482.
18. Friedman, E. Y. & Rosbash, M. (1977) *Nucleic Acids Res.* **4**, 3455–3471.
19. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acids Res.* **3**, 101–116.
20. Peacock, S. L., McIver, C. M. & Monahan, J. J. (1981) *Biochim. Biophys. Acta* **655**, 243–250.
21. Wood, W. B. (1966) *J. Mol. Biol.* **16**, 118–133.
22. Dagert, M. & Ehrlich, S. D. (1979) *Gene* **6**, 23–28.
23. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
24. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
25. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
26. Noyes, B. E., Mevarech, M., Stein, R. & Agarwal, K. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1770–1774.
27. Gregory, R. A. & Tracy, H. J. (1975) in *Gastrointestinal Hormones*, ed. Thompson, J. C. (Univ. of Texas Press, Austin, TX), pp. 13–24.
28. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851.
29. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
30. Baldwin, G. S. (1982) *FEBS Lett.* **137**, 1–5.
31. Soeda, E., Arrand, J. F., Smolar, N., Walsh, J. E. & Griffin, B. E. (1980) *Nature (London)* **283**, 445–453.