Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1931–1935, April 1985
Biochemistry

Molecular cloning of the human cholecystokinin gene by use of a synthetic probe containing deoxyinosine

(hormone gene/DNA sequence analysis/intervening sequence/brain/gut peptide)

YOOSUKE TAKAHASHI*, KIKUYA KATO*, YOSHIHIDE HAYASHIZAKI*, TOSHIAKI WAKABAYASHI*, EIKO OHTSUKA†, SHIGERU MATSUKI†, MORIO IKEHARA†, AND KENICHI MATSUBARA*

*Institute of Molecular and Cellular Biology, and †Faculty of Pharmaceutical Sciences, Osaka University, Yamadaoka, Suita, 1-3, Japan

Communicated by Sydney Brenner, November 21, 1984

ABSTRACT **A synthetic DNA based on the known amino acid sequence of the brain/gut peptide cholecystokinin (CCK) was synthesized. This DNA contained deoxyinosines at ambiguous codon positions and was used as a probe to isolate the CCK gene directly from a human genomic library. Nucleotide sequence analysis of the isolated gene revealed that human preprocholecystokinin consists of 115 amino acid residues, with 11 amino acids in common with the human gastrin precursor, another member of the gastrin-CCK family, and that the coding region is separated by a single, long intron. CCK appears to be encoded by a single-copy gene in the haploid human genome, as revealed by genomic Southern hybridization analysis, suggesting that the same gene is expressed both in gut and brain.**

Cholecystokinin (CCK) is a brain/gut peptide that has been detected in the small intestine (1) as well as in the brain (2). In the gut, CCK induces gallbladder contraction and the release of pancreatic enzymes (3). In the brain, its role is much less clear, but the high content of this organ is taken as a sign that CCK has an important function(s) in the nervous system.

Various forms of CCK that differ in amino acid chain length, but all of which are biologically active, have been reported. Among them, CCK octapeptide (CCK8) is believed to be the major functional unit of this peptide. These peptides apparently are formed by processing, mostly at amino termini, from a single precursor polypeptide. Their carboxyl termini are all the same, however, ending with an amidated phenylalanine. The multiple molecular forms and functions of brain and gut CCK raise the question whether they stem from a single or multiple genes. If the same gene functions, then another question emerges as to how the different peptides are produced in different tissues.

Recently, the amino acid sequence of porcine CCK58 and the cDNA sequence for rat CCK have been reported (4, 5). Considering the large accumulation of data on human CCK (its suspected biological role in the brain, its possible connection with diseases, etc.), it is very important to determine the amino acid sequence of human CCK and its precursor. The carboxyl-terminal five amino acid residues of CCK are identical with those of gastrin (6), another brain/gut peptide, which induces the release of gastric juices and which is detected also in the brain, although its function there is not known. These two peptide hormones, in addition to caerulein found in frog skin (7), comprise the gastrin-CCK family. The related structures of these peptide hormones are also of interest in terms of molecular evolution. These concerns prompted us to isolate and study the human CCK gene.

Lack of a CCK-producing tumor cell line precluded our obtaining human CCK mRNA for analysis, because its concentrations in the available human tissues are extremely low. We therefore isolated the gene directly from a genomic library by using an appropriate oligonucleotide probe. For cDNA cloning (8), mixed-sequence oligonucleotide probes of 14–17 nucleotides often have been used. These probes typically consist of mixtures of 8–32 oligodeoxynucleotides, in order to represent every possible codon combination for the stretch of amino acid sequence in question. For the direct cloning of a single-copy gene in such a complex population as a human genomic library, however, a much longer probe is needed (9). One difficulty is that mixed probes cannot be used effectively because they necessarily contain too many varieties of molecular species due to many codon degeneracies. We overcame this problem by using an appropriate base analog that can "pair" with any of the four natural bases at the ambiguous positions, with or without forming hydrogen bonds. Probes with inosines at the ambiguous positions in the degenerate codons were useful for this purpose. Using this novel probe, we isolated the human CCK gene directly from the genomic library and determined its primary structure.

MATERIALS AND METHODS

Oligonucleotide Probe. The deoxyinosine-containing oligodeoxynucleotide probe was synthesized by the triester method on a polymer support, as described in detail elsewhere (10). A mobility-shift analysis of the oligonucleotide was performed after partial digestion with snake venom phosphodiesterase and nuclease P1 as described (10).

Screening of the Human Genomic Library. The genomic library was provided by R. M. Lawn (11). Plaque transfer to nitrocellulose filters was performed as described by Benton and Davis (12). The filters prepared in this way were prehybridized for 2 hr at 65°C in 6 × NET (13)/1× Denhardt's solution (1× NET is 0.15 M NaCl/1 mM EDTA/0.03 M Tris Cl, pH 8.0. 1× Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrollidone/0.02% bovine serum albumin.) Hybridization with the 5′-^{32}P end-labeled oligonucleotide (5–10 × 10^5 cpm/ml) was at 55°C for 16–18 hr in 6× NET/1× Denhardt's solution. The specific activity of the probe was $\geq 10^6$ cpm/pmol. The filters were washed four times at room temperature in 6× NaCl/Cit (1× is 0.15 M NaCl/15 mM sodium citrate, pH 7.0) and then once for 1 min at 55°C in 6× NaCl/Cit. Filters were exposed at −70°C for 50 hr to Kodak XAR-5 x-ray film with a Lightning Plus intensifying screen.

Sequence Analysis. DNA fragments were subcloned in M13 vector mp8 and mp9 or in mp10 and mp11. The nucleotide sequences were determined by the dideoxy method (14).

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

Abbreviations: kb, kilobase(s); CCK, cholecystokinin (the numeral suffix refers to the number of amino acid residues—e.g., CCK8 is the cholecystokinin octapeptide).

```
CCK8
Amino Acids     Asp Tyr Met Gly Trp Met Asp Phe NH2
                                              ?
mRNA            GAU UAU AUG GGA UGG AUG GAU UUU GGN
                  C   C       C           C   C
                            G
                            U

I26mer Probe    CTI ATI TAC CCI ACC TAC CTI AAI CC
```

FIG. 1. A synthetic oligodeoxyribonucleotide used as a probe for cloning the human CCK gene. The mRNA nucleotide sequence was deduced from the known amino acid sequence of porcine CCK8. The probe contained deoxyinosine (I) at positions at which the nucleotide could not be assigned due to codon degeneracy. For other arguments used in arriving at this sequence, see the text. I residues are underlined.

Enzymes. Restriction endonucleases, T4 polynucleotide kinase and DNA polymerase I (Klenow fragment) were purchased from Takara Shuzo (Kyoto, Japan). T4 ligase was a gift from T. Tsurimoto.

RESULTS

The Probe. The 26mer oligonucleotide probe designed for cloning the human CCK gene has the sequence depicted in Fig. 1, which was deduced from the amino acid sequence of porcine CCK8 (Asp-Tyr-Met-Gly-Trp-Met-Asp-Phe-NH_2). Because CCK8 is the functional unit of CCK, we assumed that this sequence is conserved. The following points were taken into account in designing the probe: First, we chose deoxyinosine when the nucleotide could not be assigned because of codon degeneracy. Preliminary experiments showed that deoxyinosine behaved as an "inert" base that neither destabilized nor contributed much to the DNA duplex structure (10). Second, carboxyl-terminal amidation of CCK8 was taken to mean that the juxtaposed amino acid on its carboxyl side is glycine, as is common for other prohormones with amidated carboxyl termini (15). Third, we assumed that no intron is present in the sequence that encodes CCK8, the functional unit of CCK, because functional domains of known peptide hormones are almost always covered by a single exon.

Screening of the Human Genomic Library. To determine the appropriate temperature for DNA·DNA hybridization when a probe containing deoxyinosine is used, synthetic DNA was cloned in pBR322 (10), and the resulting plasmid DNA was examined for its dissociation temperature in blot–hybridization at various temperatures with the synthetic DNA as the radioactive probe. The dissociation temperature was 55–60°C (10). Thus, we took 55°C as the temperature for plaque hybridizations with the 5′-^{32}P-labeled probe. The human genomic library is λ Charon 4A carrying human fetal liver DNA fragments from partial digestion with *Hae* III and *Alu* I (11). From about 3×10^5 plaques, we isolated four positive clones. Restriction enzyme analyses revealed that three of the four clones carried identical 12-kilobase (kb) inserts and that the fourth carried an overlapping fragment. One clone, λCK58, which belongs to the first group (12-kb insert) was chosen for further analyses.

Restriction Mapping and Sequencing Analysis. The restriction map of the insert in λCK58 is shown in Fig. 2. Nucleotide sequence analysis showed that λCK58, as expected, carried the human CCK gene as judged from the exact correspondence of its sequence with the amino acid sequences of porcine (6) and rat CCK (7). The correspondence extends from the carboxyl terminus to the region upstream of the second amino acid of porcine CCK33 (Fig. 3), where it ends abruptly, suggesting that an intron starts in the human gene sequence at this position. The typical splicing acceptor sequence (16) is also found at this position.

We then synthesized a mixed oligonucleotide whose sequence we based on the amino terminal pentapeptide sequence of porcine CCK39, Tyr-Ile-Gln-Gln-Ala. This probe consisted of 24 tetradecanucleotides representing every possible codon combination, A-T-(A/G)-T-A-(A/G/T)-G-T-(T/C)-G-T-(T/C)-C-G. It was 5′-end-labeled and then was used to detect the upstream exon; again we assumed that human and porcine CCK share this amino acid sequence. Southern hybridization of variously cleaved λCK58 with this probe showed that a 4.0-kb *Eco*RI fragment (Fig. 2) was hybridized, evidence that the upstream exon is located in this fragment. The two exons are separated by ≈6.2 kb. The 4.0-kb *Eco*RI fragment was subcloned in pBR322 for use in further studies.

The relevant nucleotide sequence is shown in Fig. 3. The location of the initiator methionine codon is proposed for the following reasons: (*i*) it is immediately followed by a putative signal-peptide sequence encoding hydrophobic amino acids; (*ii*) its downstream sequence, but not its upstream sequence, matches very well with the sequence of the rat CCK-coding region; and (*iii*) further upstream, a methionine codon does not appear before a termination codon (UAA) appears. The splice junction A-A-A-G-[G-T- ..... -T-C-C-T-T-G-T-A-G] was assigned because it conforms to the consensus sequence A-A-G-[G-T ..... -Y-Y-Y-T-T-N-C-A-G] and because only within the bracketed region do the human genomic and rat cDNA sequences fail to show homology.

Our deduced human preprocholecystokinin consists of 115 amino acid residues, the same number as recently determined for rat preprocholecystokinin (7). Starting with methionine (residue 1), the first 20 amino acid residues are hydrophobic; such a feature is characteristic of signal peptides

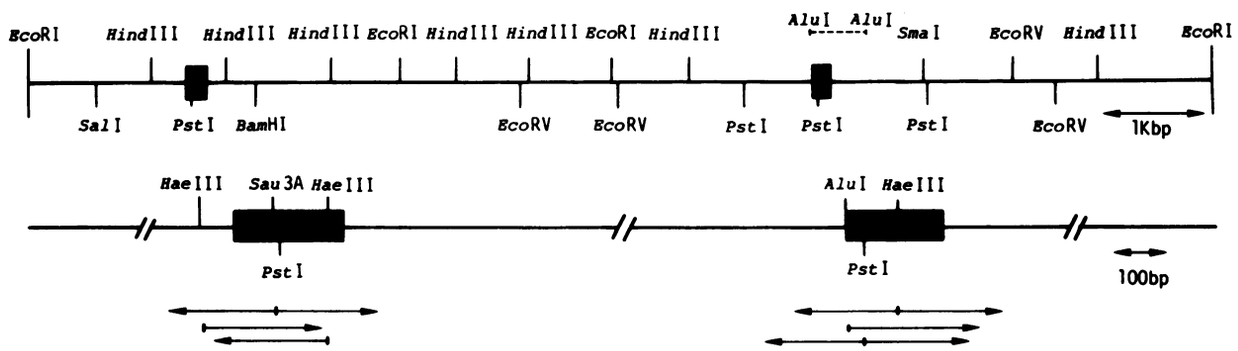

FIG. 2. Restriction map of λCK58, which carries the human CCK gene. Only relevant restriction sites are displayed. Relative positions of exons, deduced from nucleotide sequence data, are indicated by boxes. The directions and extents of the sequence determinations are shown by horizontal arrows in the lower, expanded map. The dashed line above represents the region (*Alu* I fragment) used as a probe for genomic Southern hybridization (Fig. 4). bp, Base pairs.

```
AGCCAGGTGCCCGGGGCGCGCGCCATGGCACTCGGCTGGGTCAGCGCTTGGCGAGCCCTCGTGCTTGCGGCGGCCTGGCGGGTCTGGAGCACCTGTGTTTGTCCTAAGCCCGCTTCTGG

                                                                   1                                   10
GTGCCCGGCCCGCCGCAGGAGCTCTGTTGCCCAGCCTTTCAGTTAACGCGCCCTCCTCCTTTGCTCCCTACAGCC ATG AAC AGC GGC GTG TGC CTG TGC GTG CTG ATG
                                                                           Met Asn Ser Gly Val Cys Leu Cys Val Leu Met

                            20                                      30                                      40
GCG GTA CTG GCG GCT GGC GCC CTG ACG CAG CCG GTG CCT CCC GCA GAT CCC GCG GGC TCC GGG CTG CAG CGG GCA GAG GAG GCG CCC CGT
Ala Val Leu Ala Ala Gly Ala Leu Thr Gln Pro Val Pro Pro Ala Asp Pro Ala Gly Ser Gly Leu Gln Arg Ala Glu Glu Ala Pro Arg

                            50                                      60                                      70
AGG CAG CTG AGG GTA TCG CAG AGA ACG GAT GGC GAG TCC CGA GCG CAC CTG GGC GCC CTG CTG GCA AGA TAC ATC CAG CAG GCC CGG AAA
Arg Gln Leu Arg|Val Ser Gln Arg Thr Asp Gly Glu Ser Arg Ala His Leu Gly Ala Leu Leu Ala Arg|Tyr Ile Gln Gln Ala Arg|Lys
               ↓ CCK58                                                                     ↓ CCK39                ↓ CCK33

G [GT]AAGAATGCTGCCTCCCCATCCCTCACTTCTGCCCTTGTTCCCAGGCTCCCGATGCTGACCC ------ 6.0 kb ----  TTTTTCCATTGCTCTCCTTCCCAA
A
                                                                                                       80
CAAATTGAACTAGAGATGCTTTTAGATGCAATGTCCCTGTTTCCTTCCTCCCTGCTCCTTGT[AG] CT CCT TCT GGA CGA ATG TCC ATC GTT AAG AAC CTG CAG AAC
                                                                   la Pro Ser Gly Arg Met Ser Ile Val Lys Asn Leu Gln Asn

                90                                  100                                     110
CTG GAC CCC AGC CAC AGG ATA AGT GAC CGG GAC TAC ATG GGC TGG ATG GAT TTT GGC CGT CGC AGT GCC GAG GAG TAT GAG TAC CCC TCC
Leu Asp Pro Ser His Arg|Ile Ser Asp Arg|Asp Tyr Met [Gly Trp Met Asp Phe Gly Arg Arg Ser Ala Glu] Glu Tyr Glu Tyr Pro Ser
                       ↓ CCK12         ↓ CCK8
110
TAGAGGACCCAGCCGCCATCAGCCCAACGGAAGCAACCTCCCAACCCAGAGGAGGCAGAATAAGACAACAATCACACTCATAACTCATTGTCTGTGGAGTTTGACATTGAATGTATCTA
Term

TTTATTAAGTTCTCAATGTGAAAATTGTGTCTGTAAGATTGTCCAGTG
```

FIG. 3. The nucleotide sequence of the human CCK gene. The deduced amino acid sequence of the human CCK precursor also is shown. For arguments on the identification of the splice junction and initiation codon, see the text. The numbering of the preproCCK begins at the putative initiator methionine (residue 1) and terminates at serine (residue 115). The splice donor and acceptor sites are boxed. The DNA sequences that correspond to the two probes are underlined. Cleavage sites that define different forms of CCK are indicated by arrows. The carboxyl-terminal amino acid sequence that also is found in human preprogastrin is boxed.

found at the amino-terminal end of secreted proteins. The amino termini of CCK58 and CCK39 correspond to residues 46 and 65, respectively. A long intron is present in codon 72, which encodes the second amino acid of CCK33. CCK8 covers residues 96–103. Phe-103, which is to become the amidated carboxyl terminus in mature CCK, is followed by Gly-Arg-Arg, a sequence commonly observed with related amidated peptides (15).

Genomic Southern Hybridization. To examine the copy number of the CCK gene in the human genome, genomic Southern hybridization was performed. High molecular weight human leukocyte DNA was digested with *Eco*RI or *Bgl* II. The resulting fragments were separated by agarose gel electrophoresis, followed by blotting according to Southern (17). The DNA then was hybridized with the 550-base-pair *Alu* I fragment (shown in Fig. 2) as a probe that covers the entire exon that codes for the functional domain of CCK. A sharp single band was observed for both the *Eco*RI and *Bgl* II digests (Fig. 4). In other experiments, in which *Hin*dIII or *Pvu* II digestions were used, a sharp single band again was seen (not shown). These observations strongly suggest that the CCK gene exists as a single copy in the human genome.

DISCUSSION

We have described the direct isolation of the human CCK gene, a unique (single-copy) gene in the haploid genome, from a human genomic DNA library. A 26mer oligodeoxynucleotide probe containing deoxyinosines was used to screen λ phage carrying the library. The deoxyinosines located at ambiguous codon positions enabled us to prepare a single, long oligonucleotide probe. This novel type of probe should prove useful for cloning genes directly from a genomic library when an amino acid sequence of the gene product is known, even when it is not possible to obtain mRNA. At present, however, there is no reasonable way to identify transcriptional and translational initiation and termination signals or to identify introns in the sequence obtained by this gene cloning procedure. In the case of the human CCK gene, however, available data on the amino acid sequence of porcine CCK and the nucleotide sequence of rat CCK cDNA helped solve this problem. Thus, we were able to locate

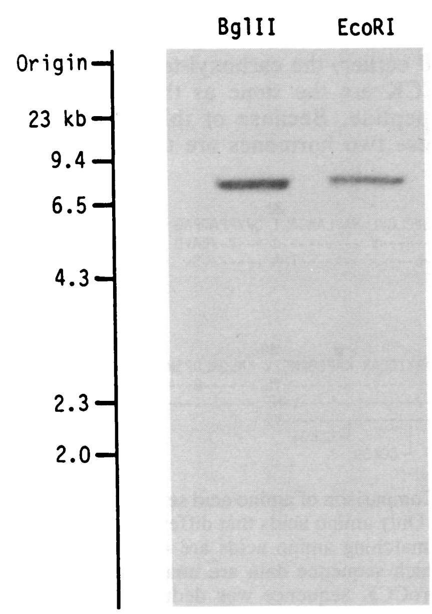

FIG. 4. Genomic Southern hybridization analysis of the human CCK gene. Human leukocyte DNA (10 μg) was digested with *Eco*RI or *Bgl* II, then electrophoresed through a 0.7% agarose gel and transferred to a nitrocellulose filter (17). Hybridization was carried out at 65°C in 6× NaCl/Cit for 16 hr, using a ^{32}P-labeled (10^8 cpm/μg) 550-bp *Alu* I fragment probe that carried a portion of the CCK gene (see Fig. 2). The filter was washed four times at 65°C in 0.1× NaCl/Cit containing 0.5% $NaDodSO_4$ and then autoradiographed for 2 days with an intensifying screen. Markers at left (in kb) indicate positions of fragments from *Hin*dIII-digested λ DNA.

splice junctions and the putative initiator codon in the human DNA sequence.

The single intron in the human CCK gene is very long (6.2 kb) and separates a 40 amino acid sequence that includes the principal functional unit (CCK8) from the rest (75 amino acids) of the polypeptide. In peptide hormone genes, such as those for pro-opiomelanocortin (18), preproenkephalin (19), and preproinsulin (20), an additional intron is common in the 5′ noncoding region. The human CCK gene also may have an intron in the 5′ noncoding region, but the unavailability of mRNA precludes our confirming this. Deschenes *et al.* (5), however, showed that the rat CCK gene has, in fact, an intron situated 2 base pairs upstream from the ATG initiation codon. This position in the human sequence carries a consensus sequence of the splicing acceptor site (16): a long pyrimidine stretch followed by the dinucleotide A-G (Fig. 3).

Available data on the CCK amino acid sequence in several species is given in Fig. 5. The deduced human CCK sequence is highly homologous to that of porcine (4), rat (5), and dog CCK (22). The carboxyl-terminal region, which carries a functional CCK8, is completely conserved. Two obviously heterogeneous regions are present: amino acids 46–52 and amino acids 79 and 80. The sequences in the human and rat show a third heterogeneous region between residues 24 and 34. The meaning of these conserved and nonconserved regions in the CCK gene is not yet clear.

Various biologically active forms of CCK that differ in their amino acid chain lengths have been reported. They include CCK58 (4), CCK39 (23), CCK33 (24), CCK12 (25), and CCK8 (26). Whether they have precursor–product relationships, are produced in different cells, or have different physiological functions has yet to be determined. Because we now have the total amino acid sequence in hand, these possibilities can be examined critically. Examination of the amino termini of these intermediate peptides shows clearly that cleavage takes place next to basic amino acid residues (15). Thus, cleavage on the carboxyl side of arginine at positions 45, 64, 70, 91, and 95 produces, respectively, CCK58, -39, -33, -12, and -8.

As stated earlier, the carboxyl-terminal five amino acids found in CCK are the same as those of gastrin, another brain/gut peptide. Because of this highly conserved sequence, these two hormones are thought to have evolved from a common ancestor gene. Comparison of the human CCK and gastrin genes (27, 28), however, revealed an apparent dissimilarity; the CCK gene has a long intron (6.2 kb) in the 72nd codon, whereas the gastrin gene has only a short intron (130 bp) at the 71st codon. The amino acid sequences of the two preprohormones have in common 11 amino acids located near the carboxyl terminus of the mature peptide, which includes the five conserved amino acids plus glycine. No obvious homologies were found in other regions. At present, the striking homology in the deduced carboxyl-terminal sequences, which is in contrast to the apparent lack of homologies in other regions, is taken to suggest that the two genes are derived from a common ancestor at least at the carboxyl-terminal region of the coding sequence. Whether man has other genes that carry the common sequence is an interesting question.

CCK appears to be encoded by a single-copy gene in the human genome as revealed by genomic Southern hybridization analysis. Provided that all the CCK peptides detected in the brain and gut, primarily by immunological means, have identical amino acid sequences other than only at the carboxyl terminus, at which the major anti-CCK activity has been directed, the same gene is expressed in both organs under different control systems.

After this work was completed, Gubler *et al.* (21) reported the cDNA sequence of porcine CCK. A comparison between porcine and human CCK shows that the preproform of the former lacks one amino acid in the heterogeneous region around amino acid 30; otherwise the two sequences are very similar.

This study was supported by grants from the Ministry of Education, Science, and Culture of Japan.

```
         1                     20                             40
Human    MNSGVCLCVL MAVLAAGALT QPVPPADPAG SGLQRAEEAP RRQLRVSQRT DGESRAH
Rat      -KC------V ---------A ---V-VEAVD PMEE------ -----AVL-P -S-P--R
Porcine  --G-L----- -------T-A -------S-  VPGAQE--AH -----AV-KV -------
Dog                                                    AQKVN S------
                                                       └ CCK58

       60              ▼     80                    100
       LGA LLARYIQQAR KAPSGRMSIV KNLQNLDPSH RISDRDYMGW MDFGRRSAEE YEYPS
       --- --------V- --------VL ------G--- ---------- ---------D -----
       --- ---------- ------V-MI ----S----- ---------- ---------D -----
       -------------------------VI
                |     └CCK33                   |   └CCK8
                └CCK39                         └CCK12
```

FIG. 5. Comparison of amino acid sequences of CCKs from various species. Only amino acids that differ from those of human CCK are shown; matching amino acids are represented by dashes. Regions for which sequence data are unavailable are left open. The human preproCCK sequence was deduced from the genomic sequence. Rat preproCCK sequence was from cDNA (5). Porcine sequence was taken from CCK58 (4) and cDNA (21). The dog sequence, which covers only residues 46–82, was determined by amino acid sequence analysis (22). Numbering begins at the putative initiator methionine (residue 1) of human CCK. The triangle shows the location of an intron in the human gene. The single-letter amino acid code used is A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine.

1. Harper, A. A. & Raper, H. S. (1943) *J. Physiol.* **102,** 115–125.
2. Dockray, G. J. (1976) *Nature (London)* **264,** 568–570.
3. Mutt, V. (1980) in *Gastrointestinal Hormones*, ed. Jerzy, G. B. (Raven, New York), pp. 169–221.
4. Eng, J., Shiina, Y., Pan, Y. C. E., Blacher, R., Chang, M., Stein, S. & Yallow, R. S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
5. Deschenes, R. J., Lorenz, L. J., Haun, R. S., Roos, B. A., Collier, K. J. & Dixon, J. E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 726–730.
6. Gregory, H., Hardy, P. M., Jones, D. S., Kenner, G. W. & Shepperd, R. C. (1964) *Nature (London)* **204,** 931–933.
7. Anastasi, A., Erspamer, V. & Endean, R. (1968) *Arch. Biochem. Biophys.* **125,** 57–68.
8. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6613–6617.
9. Anderson, S. & Kingston, I. B. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6838–6842.
10. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. & Matsubara, K., *J. Biol. Chem.*, in press.
11. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
12. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
13. Dalbadie-Mcfarland, G., Cohen, L. W., Riggs, A. D., Morin, C., Itakura, K. & Richards, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6409–6413.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Bradbury, A. F. & Smyth, D. G. (1982) *Nature (London)* **298,** 686–688.
16. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
17. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
18. Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. & Numa, S. (1981) *Eur. J. Biochem.* **115,** 429–438.
19. Noda, M., Teranishi, Y., Takahashi, H., Toyosato, M., Notake, M., Nakanishi, S. & Numa, S. (1982) *Nature (London)* **297,** 431–434.
20. Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer,

E. & Goodman, H. M. (1980) *Nature (London)* **284,** 26–32.
21. Gubler, U., Chua, A. O., Hoffman, B. J., Collier, K. J. & Eng, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4307–4310.
22. Eysselein, V. E., Reeve, J. R., Jr., Shively, J. E., Hawke, D. & Walsch, J. H. (1982) *Peptides* **3,** 687–691.
23. Rehfeld, J. F. (1978) *J. Biol. Chem.* **253,** 4022–4030.
24. Mutt, V. & Jorpes, E. (1981) *Biochem. J.* **125,** 57–58.
25. Rehfeld, J. F. (1981) *Am. J. Physiol.* **240,** 255–266.
26. Dockray, G. J. (1977) *Nature (London)* **270,** 359–361.
27. Kato, K., Hayashizaki, Y., Takahashi, Y., Himeno, S. & Matsubara, K. (1983) *Nucleic Acids Res.* **11,** 359–361.
28. Wiburg, O., Berglund, L., Boel, E., Norris, F., Norris, K., Rehfeld, J. F., Marcker, K. A. & Vuust, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1067–1069.