Molecular cloning of the human thyrotropin-β subunit gene

Yoshihide Hayashizaki*+, Kiyoshi Miyai*, Kikuya Kato+ and Kenichi Matsubara+

*Department of Laboratory Medicine and Central Laboratory for Clinical Investigation, Osaka University Medical School, 1-1-50 Fukushima, Fukushima-ku, Osaka 553, and +Institute for Molecular and Cellular Biology, Osaka University, Yamadaoka, Suita, 1-3 Osaka 565, Japan

Received 21 June 1985; revised version received 3 July 1985

Genomic DNA fragments that carried a gene for human thyrotropin-β (hTSHβ) subunit were isolated. Nucleotide sequence analysis of the gene showed that the hTSHβ subunit precursor consists of 138 amino acid residues. There is an N-terminal sequence of 20 amino acids as a signal peptide, followed by 112 amino acids, whose sequence is in agreement with that known for the secretory form of hTSHβ subunit. This is followed by an additional stretch of 6 hydrophobic amino acids, which may be eliminated post-translationally. The coding region is separated by an intron of about 460 bp. Genomic Southern blot hybridization analysis suggested that the hTSHβ gene is a unique single copy gene.

Human thyrotropin-β subunit gene *DNA cloning* *DNA sequence* *Hormone precursor* *COOH-terminal extension* *Hormone ancestor gene*

1. INTRODUCTION

Human thyroid stimulating hormone (hTSH) is a glycoprotein hormone secreted from the anterior pituitary lobe, which plays an important physiological role in regulation of the hypothalamus-pituitary-thyroid axis by stimulating production and secretion of the thyroid hormone from the thyroid gland. TSH consists of 2 non-covalently linked subunits, α and β [1,2], and is a member of a closely related group of glycoprotein hormones, which include the luteinizing hormone (LH), chorionic gonadotropin (CG), and follicle stimulating hormone (FSH). These hormones share the same α subunit, which may be encoded by a single gene [3], but have different β subunits, which may have been differentiated for the specific functions of the hormones. Excess α subunits, but not β subunits, can be found in the free form in normal pituitary [1,4] and in normal placenta [1,5], suggesting that the syntheses of the α and β subunits are regulated independently and that synthesis of the specific β subunit may be the limiting process in production of these hormones.

The cDNA and gene for the α subunit have been cloned and the nucleotide sequences of human [3,6], mouse [7], bovine [8,9] and rat [10] preparations have been determined. Only the human CGβ [18–20], human LHβ [20] and rat LHβ [10,21,26] genes for the β subunit of the glycoprotein hormones have been reported. Mouse [11] and bovine [12] TSHβ subunit cDNA have been cloned and studied. To investigate the control of hTSHβ expression, which plays a key role in the regulation of the hypothalamus-pituitary-thyroid axis, and also the mechanism of overproduction of TSH in some adenomas and its deficiency in some patients [28,29], we attempted to isolate the hTSHβ gene. The mRNA of hTSHβ should be produced only in the anterior lobe, which is not readily available as material for study. So we cloned the gene directly from human DNAs using bovine TSHβ cDNA as a probe.

Here, we report the nucleotide sequence of the hTSHβ gene. This gene codes for a polypeptide consisting of a signal sequence of 20 amino acids, 'mature' TSHβ subunit of 112 amino acids, and a sequence of 6 hydrophobic amino acids at the COOH-terminal. It is probably a single gene.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

2. MATERIALS AND METHODS

2.1. *Enzymes*

Restriction endonucleases, DNA polymerase I, and the Klenow fragment were purchased from Takara Shuzo (Kyoto, Japan). T_4 ligase was a gift from Dr T. Tsurimoto.

2.2. *Genomic DNA Southern blot hybridization*

Human leukocyte DNA was isolated as described [27]. Approx. 10 μg of this DNA was digested with *Eco*RI, *Pvu*II, *Hin*dIII or other appropriate restriction enzymes, and the digests were subjected to electrophoresis and Southern filter blot hybridization [13]. A bovine TSHβ cDNA insert fragment, kindly provided by Dr R. Maurer [12], was labelled with [α-^{32}P]dCTP by nick-translation to a specific activity of 1–5 $\times$ 10^8 cpm/μg DNA. The filters were prehybridized [27], and then hybridization was performed at 65°C, for 17–18 h in 6 $\times$ SSC/0.1% SDS/1 $\times$ Denhardt's solution [27]/10 μg per ml of salmon sperm DNA. The filters were washed and subjected to autoradiography as described in [27].

2.3. *Human genomic DNA libraries*

Two human genomic libraries were used. The first library was constructed by joining complete *Eco*RI cleaved human liver genomic DNA to λ Charon 28 arms (kindly provided by Mr T. Nagaya and Mr T. Nakamura). The initial titer of recombinant clones was 8.0 $\times$ 10^5. This library was screened without amplification by the method of Benton and Davis [14].

The second library, in which the hTSHβ gene was partially enriched, was constructed from *Eco*RI-digested human leukocyte DNA. The digest of 1200 μg DNA was subjected to electrophoresis on 0.7% agarose gel and vertical sections of the agarose were subjected to Southern filter blot hybridization using bovine TSHβ cDNA [12] as a probe. Two bands of 2.2 and 3.2 kbp appeared (see section 3). The DNAs of corresponding sizes were eluted from the remaining agarose gel by electrophoresis and purified in a CsCl-ethidium bromide density gradient. Each DNA was inserted into the *Eco*RI site of pBR325 and transformed into *E. coli* RRI. About 4 $\times$ 10^5 colonies were produced, and screened by the method of Hanahan and Meselson [15] with bovine TSHβ cDNA as a probe.

2.4. *DNA sequence analysis*

Sequencing was done by the method of Sanger et al. [16] with M13 mp10, and mp11 as cloning vectors.

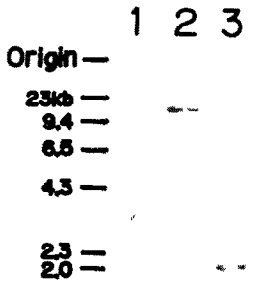


Fig.1. Genomic Southern blot hybridization analysis of the hTSHβ subunit gene. Human leukocyte DNA was digested with *Eco*RI (lane 1), *Hin*dIII (lane 2), and *Pvu*II (lane 3). After electrophoresis on 0.7% agarose gel, the DNA was transferred to a nitrocellulose filter, and then hybridized to the α-^{32}P-labelled 320 bp cDNA fragment encoding the bovine TSHβ subunit as described in section 2. The positions of marker DNAs in *Hin*dIII-digested λ phage DNA are shown on the left.

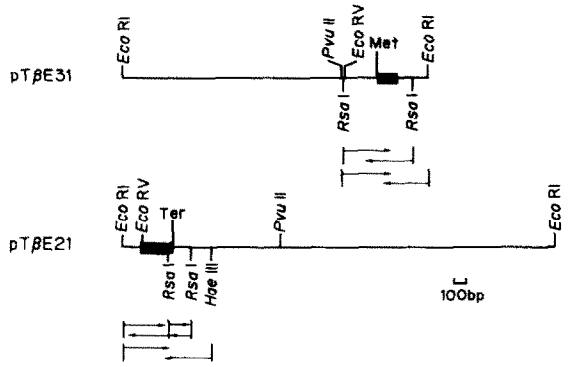


Fig.2. Restriction map and sequencing strategy of pTβE21 and pTβE31 that, respectively, carry the 2.2 and 3.2 kbp DNA fragments of the hTSHβ gene. Only the restriction sites used in this study are displayed. Solid bars indicate the amino acid coding regions as deduced from the nucleotide sequence (see fig.3). Met and Ter represent the initiator methionine and termination codons, respectively.

3. RESULTS AND DISCUSSION

3.1. *Genomic Southern blot hybridization*

The 320 bp fragment of the bovine TSHβ cDNA, provided by Dr R. Maurer [12], encodes part (9 amino acid residues) of the signal peptide and 99 amino acid residues of the mature TSHβ subunit. With this fragment as a probe, Southern blot hybridizations of *Hin*dIII- and *Pvu*II-fragments of human genomic DNA from leukocytes each gave a single band (fig.1, lanes 2 and 3), while *Eco*RI-fragments of the DNA gave bands of 3.2 and 2.2 kbp (fig.1, lane 1). Fragments of the corresponding size were cut out from the gel and cloned into the *Eco*RI site of pBR325 to construct a pBR325 library.

```
CAGCTGTACATATTTCCACCTTAAAGGGATATCCTAAGGGTTTGGAAGTGGGATCAGGGGGTTCCTAGATTTCTGAGTTAGCCCCCTTAACACCAGTTGTA

ATTTCAGTTGACCTTTTTTGGACTTTATCTTTCTGGTGTCTTCCTTGACCAAATGGTAGAATTATAAGCATGATCATATGCATTGGGATGGTACTGAAGT

                                                                                    *a
TTGGTTATACTTTTTCTTGGTTTCTTTGCCCTTTCTGATTTTAACAAATAGGTTCTTTAATTTTATCTTTGATTTAGC ATG ACT GCT CTC TTT CTG
                                                                                      Met Thr Ala Leu Phe Leu
                                                                                       +1

 ATG TCC ATG CTT TTT GGC CTT GCA TGT GGG CAA GCG ATG TCT TTT TGT ATT CCA ACT GAG TAT ACA ATG CAC ATC
 Met Ser Met Leu Phe Gly Leu Ala Cys Gly Gln Ala Met Ser Phe Cys Ile Pro Thr Glu Tyr Thr Met His Ile
             +10                                  +20                                      +30

                                                                                        *b
 GAA AGG AGA GAG TGT GCT TAT TGC CTA ACC ATC AAC ACC ACC ATC TGT GCT GGA TAT TGT ATG ACA CGG GTATGTA
 Glu Arg Arg Glu Cys Ala Tyr Cys Leu Thr Ile Asn Thr Thr Ile Cys Ala Gly Tyr Cys Met Thr Arg
                                +40                                  +50            +54

 GTTCATGTCACTTCTTTTGGCTG.....0.46 kbp intron.....ATTATGCTCTCTTTTCTGTTCTTTCCCCAG

*c
 GAT ATC AAT GGC AAA CTG TTT CTT CCC AAA TAT GCT CTG TCC CAG GAT GTT TGC ACA TAT AGA GAC TTC ATC TAC
 Asp Ile Asn Gly Lys Leu Phe Leu Pro Lys Tyr Ala Leu Ser Gln Asp Val Cys Thr Tyr Arg Asp Phe Ile Tyr
 +55                 +60                                  +70

 AGG ACT GTA GAA ATA CCA GGA TGC CCA CTC CAT GTT GCT CCC TAT TTT TCC TAT CCT GTT GCT TTA AGC TGT AAG
 Arg Thr Val Glu Ile Pro Gly Cys Pro Leu His Val Ala Pro Tyr Phe Ser Tyr Pro Val Ala Leu Ser Cys Lys
 +80                                  +90                                      +100

 TGT GGC AAG TGC AAT ACT GAC TAT AGT GAC TGC ATA CAT GAA GCC ATC AAG ACA AAC TAC TGT ACC AAA CCT CAG
 Cys Gly Lys Cys Asn Thr Asp Tyr Ser Asp Cys Ile His Glu Ala Ile Lys Thr Asn Tyr Cys Thr Lys Pro Gln
                   +110                                  +120

                                    *d
 AAG TCT TAT CTG GTA GGA TTT TCT GTC TAA TAGTGATATAATTTGCAATTTGGTTAAATGTGCTTGCCTGAAATAAAGCTAATAAAAAT
 Lys Ser Tyr Leu Val Gly Phe Ser Val Ter
+130    +132                  +138

 ATTATGTTTCACATTATCTTCTGTTCATTTTGAGTACTATTTAATCCATACCC
```

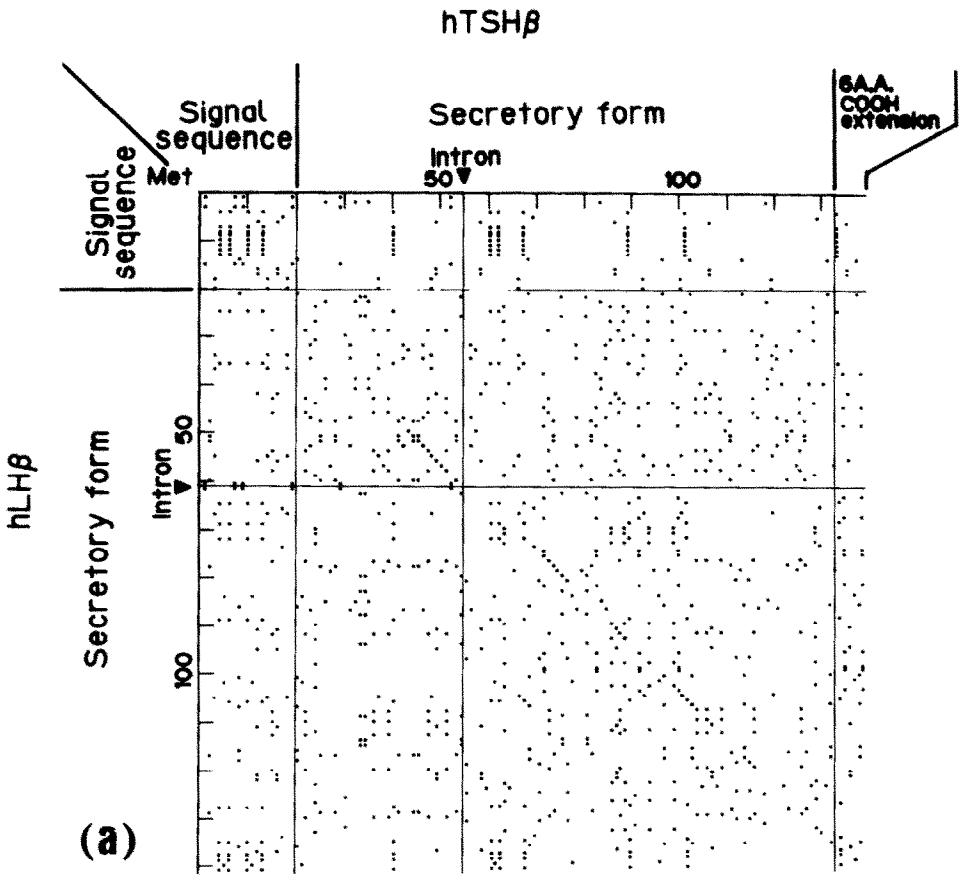
Fig.3. Nucleotide sequence of the hTSHβ subunit gene. Nucleotide sequencing was done as described in section 2. The deduced amino acid sequence is also displayed and the putative initiator methionine in the prepro TSHβ is numbered +1. The nucleotide sequences from *a to *b and from *c to *d represent the entire sequence coding for 138 amino acids, amino acids +54 and +55 being interrupted by an intron. The first 20 amino acids may be a signal peptide. The amino acid sequence from Phe at residue +21 to Tyr at residue +132 matches that of mature hTSHβ polypeptide, as determined by amino acid sequencing [17].

3.2. *Cloning of human DNA fragments carrying the hTSHβ gene*

The λ Charon 28 genomic DNA library carrying *Eco*RI digested human liver DNA and the pBR325 DNA library carrying *Eco*RI digested human leukocyte DNA, were employed for screening clones for the TSHβ gene. About 8.0×10^5 recombinant phages were surveyed, with bovine TSHβ cDNA [12] as a probe, and one positive clone was obtained. Southern blot hybridization of this phage clone showed that it carried the 3.2 kbp component. This 3.2 kbp *Eco*RI fragment was subcloned into the *Eco*RI site of pBR322 to yield pTβE21. With the same bovine cDNA probe, 5 positive clones were obtained from approx. 4×10^5 colonies in one of the pBR325 libraries carrying the 2.2 kbp *Eco*RI fragment. One clone (pTβE31), that was confirmed to carry the 2.2 kbp *Eco*RI fragment, was used for subsequent analyses.

3.3. *Nucleotide sequence of the hTSHβ gene*

Restriction maps of pTβE21 and pTβE31 were constructed and are shown in fig.2. Nucleotide sequencing was done with the whole cloned fragments or parts of them by the strategies shown in the same figure. The nucleotide sequence is shown in fig.3. In the 2.2 kbp DNA fragment carried in pTβE31, the putative initiator methionine of the prepro TSHβ was detected and was numbered as the amino acid residue + 1. An open reading frame is seen in the region between *a and *b (see legend to fig.3). The first part of this region, which may be a single peptide, consists of 20 amino acids and is rich in hydrophobic residues. The deduced amino acid sequence from +21 to +54 matches the amino acid sequence determined by analysis of the hTSH subunit protein [1,17] except for +28 and +29 as described below. The amino acid coding sequence is interrupted after



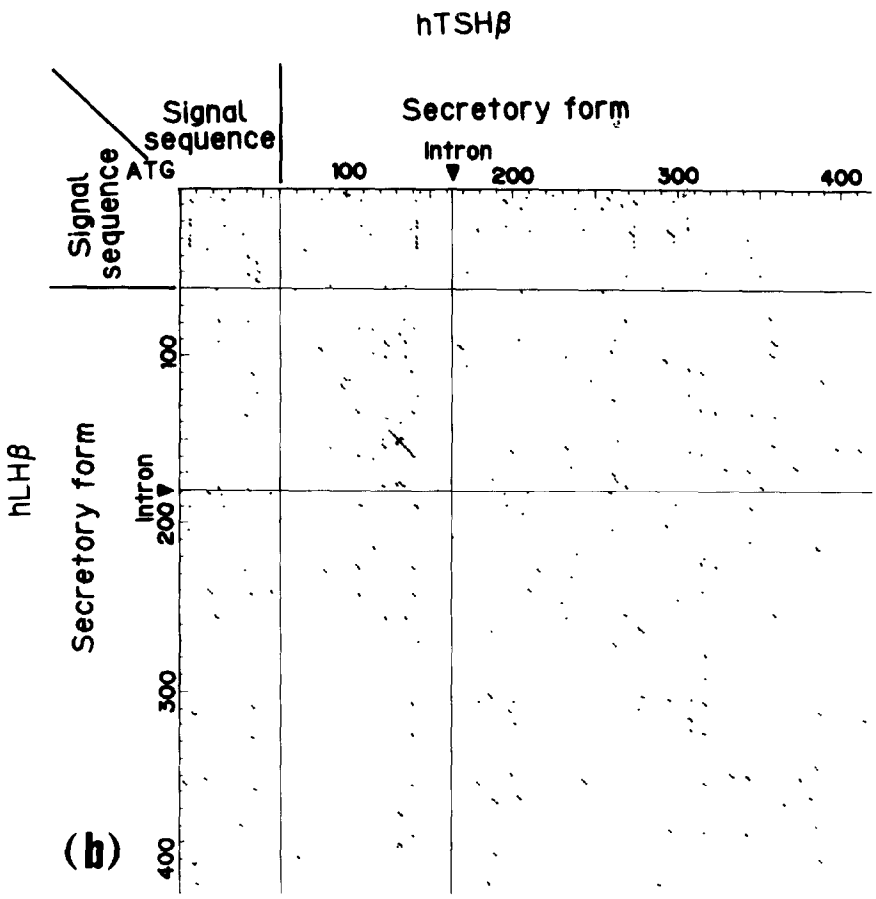


Fig.4. Harr plot analysis for comparison of the nucleotide and amino acid sequences of hTSHβ and hLHβ. Harr plots to demonstrate amino acid (a) and nucleotide (b) sequence homologies between hTSHβ and hLHβ and their corresponding genes. The amino acid and nucleotide sequences of hLHβ and its gene are from Maghuin-Rogister et al. [31] and Talmadge et al. [20]. Each dot represents the position where 1 amino acid or 4 nucleotides match in the 2 sequences. Arrowheads represent the positions of introns.

amino acid +54 from which the GT sequence, corresponding to the consensus intron donor site sequence, starts. The nucleotide sequence matching the rest of the hTSHβ amino acid sequence from residues +55 to +132 is present in the 3.2 kbp DNA fragment carried in pTβE21, preceded by the consensus intron acceptor site sequence AG. The codon of amino acid +132 is followed by 6 codons for hydrophobic amino acids and then by the termination codon TAA. These data indicate that the human DNA fragments carried in pTβE31 and pTβE21 cover the hTSHβ gene and that they can be aligned in this order with a DNA region of about 0.46 kbp located between codon +54 and +55. We conclude that this region represents an intron. We cannot exclude the possibility that a small *Eco*RI fragment may exist between the 2.2 and 3.2 kbp DNA fragments. Our deduced amino acid sequence differs from that obtained by amino acid sequencing [1,17] at amino acids +28 and +29, where we found threonine and methionine residues instead of methionine and threonine residues.

The initiator methionine codon at +1 was assigned for the following reasons: (i) there is a termination codon TGA 6 bp 5′-upstream from the methionine codon; (ii) among the 3 methionine codons preceding phenylalanine at residue +21, the methionine at +1 is followed by a sequence of hydrophobic amino acids with a size correspond-

ing to that of a signal peptide; (iii) comparison of this sequence with that of mouse TSHβ cDNA, whose initiator methionine codon has been unambiguously assigned [11], strongly suggests that the first methionine is the initiator. So far, the TSHβ coding sequences have been analysed only with mouse and bovine cDNAs. Our work demonstrated, for the first time, the location of the intron in the TSHβ gene.

The nucleotide sequence of the coding region of hTSHβ shows 84.7% and 89.9% homologies with the mouse and bovine coding region, respectively. However, there was no appreciable homology in the 5′-untranslated regions of the hTSHβ gene and bovine TSHβ cDNA [12] as judged by cross-hybridization tests. The absence of homology in nucleotide sequences and the fact that hTSHβ mRNA is not readily available, prevented us from determining the 5′- and 3′-untranslated regions of this gene, or its promoter region. Similarly, the possible presence of another intron in this region cannot be excluded. For the same reason, we think it is premature to discuss the 3′-untranslated regions. These problems require further study.

The hTSHβ precursor deduced from the nucleotide sequence has a stretch of 6 hydrophobic amino acids attached to the carboxyl-terminus of the mature TSHβ polypeptide. The presence of such an extension at the carboxyl-terminus has also been noted with bovine TSHβ cDNA [12] in which, however, the extension is 5 amino acid residues. This extension seems unlikely to be a cloning artifact, because the sequence of human and bovine extensions show high homology. Moreover, from mouse cDNA studies a similar extension polypeptide has been predicted [11]. The mature hTSHβ subunit may, then, be produced by posttranslational processing. Elimination of the extension polypeptide might induce a conformational change of the protein molecule, as suggested by Pierce and Parsons [1].

3.4. *Copy number of the hTSHβ gene*

As shown in fig.1, the Southern blot patterns of *Pvu*II and *Hin*dIII fragments of human genomic DNA each gave single bands of about 2.0 and 14 kbp, respectively. The *Eco*RI fragments of the DNA gave two bands, in accordance with the notion that the short intron carries an *Eco*RI site(s). These observations strongly suggest that the human TSHβ gene is a single copy gene.

3.5. *Comparison of the hTSHβ gene with the β genes of other glycoprotein hormones*

The general features of the hTSHβ gene are similar to those of genes for the β subunits of other glycoproteins such as hCG or hLH [18–21]. The location of the intron within the amino acid coding region of hTSHβ gene is identical with those known for hCGβ and hLHβ genes. Another intron is located 15 bp downstream from the translational initiator methionine codon in both the hCGβ and hLHβ genes, as shown elsewhere [20]. This intron is not present in the hTSHβ gene, suggesting that the hTSHβ gene diverged from a common ancestral gene before the other 3 glycoprotein β subunit genes during evolution. The amino acid sequence is conserved to some extent and homologous regions are found in both exon 2 and 3 of hLHβ and hTSHβ (fig.4a). The nucleotide sequence, on the other hand, is poorly conserved with homologous regions only in exon 2 of the hLHβ gene (fig.4b). This narrow but highly conserved region, both in terms of amino acid sequence and nucleotide sequence, was identical to the corresponding region in β subunit of other glycoproteins [30] that is known as the CAGY region. From the standpoint of molecular evolution, it is of interest that the region of complete identity in nucleotide sequence is just limited to the CAGY region among other observed regions of homology in amino acid sequence.

ACKNOWLEDGEMENTS

We thank Dr R. Maurer for providing us with bovine TSHβ cDNA, Mr Y. Takahashi and Dr K. Ichihara for stimulating discussion and valuable advice throughout this work and Mr T. Nagaya and Mr T. Nakamura for providing the human genomic library. We also thank Dr S. Hayashizaki for her technical assistance and Miss Nishimura and Miss Yamada for secretarial assistance.

REFERENCES

[1] Pierce, J.G. and Parsons, T.F. (1981) Annu. Rev. Biochem. 50, 465–495.
[2] Pierce, J.G. (1971) Endocrinology 89, 1331–1344.

[3] Fiddes, J.C. and Goodman, H.M. (1981) J. Mol. Appl. Genet. 1, 3–18.
[4] Kourides, I.A., Weintraub, B.D., Ridgway, E.C. and Maloof, F. (1975) J. Clin. Endocrinol. Metab. 40, 872–885.
[5] Vaitukaitis, J.L., Ross, G.T., Braunstein, G.D. and Rayford, P.L. (1976) Recent Prog. Horm. Res. 32, 289–331.
[6] Fiddes, J.C. and Goodman, H.M. (1979) Nature 281, 351–356.
[7] Chin, W.W., Kronenberg, H.M., Dee, P.C., Maloof, F. and Habener, J.F. (1981) Proc. Natl. Acad. Sci. USA 78, 5329–5333.
[8] Erwin, C.R., Croyle, M.L., Donelson, J.E. and Maurer, R.A. (1983) Biochemistry 22, 4856–4860.
[9] Goodwin, R.G., Moncman, C.L., Rottman, F.M. and Nilson, J.H. (1983) Nucleic Acids Res. 11, 6873–6882.
[10] Godine, J.E., Chin, W.W. and Habener, J.F. (1982) J. Biol. Chem. 257, 8368–8371.
[11] Gurr, J.A., Catterall, J.F. and Kourides, I.A. (1983) Proc. Natl. Acad. Sci. USA 80, 2122–2126.
[12] Maurer, R.A., Croyle, M.L. and Donelson (1984) J. Biol. Chem. 259, 5024–5027.
[13] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[14] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[15] Hanahan, D. and Meselson, M. (1980) Gene 10, 63–67.
[16] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[17] Sairam, M.R. and Li, C.H. (1977) Can. J. Biochem. 55, 755–760.
[18] Fiddes, J.C. and Goodman, H.M. (1980) Nature 286, 684–687.
[19] Boorstein, W.R., Vamvakopoulos, N.C. and Fiddes, J.C. (1982) Nature 300, 419–422.
[20] Talmadge, K., Vamvakopoulos, N.C. and Fiddes, J.C. (1984) Nature 307, 37–40.
[21] Policastro, P., Ovitt, C.E., Hoshina, M., Fukuoka, H., Boothby, M.R. and Boime, I. (1983) J. Biol. Chem. 258, 11492–11499.
[22] Fontaine, Y.A. and Burzuwa-Gerard, E. (1977) Gen. Comp. Endocrinol. 32, 341–347.
[23] Giudice, L.C. and Weintraub, B.D. (1979) J. Biol. Chem. 254, 12679–12683.
[24] Giudice, L.C. and Pierce, J.G. (1976) J. Biol. Chem. 251, 6392–6399.
[25] Giudice, L.C. and Pierce, J.G. (1978) Biochim. Biophys. Acta 533, 140–146.
[26] Chin, W.W., Godine, J.E., Klein, D.R., Chang, A.S., Tan, L.K. and Habener, J.F. (1983) Proc. Natl. Acad. Sci. USA 80, 4649–4653.
[27] Takahashi, Y., Kato, K., Hayashizaki, Y., Watabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. and Matsubara, K. (1985) Proc. Natl. Acad. Sci. USA 82, 1931–1935.
[28] Miyai, K., Azukizawa, M. and Kumahara, Y. (1971) N. Engl. J. Med. 285, 1043–1048.
[29] Kohno, H., Watanabe, N., Ootsuka, M., Kajiwara, M. and Gohya, N. (1980) Arch. Dis. Child. 55, 725–727.
[30] Mullin, B.R., Fishman, P.H., Lee, G., Aloj, S.M., Ledley, F.D., Winand, R.J., Kohn, L.D. and Brady, R.O. (1976) Proc. Natl. Acad. Sci. USA 73, 842–846.
[31] Maghuin-Rogister, G., Combarnous, Y. and Hennen, G. (1973) Eur. J. Biochem. 39, 255–263.