ISOLATION AND STRUCTURAL ORGANIZATION OF THE HUMAN CORTICOTROPIN–β-LIPOTROPIN PRECURSOR GENE

Hideo TAKAHASHI, Yutaka TERANISHI, Shigetada NAKANISHI and Shosaku NUMA
Department of Medical Chemistry, Kyoto University, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606, Japan

Received 12 October 1981

1. Introduction

The common precursor of corticotropin (ACTH) and β-lipotropin (β-LPH) [1–8] contains multiple biologically active peptides. We have elucidated the whole primary structure of the bovine ACTH–β-LPH precursor by determining the nucleotide sequence of cloned DNA complementary to the mRNA coding for the precursor protein [9]. Using this bovine cDNA as a hybridization probe, we have then isolated the entire bovine gene encoding the ACTH–β-LPH precursor as a set of genomic DNA fragments and have determined its structural organization [10,11]. We have now isolated and characterized a human genomic DNA fragment containing the entire ACTH–β-LPH precursor gene. The human gene, like the bovine counterpart, consists of 3 exons (mRNA-coding sequences) divided by 2 large introns (intervening sequences). Nucleotide sequence analysis of the whole exons has revealed the complete mRNA and amino acid sequences of the human ACTH–β-LPH precursor.

2. Materials and methods

DNA was prepared from a human placenta as in [12]. Human genomic DNA clones were isolated as in [10,11,13]. All cloning procedures were performed under P2 containment by using the approved EK2 host/vector systems, λgtWES · λB/DP50supF and pBR322/X1776, according to the Japanese guidelines for recombinant DNA research. Restriction mapping was accomplished by combining data obtained by digestion and subsequent gel electrophoresis of both 5′-end-labelled and unlabelled DNA fragments and by blot hybridization analysis [14]; the procedures used were as in [10,11], except that prehybridization, hybridization and filter washing were conducted at 60°C. DNA sequencing was carried out by the procedure in [15]. Reagents were obtained as in [11].

Abbreviations: ACTH, corticotropin; β-LPH, β-lipotropin; γ-LPH, γ-lipotropin; α-MSH, α-melanotropin; β-MSH, β-melanotropin; γ-MSH, γ-melanotropin; CLIP, corticotropin-like intermediate lobe peptide

Dedicated to Professor Dr Helmut Holzer on the occasion of his 60th birthday

3. Results and discussion

Human placental DNA was digested to completion with *Eco*RI and electrophoresed on 0.7% agarose gel. Blot hybridization analysis with the cloned bovine cDNA (*Pst*I-excised insert of plasmid pSNAC20 [8]) as a probe displayed a single hybridization-positive band corresponding to a size of ~11.5 kilobase pairs. Therefore, a fraction of *Eco*RI-digested human placental DNA having a size of 10–12 kilobase pairs was isolated by preparative agarose gel electrophoresis and used for cloning into bacteriophage λgtWES. Recombinant phage that exhibited a positive hybridization signal with the bovine cDNA probe were screened, and the 11.5-kilobase pair *Eco*RI fragment derived from the phage isolated was subcloned in plasmid pBR322.

A restriction map of the cloned 11.5-kilobase pair *Eco*RI fragment was constructed (fig.1A) and DNA fragments containing an mRNA-coding sequence were identified by blot hybridization analysis with the bovine cDNA or with a bovine genomic DNA fragment containing exon 1 (1.7-kilobase pair *Bam*HI fragment [11]) as a probe. The 3 hybridization-positive

Published by Elsevier/North-Holland Biomedical Press
00145793/81/0000–0000/$02.75

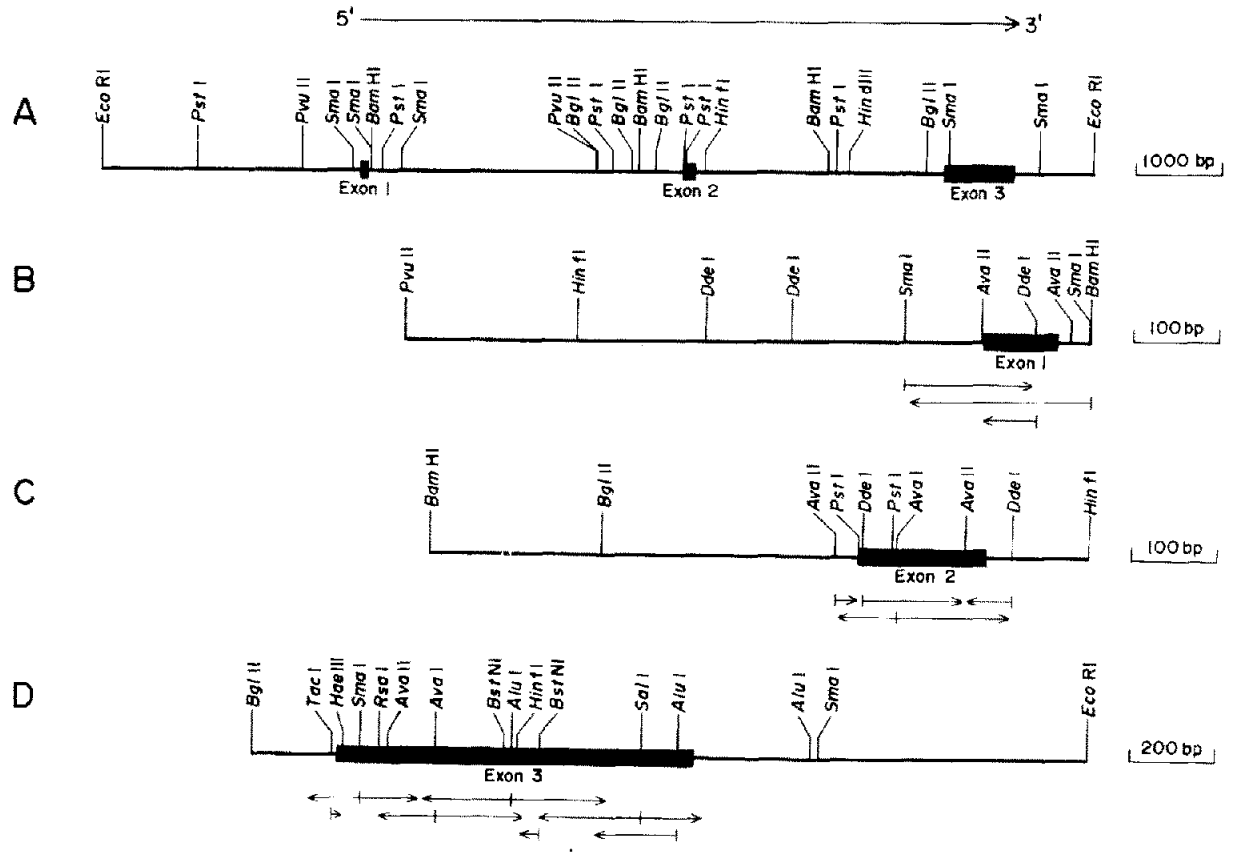

Fig.1. Restriction mapping of cloned human genomic DNA containing the ACTH–β-LPH precursor gene and sequencing strategy. (A) 11.5-kilobase pair *Eco*RI fragment; (B–D) portions of the 11.5-kilobase pair *Eco*RI fragment containing exon 1, exon 2 and exon 3, respectively. The direction of transcription is from left–right. Scales are given in basepairs (bp) on the right side of each restriction map. For reference, the locations of exons are shown by closed boxes. Only relevant restriction sites are displayed for: *Pvu*II and *Hin*fI (A); *Dde*I (B); *Dde*I and *Ava*II (C); and *Tac*I, *Hae*III, *Rsa*I, *Ava*II, *Ava*I, *Bst*NI, *Alu*I and *Hin*fI (D). The horizontal arrows beneath the restriction maps in (B–D) indicate the direction and extent of sequence determinations.

Fig.2. Nucleotide sequences of the exons and their surrounding regions of the human ACTH–β-LPH precursor gene and comparison with those of the bovine gene. The nucleotide sequence of the human message strand, together with the deduced amino acid sequence, is shown, and the nucleotide and amino acid differences found in the bovine sequences (data from [10,11]) are displayed under the human sequences; the absence of a nucleotide or an amino acid in the bovine sequence indicates that the human and bovine sequences are the same; the presence of a colon in either nucleotide sequence indicates a gap; the amino acid residues in the human sequence are numbered, beginning with the tryptophan residue encoded by the 27th codon, and the preceding residues constituting the signal peptide are indicated by negative numbers. Large and small capital letters represent the sequences of exons and their surrounding regions, respectively. The exon/intron junctions are positioned according to the GT/AG rule [16]; the redundant nucleotide residues at the junctions are underlined. The sequences at both ends of the introns that are complementary to the 5'-end of U1 small nuclear RNA are overlined. The putative sites of capping and poly(A) addition are indicated. The 'Hogness box' [17] is enclosed with a box. An imperfect palindrome located in the 5'-flanking region is shown by arrows starting at its centre above the nucleotide sequence. The sequence of exon 3 of the human gene in [18] differs from ours as follows (the numbers in parentheses refer to the nucleotide numbers of the human exon 3 sequence we determined, beginning with its 5'-end): substitutions, A (the residue immediately preceding exon 3), T (6), G (10), C (12), G (375), G (414), C (450), C (455), A (618), C (695), A (696); deletions (456, 472, 477, 760); additions A (703–704), T (722–723), G (729–730), T (796–797).

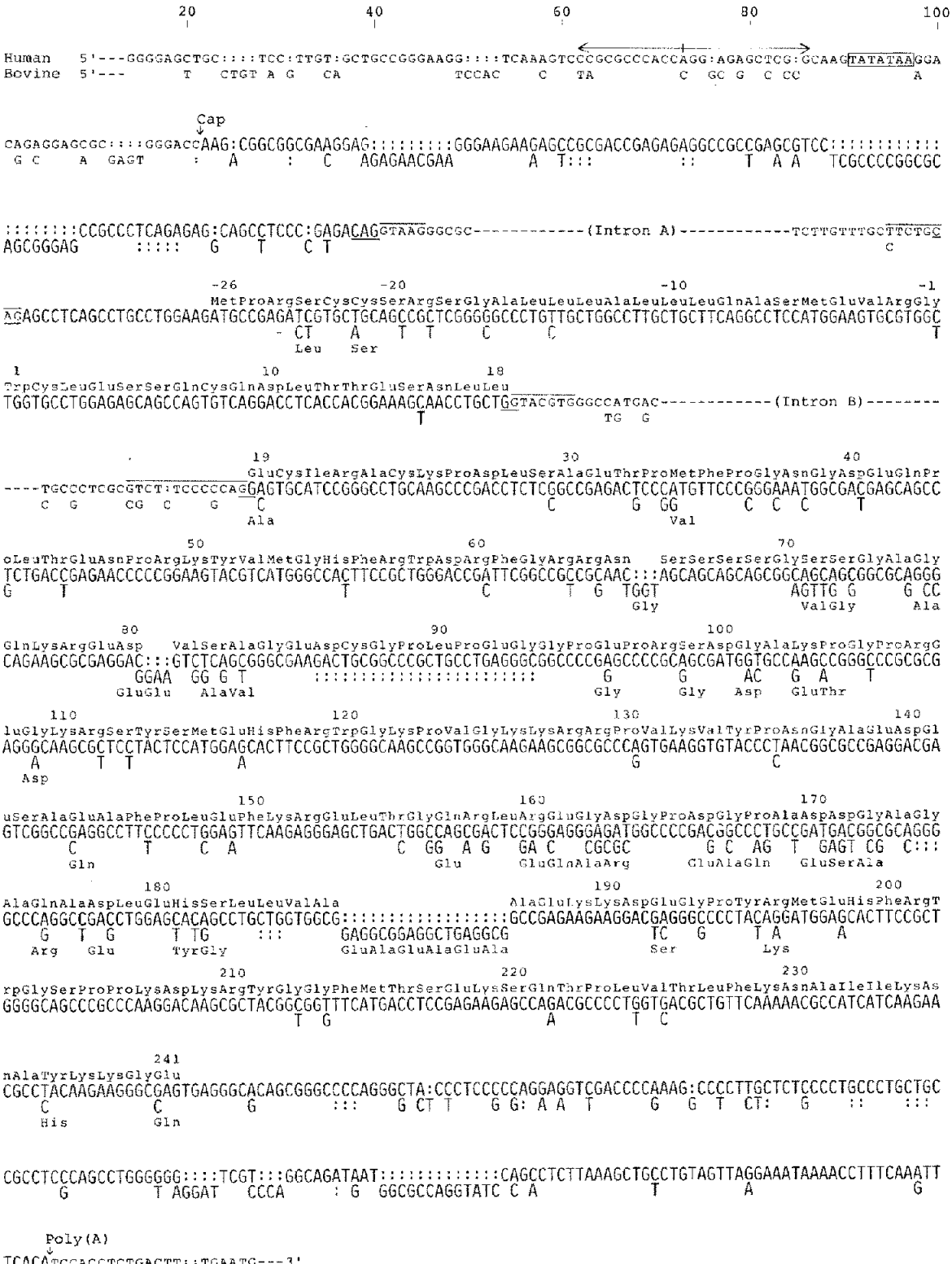

Fig.2.

DNA segments shown in fig.1B–D were subjected to nucleotide sequence analysis according to the strategy indicated.

The human DNA sequence determined is presented in fig.2. Comparison with the nucleotide sequence of the bovine ACTH–β-LPH precursor gene [10,11] enables us to determine the structural organization of the human gene. The mRNA-coding sequence of the human ACTH -β-LPH precursor gene is divided by 2 large introns into 3 exons (exon 1, exon 2 and exon 3 oriented in the direction of transcription). An intron of ~3.6 kilobase pairs (intron A) interrupts the segment encoding the 5′-untranslated region of the mRNA between 20 and 21 nucleotide residues upstream from the translational initiation site. The other intron of ~2.9 kilobase pairs (intron B) is located within the protein-coding sequence, separating the segment encoding the amino-terminal 44 amino acid residues including the signal peptide from that encoding the whole remaining amino acid residues and the 3′-untranslated region of the mRNA. Thus, the introns of the human gene are located at exactly the same positions as those of the bovine counterpart; the lengths of intron A and intron B of the bovine gene are ~4 and 2.2 kilobase pairs, respectively [10,11]. By comparison with the bovine gene sequence, the capping site and the poly(A) addition site of the human gene have been assigned tentatively. On the basis of this assignment and the GT/AG rule for exon/intron boundaries [16], the lengths of exon 1, exon 2 and exon 3 of the human gene are 87, 152 and 833 basepairs, respectively; the lengths of the bovine counterparts are 108, 152 and 838 basepairs, respectively.

The complete amino acid sequence of the human ACTH–β-LPH precursor has been deduced from the mRNA-coding sequence. The human precursor protein is composed of 267 amino acid residues, having a calculated M_r of 29 455; the bovine precursor protein consists of 265 amino acid residues. The signal peptide, i.e., the amino-terminal 26 amino acid residues (review [19]), of the human precursor protein exhibits 2 amino acid substitutions as compared with the bovine counterpart (Ser instead of Leu at residue −23; Cys instead of Ser at residue −21). The following 18 amino acid residues encoded by exon 2 are identical in the 2 species. A human genomic DNA segment containing exon 3 has been isolated and sequenced in [18]. There are 19 differences between their nucleotide sequence data and ours as listed in fig.2. Owing to some of these differences, our deduced amino acid sequence shows the substitution of Arg (residue 22) for Gly and that of Asp–Asp–Gly–Ala–Gly–Ala–Gln–Ala (residues 170–177) for Ala–Thr–Ala–Gln–Gly–Pro–Gly. The latter replacement located in the region between ACTH and β-melanotropin (β-MSH) is due to shifts of the reading frame resulting from the deletion of 3 nucleotide residues in the sequence of [18]. The other nucleotide differences in the protein-coding sequence lead to synonymous codons. We have assigned an additional amino acid residue (Glu) at the 5′-end of exon 3.

The sequence of the 76 amino acid residues following the signal peptide of the human ACTH–β-LPH precursor deduced from our nucleotide sequence is in complete agreement with the sequence of a human pituitary glycopeptide determined in [20]. However, the amino acid sequence of γ-lipotropin (γ-LPH) deduced from our nucleotide sequence disagrees partly with that determined in [21], which exhibits differences in 5 residues (Gln for residue 161; Asn for residue 166; Ala for residue 167; Gly for residue 168; Pro for residue 173). The differences observed may be due to possible polymorphism in the human gene because the DNA sequence we report here is based on analysis of a single DNA clone. It cannot be excluded either that these differences are due to possible nucleotide changes that have occurred during replication of recombinant DNA.

Fig.3 illustrates schematically the degree of nucleotide sequence homology in the different structural and functional regions of the human and bovine ACTH–β-LPH precursor genes. A high degree of conservation is observed in the segment encoding the sequence extending from the signal peptide to γ-melanotropin (γ-MSH), the ACTH region and the segment encoding β-MSH and β-endorphin. The segments preceding and following the ACTH region as well as the 5′- and 3′-untranslated regions and the segment preceding the capping site are moderately conserved. Less homology is found in the segment following the poly(A) addition site [10,11,18]. The homology map suggests that γ-MSH and the amino-terminal fragment lying between the signal peptide and γ-MSH may be of physiological importance, as is the case for ACTH and β-endorphin. Examination of the nucleotide sequences surrounding the sites of deletion/addition in fig.3 reveals the presence of short direct repeats, which may be involved in the generation of these divergences in the 2 genes as pointed out in the case of β-like globin genes [23]. The fact that the intron

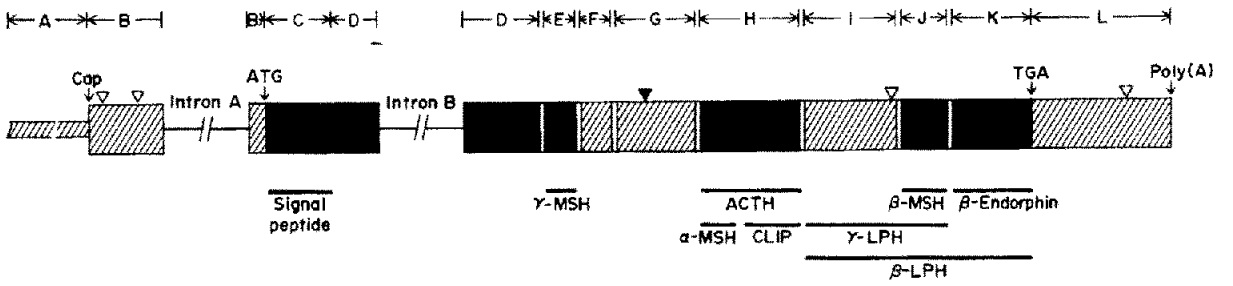

Fig.3. Nucleotide sequence homology in different regions of the human and bovine ACTH–β-LPH precursor genes. The mRNA-coding segment (thick blocks) and the segment preceding it (thin block) are divided into 12 structural and functional regions indicated by (A–L). Regions of high and moderate homology, based on the corrected percent divergence calculated for each region, are shown by closed and shaded bars, respectively. The open slits represent codons encoding paired basic amino acid residues that separate the component peptides, the locations of which are displayed beneath the gene sequence. The sites of capping, translational initiation (ATG), translational termination (TGA) and poly(A) addition are shown. (▽,▼) Positions of deletion and addition, respectively, of ≥9 basepairs in the human sequence as compared with the bovine sequence. The lines indicating introns do not represent their real lengths. The corrected percent divergences of replacement substitution sites (r), silent substitution sites (s) and total sites (t) of the protein-coding regions and those of the non-coding regions were calculated as in [22]; codons corresponding to gaps were excluded from the calculation for replacement and silent substitution sites, whereas gaps were counted as one substitution regardless of their length for the calculation for total sites of the protein-coding regions as well as for the non-coding regions. The corrected percent divergences calculated for the regions (A–L) are as follows: A, 29.3; B, 19.5; C, 4.8 (r), 41.6 (s) 11.0 (t); D, 1.9 (r), 38.5 (s), 8.8 (t); E, 0 (r), 46.1 (s), 5.8 (t); F, 17.6 (r), >83.9 (s), 37.9 (t); G, 19.3 (r), 27.2 (s), 24.8 (t); H, 1.2 (r), 39.4 (s), 7.2 (t); I, 30.4 (r), 69.9 (s), 42.4 (t); J, 7.6 (r), 55.0 (s), 12.0 (t); K, 2.9 (r), 31.0 (s), 7.9 (t); L, 26.0.

sequences adjacent to the exon/intron junctions are highly conserved suggests the relevance of these sequences to RNA splicing. The complementarity between the 5′-terminal sequence of U1 small nuclear RNA and the consensus sequences at both ends of introns is considered to be involved in the processing of hnRNA [24,25].

Acknowledgements

This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

References

[1] Nakanishi, S., Taii, S., Hirata, Y., Matsukura, S., Imura, H. and Numa, S. (1976) Proc. Natl. Acad. Sci. USA 73, 4319–4323.
[2] Nakanishi, S., Inoue, A., Taii, S. and Numa, S. (1977) FEBS Lett. 84, 105–109.
[3] Taii, S., Nakanishi, S. and Numa, S. (1979) Eur. J. Biochem. 93, 205–212.
[4] Kita, T., Inoue, A., Nakanishi, S. and Numa, S. (1979) Eur. J. Biochem. 93, 213–220.
[5] Mains, R. E. and Eipper, B. A. (1976) J. Biol. Chem. 251, 4115–4120.
[6] Mains, R. E., Eipper, B. A. and Ling, N. (1977) Proc. Natl. Acad. Sci. USA 74, 3014–3018.
[7] Roberts, J. L. and Herbert, E. (1977) Proc. Natl. Acad. Sci. USA 74, 5300–5304.
[8] Nakanishi, S., Inoue, A., Kita, T., Numa, S., Chang, A. C. Y., Cohen, S. N., Nunberg, J. and Schimke, R. T. (1978) Proc. Natl. Acad. Sci. USA 75, 6021–6025.
[9] Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A. C. Y., Cohen, S. N. and Numa, S. (1979) Nature 278, 423–427.
[10] Nakanishi, S., Teranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kakidani, H., Jingami, H. and Numa, S. (1980) Nature 287, 752–755.
[11] Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429–438.
[12] Polsky, F., Edgell, M. H., Seidman, J. G. and Leder, P. (1978) Anal. Biochem. 87, 397–410.
[13] Lacy, E., Hardison, R. C., Quon, D. and Maniatis, T. (1979) Cell 18, 1273–1283.
[14] Southern, E. M. (1975) J. Mol. Biol. 98, 503–517.
[15] Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[16] Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853–4857.
[17] Goldberg, M. (1979) PhD Thesis, Stanford University.
[18] Chang, A. C. Y., Cochet, M. and Cohen, S. N. (1980) Proc. Natl. Acad. Sci. USA 77, 4890–4894.

[19] Numa, S. and Nakanishi, S. (1980) Biochem. Soc. Trans. 8, 749–751.
[20] Seidah, N. G. and Chrétien, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4236–4240.
[21] Li, C. H., Chung, D. and Yamashiro, D. (1980) Proc. Natl. Acad. Sci. USA 77, 7214–7217.
[22] Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555–566.
[23] Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980) Cell 21, 653–668.
[24] Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980) Nature 283, 220–224.
[25] Rogers, J. and Wall, R. (198) Proc. Natl. Acad. Sci. USA 77, 1877–1879.