

Structural features of the acetyl-CoA carboxylase gene: Mechanisms for the generation of mRNAs with 5' end heterogeneity

XIAOCHUN LUO, KEERANG PARK, FERNANDO LOPEZ-CASILLAS, AND KI-HAN KIM*

Department of Biochemistry, Purdue University, West Lafayette, IN 47907

Communicated by P. K. Stumpf, March 7, 1989

ABSTRACT Acetyl-CoA carboxylase [acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2] is the rate-limiting enzyme in the biogenesis of long-chain fatty acids. We have previously characterized five acetyl-CoA carboxylase mRNA species that differ in their 5' untranslated regions but not in the coding region. We have now characterized the exon-intron structure of the genomic DNA that encodes the 5' untranslated region of the mRNA. Generation of different forms of the mRNA is the result of the selective use of two promoters and differential splicing of five different exons. These five exons contain a total of 645 nucleotides and they are scattered over a 50-kilobase-pair genomic DNA region that we have characterized.

Acetyl-CoA carboxylase [ACC; acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2], the enzyme that catalyzes the carboxylation of acetyl CoA to malonyl CoA, is the rate-limiting enzyme for long-chain fatty acid synthesis (1, 2). Long-chain fatty acids are integral components of cellular membranes and energy storage in all organisms. Thus, under various physiological conditions, a variety of regulatory mechanisms are required for the precise regulation of ACC activity and of fatty acid synthesis. Although extensive studies have elucidated how allosteric controls (1) and covalent modification of the enzyme (3) modulate the short-term regulation of carboxylase activity, virtually no information is available on how this enzyme is regulated at the level of the gene. Under different nutritional and hormonal lipogenic conditions (4–8), ACC induction is accompanied by an increase in the amount of ACC mRNA (9). This increase may be due to an increase in the rate of transcription (10). We have previously determined that there is only one copy of the ACC gene per haploid chromosome set (11). How this gene is regulated to meet the changing and specific metabolic needs for fatty acids is an intriguing question.

To understand the mechanisms involved in the expression of the ACC gene, cDNA clones from rat mammary gland ACC mRNA were obtained (12) and the coding region of the enzyme was sequenced (11). ACC mRNA is ≈ 10 kilobases (kb) long, with an open reading frame of 7035 bases encoding a protein of 2345 amino acids (11).

Our studies of the structure of the 5' untranslated region of the ACC mRNA revealed several species of the mRNA that share the same coding region but differ in their 5' untranslated regions (13, 14). We have characterized three ACC mRNA species in mammary glands—namely, the FL56, the FL63, and the hypothetical species (13). The FL56 and the FL63 species differ at their 5' untranslated end by the presence (in the FL63 species) of a 61-base insert located between the 5' G+C-rich leader segment and the coding region. The analysis of hepatic and adipose tissue mRNAs showed additional forms of ACC mRNA, the pAU and band B species, whose first 242 nucleotides replace the 5' end leader of the FL56 and

FL63 species (14). These studies strongly suggested the possibility that the ACC mRNA variants are generated by the differential splicing of at least two kinds of ACC transcripts and that transcription may be physiologically regulated by the use of tissue-specific promoters of the ACC gene (14).

To understand the origin of these heterogeneous ACC mRNAs and to determine the mechanisms of ACC gene expression, we have isolated and characterized rat genomic DNA clones spanning the 5' end of the ACC gene.† We report here the exon-intron organization of that portion of the ACC gene and evidence that alternative initiation and differential splicing of the ACC gene transcripts is the cause of multiple forms of ACC mRNA.

MATERIALS AND METHODS

Materials. Enzymes and chemicals were purchased from the following suppliers: T4 DNA polynucleotide kinase from United States Biochemical; Klenow fragment, calf intestinal alkaline phosphatase, and dideoxynucleoside triphosphates from Boehringer Mannheim; T4 DNA ligase and restriction endonucleases from Bethesda Research Laboratories; S1 nuclease, T7 RNA polymerase, RNase-free DNase I, deoxynucleoside triphosphates, and ribonucleoside triphosphates from Pharmacia; T4 DNA polymerase from International Biotechnologies; RNasin and pGEM vectors from Promega Biotec; Giga Pack systems from Stratagene; [γ - 32 P]ATP (6000 Ci/nmol; 1 Ci = 37 GBq) from New England Nuclear; [α - 32 P]UTP (3000 Ci/nmol) from ICN; dATP[α - 35 S] from Amersham. All other chemicals were reagent grade.

Methods. Construction of genomic library. High molecular weight genomic DNA (60 μ g) purified from rat liver was digested with *Mbo* I (3 units) in a total vol of 600 μ l. Aliquots (100 μ l) were removed and EDTA was added (8 μ l, 0.5 M) to stop the reaction at intervals of 10, 15, 25, 35, 45, and 55 min after initiation. All the aliquots were then pooled and subjected to fractionation by 0.35% agarose gel electrophoresis at 4°C for 24 hr. The fragments with a size of 15–30 kilobase pairs (kbp) were eluted with an electroeluter (IBI). DNA fragments (1 μ g) were ligated at 15°C overnight with EMBL3 vector (2 μ g) that had been digested with both *Eco*RI and *Bam*HI. The ligation products were packaged *in vitro* according to the procedure described by Stratagene. The original genomic library of $\approx 1.3 \times 10^6$ recombinants was directly screened without amplification.

Probe labeling. cDNA fragments used to screen the genomic library were subcloned into either pGEM3 or pGEM4 plasmids, and 32 P-labeled RNA probes were obtained by transcribing the linearized subclones *in vitro*.

Genomic library screening. For screening, the genomic library was plated at a density of $\approx 3 \times 10^4$ plaques per

Abbreviations: ACC, acetyl-CoA carboxylase; CAT, chloramphenicol acetyltransferase.

*To whom reprint requests should be addressed.

†The sequences reported in this paper are being deposited in the EMBL/GenBank data base (accession nos. M23734, M23735, M23736, M23737, and M23738).

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.

150-mm plate. Nitrocellulose filter lifts were prepared from the plates as described (15). The filters were prehybridized for 4 hr in a solution containing 5× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate), 50% formamide, 5× Denhardt's solution (1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone), 250 μg of tRNA, and 1 mM EDTA at 50°C. The ³²P-labeled probe was added and the hybridization continued for at least 15 more hr at 50°C. Nitrocellulose filters were then washed two times at room temperature with 1× SSC containing 0.1% SDS, followed by another two washes in 0.1× SSC containing 0.1% SDS at 68°C. Filters were air dried and subjected to autoradiography. The positive clones were plaque purified.

Restriction mapping of the genomic clones. λ phage DNA (15) (≈1 μg) was digested with a single restriction enzyme and precipitated with ethanol using glycogen (25 μg) as a carrier. After washing with 75% ethanol, the DNA in the pellets was labeled with [γ -³²P]ATP and T4 polynucleotide kinase as described by Taylor *et al.* (16). The end-labeled DNA fragments were precipitated with ethanol and resuspended in 50 μl of water. Aliquots (10 μl) were subjected to the second restriction enzyme digestion. After digestion, aliquots of the digestion mixture (3 μl, ≈1.5 × 10⁶ cpm) were fractionated by 0.8% agarose gel electrophoresis. The gel was dried between stacks of paper towels and then subjected to autoradiography. A restriction map of the genomic DNA was obtained by comparison of the digestion pattern of single restriction enzyme with that of double enzymes (15).

RNA preparations. Total RNA was prepared from the mammary glands of lactating rats on days 5–7 postpartum, using a slight modification (17) of the method described by Cathala *et al.* (18).

DNA sequencing. DNA fragments were routinely subcloned into pUC19 and then into M13mp19 or M13mp18 (15). DNA was sequenced in both orientations by the dideoxynucleotide chain-termination method (19) using original M13 subclones and their deletion derivatives generated according to the method of Dale *et al.* (20).

Expression plasmid construction and transfection. A 1147-bp fragment flanking the 5' end of exon 1 (including 136 bp of

the exon) and a 1052-bp fragment flanking the 5' end of exon 2 (including 57 bp of the exon) were inserted in front of the chloramphenical acetyltransferase (CAT) gene of the pCAT-B' plasmid (21). The resulting plasmids, P1-CAT and P2-CAT, were introduced into 30A-5 adipocyte (22) by the calcium phosphate precipitation method (23). Expression of the plasmids was examined by assaying CAT enzyme activity (23).

RESULTS

Isolation and Restriction Map of the 5' End of the Rat ACC Gene. To obtain genomic DNA containing the exons that contribute to the 5' ends of the diverse forms of ACC mRNA, we used the EMBL3 vector to construct a rat genomic DNA library. The strategy used to isolate those clones was to "walk" the library starting at the 5' end of the ACC mRNA coding region. To orient the walk toward the 5' end of the gene and identify the exon sequences present in the clones, we probed all positive genomic clones with cDNA sequences belonging to the 5' untranslated regions of the different ACC mRNA species (13, 14). Definitive assignment of exon sequences was accomplished by DNA sequencing of the restriction fragments identified by the cDNA probes.

A series of overlapping clones was obtained and assembled into a coherent physical unit based on the information obtained from restriction mapping and DNA sequencing of the overlapping regions. Four clones (G-3, EM-5, EM-13, and EM-14) that span 52.5 kbp of genomic DNA were shown to contain the five exons that constitute the diverse segments in the multiple species of ACC mRNA (Fig. 1). Fig. 1 also shows the position of the identified exons and the probes (P1, P2, and P3) used in the walk.

Probe 1, which includes the first 200 bases of the coding region, detected the clone G-3, and the 3-kbp *Eco*RI fragment of this clone contained the previously identified exon H2 (13) (exon 5 in Fig. 1). DNA sequencing confirmed the presence of exon H2 in the *Eco*RI fragment. This exon is 250 nucleotides long and contains the first 224 bases of the coding region as well as 26 bases of the upstream noncoding region.

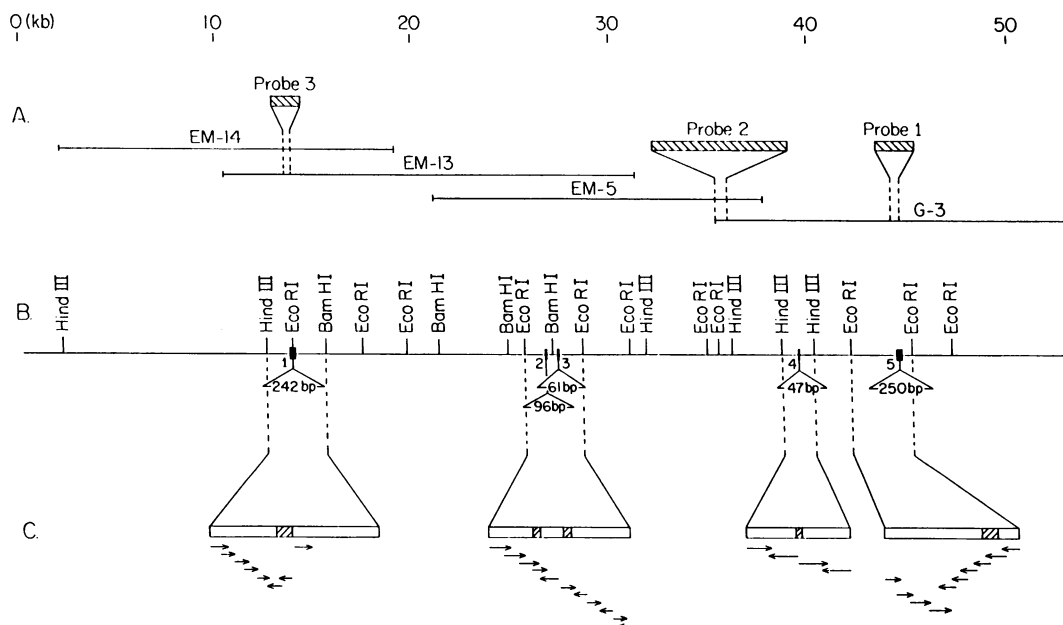


FIG. 1. Isolation and restriction map of genomic DNA around the 5' untranslated region of the ACC gene. (A) Below the scale, the relative positions of clones isolated from the genomic library are shown. The probes used to identify these clones are designated probe 1, probe 2, and probe 3. (B) The restriction map of genomic DNA around the 5' untranslated region is shown. The positions of five exons are indicated as 1, 2, 3, 4, and 5. The length of each exon is specified below the exons. (C) The restriction fragments containing exons have been subcloned into an M13 vector. The sequencing strategy to determine the exon/intron boundaries and the flanking region of each exon is shown. The positions of exons are indicated by the hatched boxes. Arrows depict the direction and extent of dideoxynucleotide sequencing reactions.

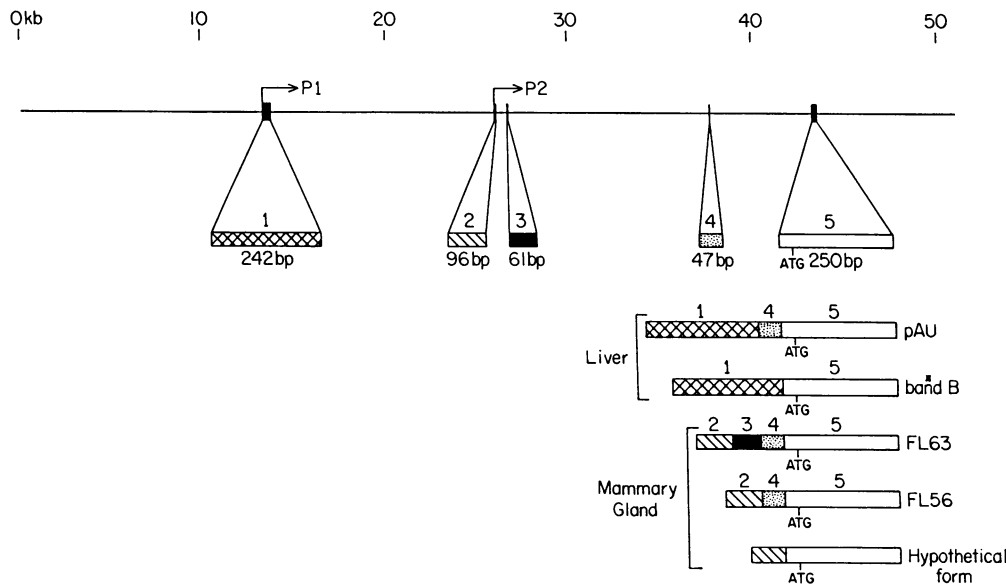


FIG. 3. Organization of genomic DNA at the 5' untranslated region of the ACC gene and the alternatively spliced transcripts. The scale of exons is enlarged and the size of each exon is shown below. pAU is the major form in both rat liver and epididymal adipose tissue. The FL56 is the major form in mammary gland and cultured adipocytes (30A-5). P1 and P2 indicate two transcription initiation sites for promoters 1 and 2 for the ACC gene.

in the figure. The nucleotide sequence of the ACC mRNA at the nontranslated 5' end and the corresponding sequences of the genomic DNA are aligned in Fig. 2. Exons 1-4 belong to the 5' untranslated region of ACC mRNA; exon 5 contains a 224-bp coding sequence including the translation initiation codon AUG. The DNA sequences of all exon/intron junctions follow the GT-AG rule (24).

Comparisons of the nucleotide sequences of the five different species of ACC mRNA we have described (13, 14) with those of genomic DNA suggest the following mechanism for generating heterogeneity in the mRNA. As shown in Fig. 3, band B (13) is generated by splicing exon 1 directly to exon 5 and skipping exons 2, 3, and 4. For the production of FL56 mRNA, which is the main species in the lactating rat mammary gland, exons 2, 4, and 5 join together without 1 and 3. The FL63 species has exactly the same splicing order as FL56, except that exon 3 is inserted between exons 2 and 4. A minor form detected in the mammary gland (13) results from the joining of exon 2 directly to exon 5. In contrast, pAU, the predominant form in rat liver, does not splice in

exons 2 and 3, but exon 4 directly splices to exon 1 (Fig. 3). The fact that pAU types of ACC mRNA are not found in the rat mammary gland (13, 14) indicates that the upstream promoter P1 remains inactive while this tissue is in its lactation-induced lipogenic state.

The ACC Gene Is Transcribed from Two Promoters. Based on our studies of the 5' untranslated regions of the different ACC mRNAs, we have proposed that the ACC gene is alternatively transcribed from two promoters (13, 14). The structural features of the regions upstream of exons 1 and 2 seem to fit this hypothesis—i.e., that the FL56 and the FL63 ACC mRNA are transcribed starting at exon 2, and the pAU ACC mRNA is transcribed starting at exon 1. If this hypothesis is correct, the 5' flanking regions of such exons must exhibit promoter activity. To test this hypothesis, we have constructed plasmids that put the CAT gene under transcriptional control of the regions upstream of exons 1 and 2 (Fig. 4A). Plasmid P1-CAT has 1147 of the 5' flanking region of exon 1, and plasmid P2-CAT has 1052 bp of the 5' flanking region of exon 2, in front of the region encoding CAT; both

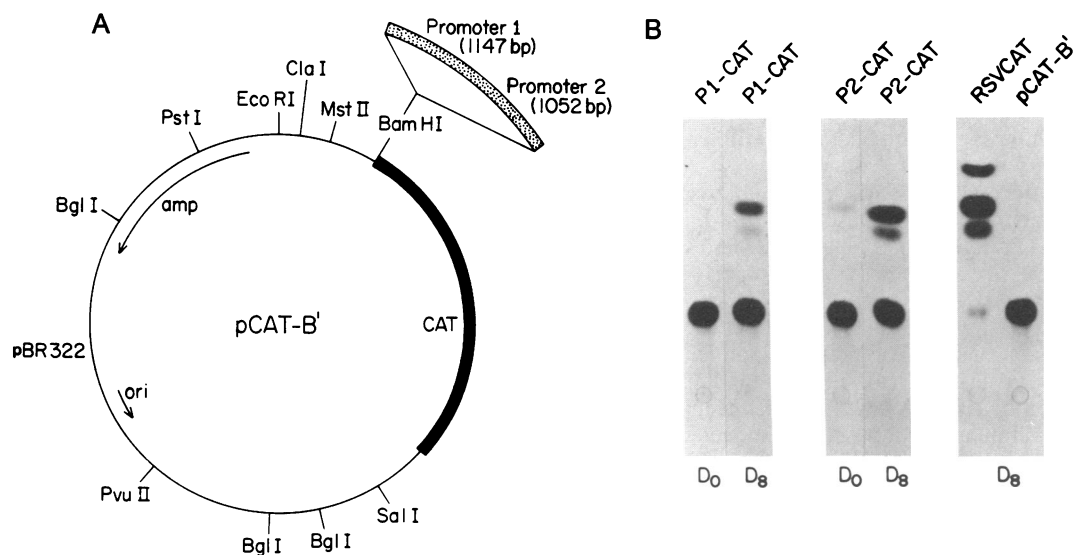


FIG. 4. The promoters of the ACC gene. (A) Construction of ACC promoter 1-CAT (P1-CAT) and ACC promoter 2-CAT (P2-CAT). (B) ACC promoter 1-CAT and ACC promoter 2-CAT constructs are transfected into 30A-5 cells at two different developmental stages (day 0 and day 8). Transient expression of CAT activity from transfected cell extracts was assayed as described (23). The pRSV-CAT served as positive and pCAT-B', which has no promoter sequence, served as negative controls. Nonacetylated chloramphenicol migrates slowest, as shown in lane P1-CAT at day 0 (D₀) and pCAT-B' at day 8 (D₈). The middle two bands (P1-CAT at day 8, P2-CAT at day 8, and RSV-CAT at day 8) are two different forms of monoacetylated chloramphenicol. The furthest moving band is the diacetylated product (RSVCAT).

plasmids contained the putative promoter DNA fragments in the correct 5' → 3' orientation. These plasmids were transfected into the 30A-5 preadipocyte cell line (22) at day 0 and day 8 of differentiation. The transient expression of CAT activity was determined to assess the strength of the promoter activity of the inserted fragments. As shown in Fig. 4B, these fragments, particularly the sequence flanking the exon 2, show strong transcriptional promoter activity. Two additional observations may be significant: (i) the liver-specific promoter can function in the fat cells, and (ii) in their undifferentiated state, 30A-5 cells show very little CAT activity, suggesting that promoter activity of these sequences is differentiation dependent.

Detailed analysis of these sequences as promoters remains to be done. The present results clearly show that the generation of multiple forms of ACC mRNA at the 5' untranslated region is due to differential splicing and alternative usage of the promoters.

DISCUSSION

Our previous studies of the 5' end of ACC mRNA have shown that the corresponding region of the gene must be complex (13, 14). The exon-intron arrangement of the unusually long 5' end of the ACC gene provides a framework that accounts for the occurrence of five different ACC mRNA species we have characterized (13, 14), as shown in Fig. 3. It is clear from the structure of the gene that both the alternative use of the two promoters and the differential splicing of the resulting transcripts must be involved in the generation of the multiple ACC mRNAs. One remarkable observation is that exons 1–5 represent only 646 bases of ACC mRNA and are encoded in the 50 kbp of genomic DNA that we have characterized. The total length of ACC mRNA is ≈10 kbp. Although the intron-exon structure of the coding region has yet to be examined, this information provides some indication of the enormity of the ACC gene size and its complexity.

The ACC gene exhibits a structural arrangement similar to the mouse and rat *Amy-1* gene (25) in which two promoters generate tissue-specific mRNAs that differ solely at their 5' untranslated regions. In cases in which promoter switching does not affect the coding region, but generates different 5' leaders, the translational efficiency of the mRNAs can be affected. Kozak (26) has discussed how the alternative transcription of many protooncogenes results in the production of two kinds of transcripts, one with a short simple 5' nontranslated region and another with a longer G+C-rich leader. The longer leader impairs the translational efficiency of the mRNA.

In the case of the ACC gene, the region upstream to exon 1 exhibits all the typical structural features that define a promoter region (TATA box at position –25 and CAAT box at position –61). We have previously reported that the synthesis of the pAU liver-specific mRNA was enhanced upon a diet-induced lipogenesis (14). On the other hand, the sequence upstream of exon 2 contains a high frequency of CpG and numerous sites for Sp1 nuclear binding proteins, which are frequently associated with "housekeeping genes" (27, 28). Further refinement of the specific sequences that are responsible for the promoter activities of both regions remains to be done.

The requirement for a stable secondary structure (larger than 50 kcal/mol; 1 cal = 4.184 J) in the 5' end of the mRNA plays a role in the translational efficiency of the mRNA (26); however, in some cases, this secondary structure seems to be dispensable, and specific interactions between protein cytoplasmic factors and specific sequences in the mRNA seem to have a commanding role in translational efficiency (29).

Although none of the many sequences exhibited by ACC mRNA at its 5' end exhibits a stable secondary structure, the profusion of 5' untranslated leaders in ACC mRNA strongly suggests that such variants must play a role in the translation efficiency of their mRNAs. This would add translational control to the long list of mechanisms that regulate lipogenesis through the control of ACC.

Finally, it should be noted that the fat-specific element, FSE1 and FSE2 (30), sequences are not found in the two promoters we have examined. However, the occurrence of such sequences in the regions we have not examined cannot be precluded at this time.

This research was supported by Grant DK12865 from the National Institutes of Health. F.L.-C. was a recipient of a postdoctoral fellowship from the American Heart Association, Indiana Affiliate. This is paper no. 11893 from the Purdue University Agricultural Experiment Station.

1. Lane, M. D., Moss, J. & Palakis, S. E. (1974) *Curr. Top. Cell. Regul.* **8**, 139–195.
2. Volpe, J. J. & Vagelos, P. R. (1973) *Annu. Rev. Biochem.* **42**, 21–60.
3. Kim, K.-H. (1983) *Curr. Top. Cell. Regul.* **22**, 143–176.
4. Majerus, P. W. & Kilburn, E. (1969) *J. Biol. Chem.* **244**, 6254–6262.
5. Nakanishi, S. & Numa, S. (1970) *Eur. J. Biochem.* **16**, 161–173.
6. Giffhorn, S. & Katz, N. R. (1984) *Biochem. J.* **221**, 343–350.
7. Spence, J. T. & Pitot, H. C. (1982) *Eur. J. Biochem.* **128**, 15–20.
8. Mackall, J. C. & Lane, M. D. (1979) *Biochem. J.* **162**, 635–642.
9. Pape, M. E., Lopez-Casillas, F. & Kim, K.-H. (1988) *Arch. Biochem. Biophys.* **267**, 104–109.
10. Pape, M. E. & Kim, K. H. (1989) *Mol. Cell. Biol.* **9**, 974–982.
11. Lopez-Casillas, F., Bai, D.-H., Luo, X. C., Kong, I.-S., Hermodson, M. A. & Kim, K.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5784–5788.
12. Bai, D.-H., Pape, M. E., Lopez-Casillas, F., Luo, X. C., Dixon, J. E. & Kim, K.-H. (1986) *J. Biol. Chem.* **261**, 12395–12399.
13. Lopez-Casillas, F., Luo, X. C., Kong, I. S. & Kim, K.-H. (1989) *Gene*, in press.
14. Lopez-Casillas, F. & Kim, K.-H. (1989) *J. Biol. Chem.*, in press.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Taylor, W. L., Collier, K. J., Deschenes, R. J., Weith, H. L. & Dixon, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6694–6698.
17. Lopez-Casillas, F., Pape, M. E., Bai, D.-H., Kuhn, D. N., Dixon, J. E. & Kim, K.-H. (1987) *Arch. Biochem. Biophys.* **257**, 63–68.
18. Cathala, G., Savouret, J.-F., Mendez, B., West, B. I., Karin, M., Mantial, J. A. & Baxter, J. D. (1983) *DNA* **2**, 329–335.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
21. Andrisani, O. M., Hays, T. E., Roos, B. & Dixon, J. E. (1987) *Nucleic Acids Res.* **15**, 5715–5728.
22. Pape, M. E. & Kim, K.-H. (1988) *Mol. Endocrinol.* **2**, 395–403.
23. Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1051–1066.
24. Green, M. R. (1986) *Annu. Rev. Genet.* **20**, 671–708.
25. Schibler, U. & Sierra, F. (1987) *Annu. Rev. Genet.* **21**, 143–176.
26. Kozak, M. (1988) *J. Cell Biol.* **107**, 1–7.
27. Chretien, S., Dubart, N., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goosens, M. & Romeo, P.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6–10.
28. Obaru, K., Tsuzuki, T., Setoyama, C. & Shimada, K. (1988) *J. Mol. Biol.* **200**, 13–22.
29. Leibold, E. A. & Munro, H. N. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2171–2175.
30. Spiegelman, B. M. (1988) *Trend Genet.* **4**, 203–209.