BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **231,** 106–110 (1997)
ARTICLE NO. RC966050

The Genomic Structure of a Mouse Seminal Vesicle Autoantigen[1]

Lung-Chih Yu,* Ya-Ling Hsiao,† Yun-Hsin Yang,† Marie Lin,* and Yee-Hsiung Chen†,‡,[2]

**Department of Medical Research, Mackay Memorial Hospital, †Institute of Biochemical Science, College of Science, National Taiwan University, and ‡Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan*

Received December 2, 1996

The genomic structure of an androgen-stimulated mouse seminal vesicle autoantigen was determined. Analysis of the nucleotide sequence established 2135 bp of the 5′-flanking region, four exons of 123, 136, 112, 227 bp, three introns of 1555, 1931, 316 bp, and 185 bp of the 3′-flanking region of this gene. Ten DNA segments, five in the 5′-flanking region, two in the first intron, and three in the second intron were identified to have more than 50% homology with the consensus sequence of the androgen response element (ARE). Two sets of adjacent DNA segments, one including −213 to −199 bp and −124 to −110 bp in the 5′-flanking region and the other including 2532 to 2546 bp and 2582 to 2596 bp in the second intron, are noticeable for their high degree of homology with ARE.

The seminal vesicles of adult male mammals secrete a group of products that constitute a major portion of seminal plasma. The primary structure for some of the major protein components in mouse and rat seminal vesicle secretions have been established (1-3), but still no significant clues to their function could be inferred from their protein structures, despite that their gene expression is shown to be androgen-dependent. Recently, we have identified an androgen-stimulated autoantigen (SVA) in mouse seminal vesicle (4). It is a zinc-binding 19-kDa glycoprotein with a core protein consisting of 131 amino acid residues (4,5).

Androgenic action is considered to be mediated through the interaction between a hormonal receptor complex and the androgen response element (ARE) in the responsive gene (6). In spite of the presence of androgen receptor in many sexual organs, the expression of SVA occurs exclusively in seminal vesicle (4). Thus, establishing the genomic structure of SVA becomes a prerequisite toward a better understanding of how the androgen regulation for the transcription of this gene in seminal vesicle. This work was carried out in this direction. Here, we report the genomic structure of the SVA gene and the identification of ARE-like sequences in the SVA gene.

METHODS

Genomic Cloning and Sequencing of the SVA Gene

A genomic library of mouse B6/CBAF1J spleen DNA in lambda FIX II vector (Stratagene, La Jolla, CA, U.S.A.) was screened. The hybridization probe was randomly primed from the full length cDNA of SVA (4). A positive lambda phage clone containing a 18 kb genomic DNA insert was isolated and denoted as 11S. The 18 kb DNA insert was digested with *Sac*I restriction enzyme and then analyzed by southern blotting technique using the same random-primed probe. Two DNA fragments of 4.2 and 5.5 kb were hybridized with the probe. The 4.2 kb fragment was further cleavaged into 2.7 and 1.5 kb fragments by *Eco*RI digestion. Only the 2.7 kb fragment bound to the hybridization probe in the southern blotting analysis. The 2.7 kb and 5.5 kb fragments were subcloned in pGEM 7Zf(−) vector (Promega, Madison, WI, U.S.A.).

DNA sequences were determined by the dideoxynucleotide chain termination method using an AmpliCycle Sequencing Kit (Perkin Elmer, Foster City, CA, U.S.A.).

Primer Extension Analysis

A synthetic oligonucleotide corresponding to the antisense sequence in the first exon (nucleotides +91 to +114) of the SVA gene was used in primer extension analysis. Total RNA from seminal vesicles of adult mice (10 weeks old) was isolated by guanidinium thiocyanate/cesium chloride cushion method (7). Twelve ng of the primer was hybridized to 6 μg of total RNA in reverse transcription buffer (50 mM NaCl, 34 mM Tris-HCl and 5 mM DTT, pH 8.3) at 85°C for 10 min, followed by 45°C for 12 h. dGTP, dTTP and dCTP each at a final concentration of 0.25 mM and 8 μCi of [α-^{32}P]dATP (2.7 pmole) were added to the annealing mixture and incubated

[1] The nucleotide sequence reported in this paper has been registrated in the GenBank/EMBL Data Bank (Accession number: L44117).

[2] To whom correspondence should be addressed: Institute of Biochemical Science, College of Science, National Taiwan University. P.O. Box 23-106, Taipei, Taiwan. Fax: 886-2-363-5038. E-mail: BC304@gate.Sinica.edw.tw.

0006-291X/97 $25.00

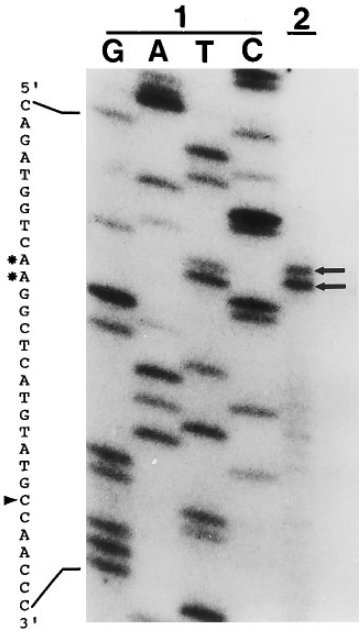


FIG. 1. Primer extention analysis of the transcription initiation sites of the SVA gene. The two major products indicated by arrows in lane 2 were determined from primer extension analysis using total RNA isolated from seminal vesicle of adult mice as template and oligonucleotide corresponding to the antisense sequence in the first exon as primer (see Methods for details). The sequence shown on the left side was read from the sequencing ladders (columns G, A, T, and C in lanes 1) of the 11S lambda DNA using the same oligonucleotide as primer. The nucleotides corresponding to the extented products, at positions −28 and −27 upstream from the ATG start codon, are marked by "*".

with 3 units of reverse transcriptase from Avian Myeloblastosis Virus (Promega) in a final volume of 6 μl at 42°C for 15 min. One μl of dNTP mixture (2 mM each) was then added to the reaction solution and chased for another 15 min at 42°C. The reaction was stopped by adding 5 μl of 98% formamide containing 10 mM EDTA, 0.3% xylene cyanole and 0.3% bromophenol blue.

The lambda DNA from 11S clone was served as template for dideoxynucleotide sequencing using the same oligonucleotide primer. The sequencing reaction products were run parallel with the primer-extended fragments on a 6% poly-acrylamide/urea sequencing gel.

RESULTS AND DISCUSSION

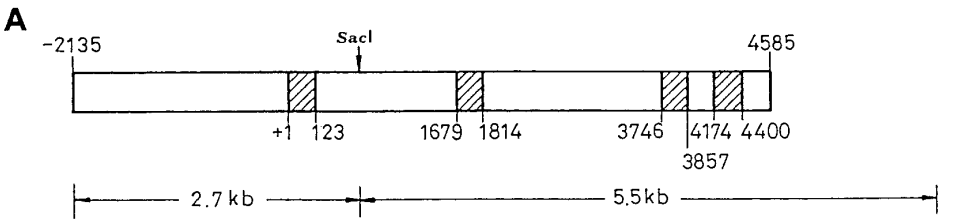
The transcription initiation site of the SVA gene was determined by primer extension analysis. The two major products were corresponding to the nucleotides at 27 and 28 bp upstream from the ATG start codon (Fig. 1), revealing the presence of two successive transcription initiation sites in the SVA gene.

Through subcloning and sequencing of the 2.7 and 5.5 kb DNA fragment (see Methods), the genomic structure of SVA was established. We determined the complete sequence of 2720 bp from the 2.7 kb fragment and the partial sequence of 4000 bp from the 5.5 kb fragment. Alignment of these sequences with that of SVA cDNA (4) supports the existence of four exons and three introns in the SVA gene (Fig. 2A). The 2.7 kb fragment covers the 5′-flanking region up to −2135 bp from the transcription initiation site, the first exon and one part of the first intron (462 bp). The partial sequence of the 5.5 kb fragment extends from the other part of the first intron to the 3′-flanking region of 185 bp. Subcloning and sequencing of a DNA fragment comprising the whole first intron amplified by polymerase chain reaction using the lambda 11S DNA as the template confirmed the linkage of *Sac*I site between the 2.7 and 5.5 kb fragments in the first intron (Fig. 2A).

The established genomic sequence of SVA gene is shown in Fig. 2B. The four exons of 123, 136, 112 and 227 bp are separated by three introns of 1555, 1931 and 316 bp. All of the exon/intron boundaries conform to the GT-AG consensus rule. A noncanonical TATA box appears at 31 bp upstream of the transcription initiation site and no CAAT box element is present within the −100 bp of 5′-flanking region. Repeats of nucleotides with purine and pyrimidine bases are present around −800 to −600 bp and −1350 to −1100 bp regions, respectively.

Functional AREs have been identified at several regions of the andron-responsive genes. Many of them are within −200 to −100 bp of the 5′-flanking region of the corresponding genes such as those encoding for the prostate-specific antigen (7), human glandular kallikrien-1 (8), rat probasin (9) and mouse vas deferens protein (10). However, functional AREs at different locations of the androgen-responsive genes have been identified also. For instance, the AREs of the genes of mouse sex-limited protein (Slp) (11) and tyrosine aminotransferase (12) locate respectively in the regions around 1.9 and 2.5 kb upstream of the transcription initiation sites, and the genes of prostatein C3 (13) and rat 20 kDa protein (14) have functional AREs in their first introns. Ten ARE-like sequences were identified in SVA gene from selection for the half-site of the consensus ARE palindrome, AGAACAnnnTGTTCT (Table 1). These potential AREs spread at several regions of SVA gene. Two sets of adjacent ARE-like sequences, one includes −213 to −199 bp and −124 to −110 bp in the 5′-flanking region and the other includes 2532 to 2546 bp and 2582 to 2596 bp in the second intron, are noticeable to have high degree of homology with the ARE consensus sequence. Further studies of nuclear protein-DNA binding analyses and the assays of reporter gene fusion plasmid are necessary to determine which one of these ARE-like sequences participates in the androgen-mediating transcription activity of SVA gene.

Many lines of evidence indicate the necessity of the cooperative interaction between hormone response



B

```
                                                GAATTCTTAATTCTGAACATCTATGCTCCAAATGCAAAGGCCCCCATATTCATAAACTTTACTAA
-2070 AGCTCAAAGCCTACATTGCACCATAGACAATAACAGTGAGAGACTTGACTTAAACACCACACTCTTAGCAATGAACAGTTCATGGAAACA
-1980 GAAACTAAGCAGAGACACAGTGAAACTTAACCGAAGTGGTGGACCAAGTGGATTTAACAGATATCTACAGAACATTTCATCCTAAAACAA
-1890 AAGAATATACCTTCTTCTCAGCATCTCACTGTACCTTCTCCAAAACTGATTATATAATAGGTCACAAAACAGGCCTCAACAGATACAAGA
-1800 AGATTGAAACAATCCCATGCATCCTATCAGAGTCACCACAGACTAAAGCTGCTCTTCAATAACAATATAAACAACAGAAAGCCCACATAC
-1710 ACATGGAAGCTGAACAATGCCCTACTCAGTGATAACTTGGTCAAGGAAGAAATAAATAAAGAAATTAAAGACTATTTTTGATTTAATTAA
-1620 CATGAAGGCCCCAGCATACCCAAACTTATGGGACACAATAAAAGCAGTGCTAAGAGGAAAATTCATAGCTCAGCTGCAAAAGAAACTGGA
-1530 AAGAGCATACTATAGCAGCTTAACAGCATACCTGAAAGCTCTAGAACAAAGAGAAGCAAATACACCCAATAGGAGTAGATGGCAGGAAAT
-1440 AAACTCAGGGCTGAAATCAACCAAGTAGAAACAAAAGGTACTATACAAAGAATCAACAAAACCAGGACCTGGTTCTTTGGTTCTTTGAGA
-1350 AAATCTTCCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTTCTTCCTTCCT
-1260 TCCTTCCTTCCTCCCTCCAGTACATAGCCATCCATCCATCCATCCTTCCTTCCTTCCTTCCTTCCTTCTTTCCTCCCTCCTTCCCTCTCT
-1170 CCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTGTCCTTTTTTGTTTTTTATCAAATACAC
-1080 TCTAACAGGAGTCAAACTGAAAAAATAAACTTCAATTGAAAAAAGGTTTCCATCATATTGTCTCGTAGGCAAAAGATCATTGAGTATTTT
 -990 CTTGACTGATGATTTACTATTGCAATCGCCATCCTGAACAGCTGGAATAAGAAATCGGTCTGCACAAGATGATTATTCTATTTCTACATG
 -900 ACCTCTTCAGTGACTGCCTCCAAAAGTCTCTGTTTGATTTCCTGTTCTGACTACCATTCATGATGTGCTGTAAGCTATAAGATGAAATAA
 -810 ACTCTTTCCTCTGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAAGGAAGAAAGGAAGGAAGG
 -720 AAGGAAGGAAGGAAGGGAGAGAGAGAGAAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA
 -630 GAGAGAGAGAGAGAGAGAGAAAGAAGAAAAAAGTGAAAAAAACAAGAGATGTATAATAATACAAAATAAAATATAATAGATAACACACAA
 -540 AATAATTCTAGTTTCGGGGCTGGAAAATTGGTTTAGGAATCAAGATCACCTACTGCTCCTCCAAGGGCCCTAGATTCAGTTCCTAGCACA
 -450 AACATGTTGGTTCACAACAGCCTGTAACTCCAGGGTATCTGATGTTTTCTTCTGGCTTCCTTGGGCATAGACATGCATGTGATATACAAA
 -360 AAGACAAATGTGTGGGCATGTGTGTGTGTACAAATAATAACATGTAAGAAATAGATATATCTACAGTTAGAGTTGTTTAAGAAGCAAAGG
 -270 ACAAAAAGATCAGATCTTCTTTAGAAAAATTGGAACTGGATCCATTCTAATAATCTTAGAACATTCTAATCTCTTATGGAGAAATCTGAT
 -180 CTTATTCACTATAGTCCATTTTGAAGTTTCATCTGAAAACTGACCTTTATGTACAAAGAACAAAGAGTGTGAAACTGGAAGGGATCCCAA
  -90 CATCTGATCATCTTAGCTGGATAGACCTGCCCATCTACTTCACTGTCTTCCTGCTGCAAATAAAAGGAGAGACCTGGTTTGCAGATGGTC

   +1 AAGGCTCATGTATGCCAACCCATACAGCATGGCTTCTCTTCAAAAGTTCTATCAACCTGGAGCAGTCACATATCTCCTTATTGTATTCTT
                                  MetAlaSerLeuGlnLysPheTyrGlnProGlyAlaValThrTyrLeuLeuIleValPheLe
                                                                  -20

   91 ACTTCAGGAGACTGCTGGGTTTTGGTACCATGGGTAAGGCACTACTTTTTCCTCAGCATCTTCCTAGAAATATTTTTACACTTAACACAG
      uLeuGlnGluThrAlaGlyPheTrpTyrHisGl
                      1

  181 AAACATAATGGATTTGTGCTGTATGTCAATCTGACTTTCTCTTTTCTAAACTCTGTGATGATGTGTCTTCTCAAACCTACATCCTGAAAT
  271 CAAATGACTGGTGACAAAAGATTGTGTTCAAACTTACTTTTCTTCCTAGATTATAGTATCATCACCATGCAATATGAAGATACAAGTGAG
  361 CTTCGCTTATATTTTAGGATAAACTATGGACAGGATATAATTTATCTCTAAGGCTAAACTTGGTGTGGAAGCAAATGATGGCATGTTCAA
  451 GAGTTCCCTCATGATAGAATAAAACATCATGTCACAACTAACCAAAACACAATGAATTAATTAGATATCTACAGGAAAATTATCTAAATT
  541 GTTTAATAAAACCTGATATCTTCATTAATTCTAATATGCAGAGCTCTGTTAACCATACATGGAACAGATTGGTTTAATAAAAAGAGAAGC
  631 CGGGCGTAGTGGCACACGCCTTTAATCCAGCACTCAGGAGGCAGAGGCAGGCGGATTTCTGAGTTCCAGGCCAGCCTGGTGTACAAAGTG
  721 AGTTCCAGGACAGCCAGGGCTGCACAGAGAAACCCTGTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
  811 AGAGAGAATGTCAGAACTTGATGCTGTACCCCAGCCTTTCTTTACTTGTTCCTTTACACAAGATTCCTTGAATAATAAATTTCTTTTCTA
  901 AATCTATAACTAAAGGGGGGAAATGTCTGATTTCAATTCCTATAAAATGAAAATAATATGTAAAGGGTCCTAAGACTTCTGACAAGAAGA
  991 ACATGTTAAGTTACAGGAAGAAATCAGAATCTACTAACAATACCTAAAATCAATAAGGAAAGGGGACTCTTCAAGATGGTAATGAATCCA
 1081 TGGATATTAGTCTGGAAAATTTTTATGTGCCAAGTAAATAATTTTCCATAATTGAAGCAATGGACTGAGCAACTTAGGAAATGATCTTAA
 1171 GGTAATTTGAGTTGGAAATCAAATTTTTTCAACTAGAACTTTGATGCTTCCTATTTCAGACCTAAAGGGCAGTCAGATGAGGAAAAAATA
 1261 TTGAGCTGAATCTGCAGGTACCATGAGACCAGGTTGACATTTATAATGAATTTAGCACTTTTGTGGCTCAATATATTTATACAAGTCCTA
 1351 CTTTTAGAAGGAAGAGAGTTGTGAAGACTTTTTTTGATACTTGTAAAAGAAACATAGGAAGGGGAGTAAGCAAAGATGGACGATGAAGTC
 1441 ATAAATCTCCACTGGGGAATGTTGCCTGGGAAAACACAAAAGCAACTTAACCAGCCCAGGACCCCACAATCTCCAGGTGACCAGGACTGC
 1531 TACTCATTGTCCTGATTTCCCTTTGCTTCATCCTCCTGGAAGTGTCTGTTGAGTAGAAATATGTTCTAACTTTGTGTGTGTGCTCTCT

 1621 CTGTCTTCCCATACGAAAACAACTAAGTCTTAGGCAGTAATGTGAGTTTCTGTTTCAGCTTACACAATGAACCCCGTAACTACACACTCA
                                                              yLeuHisAsnGluProArgAsnTyrThrLeuT

 1711 CACTGAATATGAAAATCACTCCAAGTACTAGCATTACTAACAAAAAAGGGAATGAATTTCGAGTACAACTGACAGTTAAAAACAACGTGA
      hrLeuAsnMetLysIleThrProSerThrSerIleThrAsnLysLysGlyAsnGluPheArgValGlnLeuThrValLysAsnAsnValL
                      20                                                              40

 1801 AGCAATGCCAGGAGGTAAGCGAATAATTAAAAGTTAACATGAAAACTAAAATATTATATTTAAGATAACTCAATAATTTAAATATTGTGC
      ysGlnCysGlnGlu
```

FIG. 2. Nucleotide sequence of the SVA gene. (A) Schematic diagram of the SVA gene. Cross-hatched bars are exons. Open bars are 5′-flanking, introns and 3′-flanking regions. Nucleotide numbers are relative to the first transcription initiation site. (B) The established nucleotide sequence of the SVA gene. The two successive transcription initiation sites are marked by "*" and the first one is numbered +1. Exon regions are typed in bold and coding regions are displayed with deduced amino acid sequences which are numbered according to the mature protein (4). The noncanonical TATA box at position −31 bp is double-underlined. The translation start, stop codons, and two overlapping polyadenylation signals are underlined with solid lines. The consensus sequences between the exon/intron boundaries are underlined with dashed lines. ARE-like sequences are boxed.

elements and other accessory response elements for the strong response and specificity of hormone action (15). Some important accessory response elements associated with the androgenic action have been found. A nuclear factor I (NF-I) binding site in the mouse mammary tumour virus long terminal repeat (16) and the binding sites of the octamer transcription factor 1 and the NF-I-like factor in the prostatein

FIG. 2—*Continued*

C3 gene (17) were demonstrated to modulate the androgen response of the promoters. Accessory elements or other transcription factor binding sites contributing to the regulation of the androgen response are also believed to be present in the genes of the Slp (11) and rat 20-kDa protein (14). The cooperative interaction between ARE and accessory response elements might ensure the expression of androgen-dependent genes in specific tissues. The androgen-stimulated SVA expresses exclusively in the seminal vesicle (4) despite that the androgen receptor exists in many other sexual and non-sexual organs. Regulation of the transcription of the SVA gene might be dependent on a specific machinery including a novel transcription factor(s) in seminal vesicle. Several potential binding sites for various transcription factors were identified in the SVA gene. Within the 5′-flanking region, there are potential binding sites for CACCC-binding protein (position −1468) (18), rodent PitB1 and PitB2 factor (position −582) (19) and GATA factor (position −532) (20). Whether these factors or other elements in the SVA gene participate in the control of the specific expression of SVA awaits future study.

TABLE 1

The ARE-like Sequences in SVA Gene

Position	Nucleotide no.	Sequence	Homology
ARE consensus		AGAACA nnn TGTTCT	100%
5′-Flanking	−1912 ~ −1898	AGAACA ttt caTcCT	75%
	−1488 ~ −1474	AGAACA aag aGaagc	58%
	−867 ~ −853	ttgAtt tcc TGTTCT	58%
	−213 ~ −199	AGAACA ttc TaaTCT	83%
	−124 ~ −110	AGAACA aag aGTgTg	75%
First intron	988 ~ 1002	AGAACA tgt TaagtT	67%
	1583 ~ 1597	ctAgaA ata TGTTCT	67%
Second intron	2532 ~ 2546	AGAACA gtt gGTTCa	83%
	2582 ~ 2596	gtAACt ttc TGTTCT	75%
	3593 ~ 3607	AGAACA ttc TcagCa	67%

ACKNOWLEDGMENT

This work was partially supported by Grant NSC-86-2311-B-001-081 from National Science Council, Taiwan.

REFERENCES

1. Ostrowski, M. C., Kistler, M. K., and Kistler, W. S. (1979) *J. Biol. Chem.* **254,** 383–390.
2. Mansson, P. E., Sugino, A., and Harris, S. E. (1981) *Nucl. Acids Res.* **9,** 935–946.
3. Chen, Y. H., Pentecost, B. T., Mclachlan, J. A., and Teng, C. T. (1987) *Mol. Endocrinol.* **1,** 707–716.
4. Yu, L.-C., Chen, J.-L., Tsai, W.-B., and Chen, Y.-H. (1993) *Biochem. J.* **296,** 571–576.
5. Huang, Y.-H., Luo, C.-W., Yu, L.-C., Chu, S.-K., and Chen, Y.-H. (1995) *Biophys. J.* **69,** 2084–2089.
6. Zhou, Z.-X., Wong, C.-I., Sar, M., and Wilson, E. M. (1994) *Recent Prog. Horm. Res.* **49,** 249–274.
7. Riegman, P. H. J., Vlietstra, R. J., van der Korput, J. A. G. M., Brinkmann, A. O., and Trapman, J. (1991) *Mol. Endocrinol.* **5,** 1921–1930.
8. Murtha, P., Tindall, D. J., and Young, C. Y. F. (1993) *Biochemistry* **32,** 6459–6464.
9. Rennie, P. S., Bruchovsky, N., Leco, K. J., Sheppard, P. C., McQueen, S. A., Cheng, H., Snoek, R., Hamel, A., Bock, M. E., MacDonald, B. S., Nickel, B. E., Chang, C., Liao, S., Cattini, P. A., and Matusik, R. J. (1993) *Mol. Endocrinol.* **7,** 23–36.
10. Fabre, S., Manin, M., Pailhoux, E., Veyssière, G., and Jean, C. (1994) *J. Biol. Chem.* **269,** 5857–5864.
11. Alder, A. J., Scheller, A., Hoffman, Y., and Robins, D. M. (1991) *Mol. Endocrinol.* **5,** 1587–1596.
12. Denison, S. H., Sands, A., and Tindall, D. J. (1989) *Endocrinology* **124,** 1091–1093.
13. Tan, J., Marschke, K. B., Ho, K.-C., Perry, S. T., Wilson, E. M., and French, F. S. (1992) *J. Biol. Chem.* **267,** 4456–4466.
14. Ho, K.-C., Marschke, K. B., Tan, J., Power, S. G. A., Wilson, E. M., French, F. S. (1993) *J. Biol. Chem.* **268,** 27226–27235.
15. Lucas, P. C., and Granner, D. K. (1992) *Annu. Rev. Biochem.* **61,** 1131–1173.
16. Cato, A. C. B., Skroch, P., Weinmann, J., Butkeraitis, P., and Ponta, H. (1988) *EMBO J.* **7,** 1403–1410.
17. Celis, L., Claessens, F., Peeters, B., Heyns, W., Verhoeven, G., and Rombauts, W. (1993) *Mol. Cellular Endocrinol.* **94,** 165–172.
18. Schule, R., Muller, M., Otsuka-Murakami, H., and Reinkawit, R. (1988) *Nature* **332,** 87–90.
19. Chen, R., Ingraham, H. A., Treacy, M. A., Albert, V. R., Wilson, L., and Rosenfeld, M. G. (1990) *Nature* **346,** 583–586.
20. Orkin, S. H. (1990) *Cell* **63,** 665–672.