Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 8444–8448, November 1988
Biochemistry

A gene inducible by serum growth factors encodes a member of the steroid and thyroid hormone receptor superfamily

(transcription factors/DNA-binding proteins/zinc fingers/growth-related genes)

THOMAS G. HAZEL*, DANIEL NATHANS†, AND LESTER F. LAU*‡

*Department of Molecular Biology, Northwestern University Medical School, Chicago, IL 60611; and †Howard Hughes Medical Institute Laboratory and the Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Contributed by Daniel Nathans, August 12, 1988

ABSTRACT **We previously identified, by cDNA cloning, a set of genes that are expressed during the G_0/G_1 transition (cell cycle reentry) in mouse fibroblasts. These immediate early genes are transcriptionally activated within minutes of addition of serum or purified growth factors, and their mRNAs are superinduced in the presence of protein-synthesis inhibitors. We report here that one of these genes, represented by nur/77 cDNA (originally called 3CH77), encodes a member of the superfamily of ligand-binding transcription factors that includes the steroid and thyroid hormone receptors. The nur/77 cDNA sequence encodes a protein of 601 amino acids containing two regions of sequence similarity to members of this nuclear receptor superfamily, corresponding to their DNA-binding and ligand-binding domains. These results suggest that the growth factor-inducible immediate early gene *nur/77* encodes a ligand-binding protein that regulates the genomic response to growth factors.**

The proliferation of animal cells is initiated and regulated by polypeptide growth factors. The interaction of growth factors with their specific receptors generates a cascade of intracellular biochemical events (1), leading to the sequential expression of specific genes. By analogy to the developmental program of viruses, some of the genes activated early by the actions of growth factors are likely to regulate the expression of other genes necessary for the onset of DNA replication (2). We previously identified, by cDNA cloning, a set of genes that are transiently activated within minutes after quiescent BALB/c mouse 3T3 cells are stimulated with serum, platelet-derived growth factor, or fibroblast growth factor (3, 4). These "immediate early" genes are regulated at the transcriptional and posttranscriptional levels, and their mRNAs are superinduced in the presence of protein-synthesis inhibitors (4). In this communication we report the nucleotide sequence of one of these cDNA clones, 3CH77 (hereafter referred to as nur/77), and the amino acid sequence it encodes.§ Sequence comparison shows that the nur/77-encoded protein is related to the nuclear receptor superfamily that includes the receptors for such compounds as glucocorticoids (5), mineralocorticoids (6), estrogen (7), testosterone (8, 9), progesterone (10), thyroid hormone (11, 12), retinoic acid (13–15), and vitamin D (16). These receptors function as modulators of gene expression; upon binding to their respective ligands, they exhibit increased affinity for enhancer-like DNA sequence elements associated with target genes (17). Interaction of the receptor–ligand complex with the DNA sequence elements results in altered expression of these genes (17).

The nur/77 protein and the known nuclear receptors share two regions of sequence similarity. For several receptors the first region has been shown to be the DNA-binding domain (13, 14, 18–20), and the second region the ligand-binding domain (18, 19, 21, 22). These results suggest that nur/77 cDNA encodes a ligand-dependent nuclear receptor; in the presence of its ligand, the nur/77 protein may act as a specific DNA-binding protein that regulates the genomic response to growth factors.

MATERIALS AND METHODS

Cell Culture. BALB/c 3T3 cell clone A31 was cultured as described (4).

cDNA Screening and Sequence Analysis. A cDNA library containing near-full-length copies of immediate early mRNAs (4) was screened with a nur/77 cDNA probe as described (3). Near-full-length nur/77 cDNAs were cloned into pGEM plasmids (Promega Biotec, Madison, WI). Nested 5′ and 3′ deletions were created by using BAL-31 nuclease (Bethesda Research Laboratories) and were sequenced by the dideoxynucleotide chain-termination procedure (23) with [α-[^{35}S]thio]dATP.

Primer Extension. Synthetic ^{32}P-labeled oligonucleotides complementary to nucleotides 1–25 or 24–43 of nur/77 were hybridized to 30 μg of total cellular RNA from either quiescent cells or cells stimulated with serum for 3 hr in the presence of cycloheximide (10 μg/ml). Hybridization and subsequent reverse transcription were carried out as described (24), and reaction products were resolved by electrophoresis in 8% polyacrylamide gels containing 8 M urea.

RESULTS

Nucleotide Sequence of nur/77 cDNA and Its Encoded Protein. A cDNA clone representing the 3′ protein of nur/77 (clone 3CH77 in ref. 3) was used to select a nearly full-length cDNA clone from a library prepared by using poly(A)$^+$ mRNA isolated from BALB/c 3T3 cells stimulated with serum for 3 hr in the presence of cycloheximide (4). Several cDNA clones containing inserts of about 2.7 kilobases were independently isolated (approximately full-length as estimated by electrophoresis of nur/77 mRNA) and found to have identical restriction maps. Various restriction fragments of the cDNA clones hybridized to the same RNA species as detected by RNA blot analysis. One of these clones was chosen for sequence analysis.

The cDNA analyzed is 2456 nucleotides in length, excluding the poly(A) tail (Fig. 1). It contains a major open reading frame initiating at position 112 that is 1803 nucleotides in length and encodes a protein of molecular weight 64,738 with a calculated isoelectric point of 6.72. This reading frame is

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.

Abbreviation: NGF, nerve growth factor.
‡To whom reprint requests should be addressed.
§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04113).

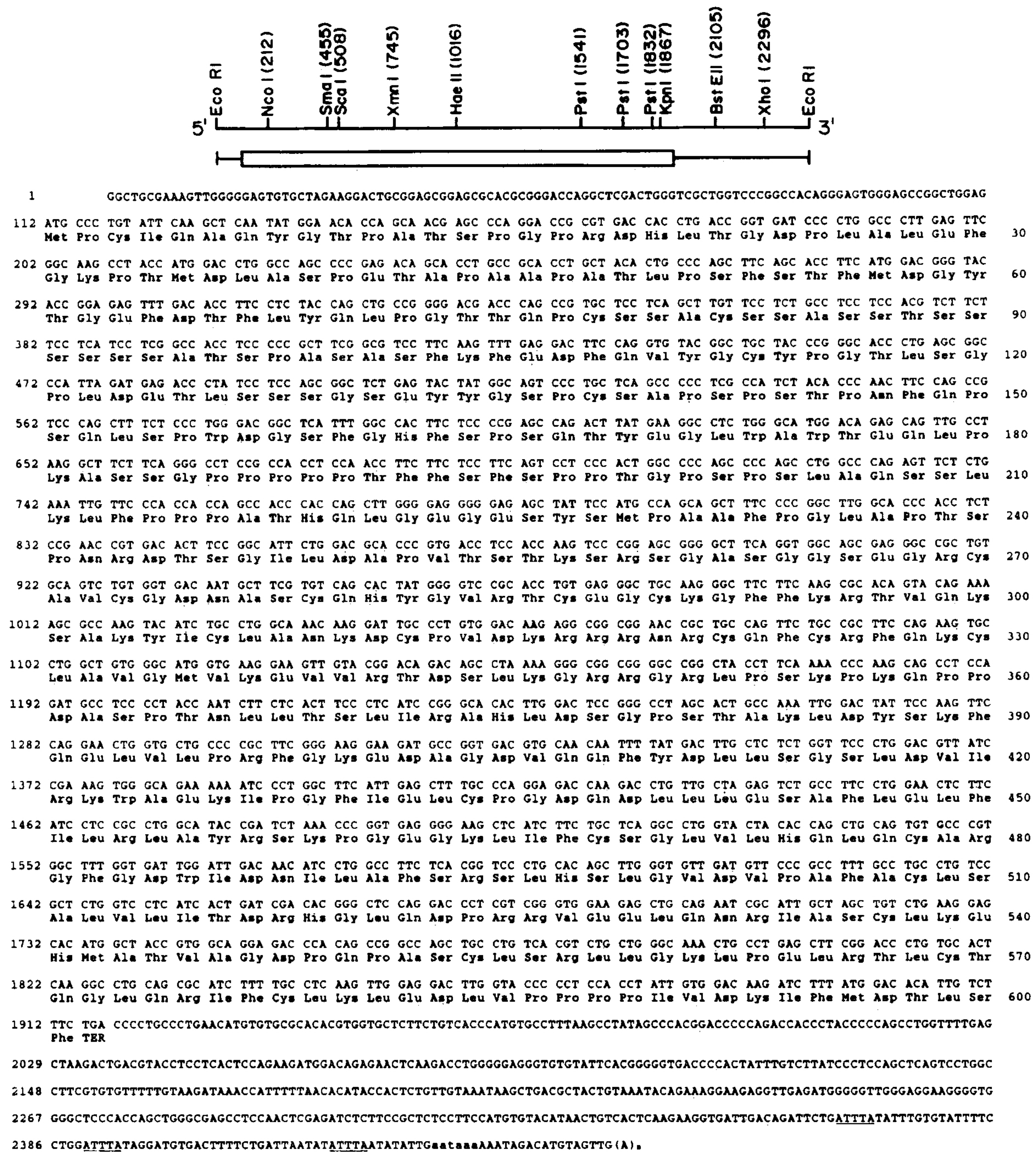

FIG. 1. Restriction map and nucleotide and deduced amino acid sequences of nur/77. (*Upper*) Schematic representation of the nur/77 cDNA showing cleavage sites of some common restriction enzymes. The open bar indicates the predicted coding region. (*Lower*) The complete nucleotide sequence and the predicted amino acid sequence of the nur/77 cDNA clone. Numbers at left refer to the first nucleotides on the lines, and numbers at right refer to the last amino acids on the lines. The ATTTA sequence motifs are underlined, and the polyadenylylation signal appears in lowercase letters. The 3'-terminal $(A)_n$ denotes the poly(A) tail.

followed by a 3' untranslated region of 542 nucleotides. A consensus polyadenylylation signal is located 18–23 bases upstream of the poly(A) tail. Like many other immediate early mRNAs, nur/77 mRNA has a short half-life (≈20 min; ref. 4). Consistent with this observation, there exist within a 60-base-pair segment of the 3' noncoding region three repeats of the ATTTA sequence motif thought to contribute to the instability of some mRNAs (25). In the deduced protein sequence, a potential N-linked glycosylation site is found at asparagine-276. Between the lysine residues at positions 32 and 104 there is a region rich in proline, glutamic acid, serine, and threonine. This type of "PEST" sequence is thought to be associated with proteins of short half-lives such as the *fos* and *myc* gene products and ornithine decarboxylase and is also found in members of the nuclear receptor superfamily (26, 27).

That the cDNA is nearly full-length was demonstrated by primer extension analysis. Reverse transcription of BALB/c 3T3 cell RNA from an oligonucleotide primer complementary to nucleotides 1–25 of the cDNA sequence indicated that the

FIG. 2. Primer extension analysis. Reactions were performed with 30 μg of total RNA from either quiescent BALB/c 3T3 cells (lanes 1 and 3) or cells stimulated with serum for 3 hr in the presence of cycloheximide (lanes 2 and 4). Two ^{32}P-labeled oligonucleotides, one complementary to nucleotides 1–25 of the cDNA sequence (lanes 1 and 2) and the other complementary to nucleotides 24–43 (lanes 3 and 4), were used. Left and right four lanes show ^{35}S-labeled sequencing-reaction products as size markers. The positions of oligonucleotides used as primers are indicated by arrows.

5′ end of the nur/77 mRNA occurs heterogeneously 9–17 nucleotides upstream of the primer sequence (Fig. 2, lanes 1 and 2). The predominant mRNA start site occurs 16 nucleotides upstream of the 5′ end of the sequence shown in Fig. 1. These results were confirmed by primer extension using an oligonucleotide complementary to nucleotides 24–43 of the cDNA; this reaction yielded products 32–40 nucleotides long (Fig. 2, lanes 3 and 4).

There are three in-frame ATG codons at the 5′ end of the nur/77 cDNA sequence occurring at positions 112, 214, and 280. In addition, an out-of-frame ATG occurs at position 134. The first ATG (position 112) has a poor flanking sequence for translation initiation according to the consensus sequence derived by Kozak (28), whereas the ATG at position 214 has a more favorable flanking sequence. However, if the ATG at position 112 is not used, the next ATG (position 134), which does have features of the consensus sequence for initiation, would be used (29), leading to translation of 13 codons of a different reading frame. Although nur/77 mRNA purified by hybrid selection is inefficiently translated in the rabbit reticulocyte lysate system (3), it does yield two detectable protein products with molecular weights of 64,000 and 58,000, consistent with initiation at positions 112 and 280 (data not shown). Further analysis will be required to determine the correct initiation site in the cell.

Relationship of nur/77 to the Nuclear Receptor Superfamily. The deduced amino acid sequence of nur/77 was compared to known protein sequences in the National Biomedical Research Foundation data base (release 14.0) and was found to share significant amino acid sequence similarity with the superfamily of nuclear ligand-binding transcription factors that includes the steroid and thyroid hormone receptors (5–16). This similarity occurs in two regions. One is a highly conserved 66-amino acid region demonstrated for some of these receptors to be the DNA-binding domain (Fig. 3; refs. 13, 14, 18–20). Most notable in this region is the presence of eight strictly conserved cysteine residues thought to form two DNA-binding "fingers," each coordinated by a zinc ion (30). The extensive conservation of these cysteines and other residues in this domain is characteristic of the members of the nuclear receptor superfamily (31). The nur/77 sequence contains 50–58% sequence identity in this domain when compared to the human nuclear receptors (Fig. 3). A second region of sequence similarity occurs carboxyl to the DNA-binding domain (Fig. 4). This region has been shown for several of the nuclear receptors to function as the ligand-binding domain (18, 19, 21, 22). Although they bind structurally distinct ligands, all members of the nuclear receptor superfamily analyzed to date share moderate sequence similarity in this region. The sequence similarity shared between nur/77 and known nuclear receptors is comparable to that shared among the known receptors. Thus the homology between nur/77 and known nuclear receptors indicates that nur/77 is a member of this superfamily.

DISCUSSION

The interaction of serum growth factors with their membrane receptors results in the sequential activation of specific genes. Some of these genes are transcriptionally activated within minutes of growth factor stimulation, even in the absence of protein synthesis (3, 4, 32–34). Among these immediate early genes are several that encode known or probable transcription factors: the protooncogenes c-*fos* (33) and c-*jun* (2, 35), *jun-B* (36), *Krox-20* (37), *zif/268* (NGFI-A, *Egr-1*, *Krox-24*) (38–41), and *fra-1* (42). In this communication we report that the immediate early gene represented by nur/77 cDNA is another member of this group, encoding a protein that is related to steroid receptors and other ligand-dependent transcription factors.

Such hormones as glucocorticoids, mineralocorticoids, estrogen, progesterone, testosterone, the morphogen reti-

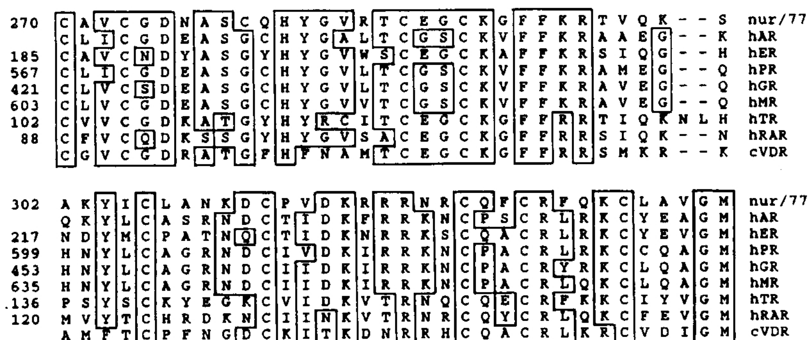

FIG. 3. Homology between the predicted nur/77 sequence and the DNA-binding domain of hormone receptors. Numbers on the left refer to the first amino acids (one-letter code) of the lines. Amino acids that are identical in five or more sequences are boxed. hAR, human androgen receptor (8, 9); hER, human estrogen receptor (7); hPR, human progesterone receptor (10); hGR, human glucocorticoid receptor (5); hMR, human mineralocorticoid receptor (6); hTR, human thyroid hormone receptor (11, 12); hRAR, human retinoic acid receptor (13–15); cVDR, chicken vitamin D receptor (16).

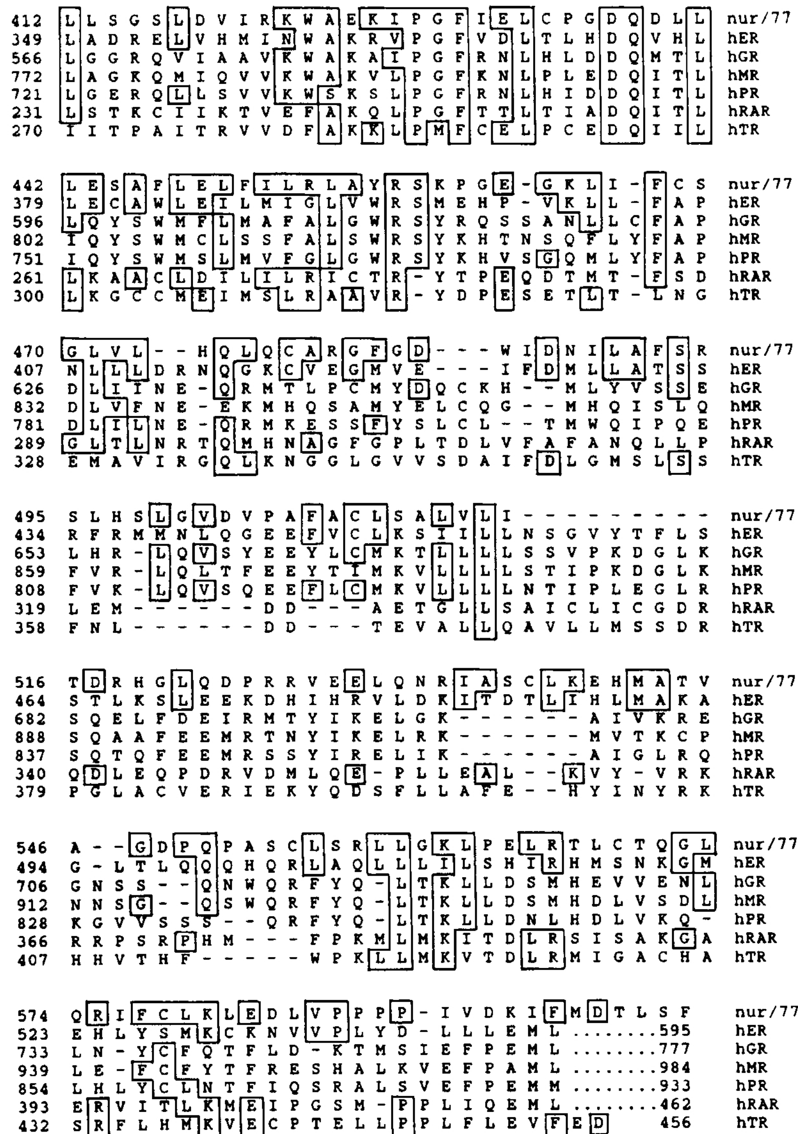

FIG. 4. Sequence comparison by a computer-generated alignment of nur/77 with the ligand-binding domains of hormone receptors. Numbers on the left refer to the first amino acids of the lines. Those amino acids of nur/77 that are identical to at least one other receptor sequence are boxed. Numbers at the ends of sequences refer to the total numbers of amino acids in those sequences. Receptor designations are as given in the legend of Fig. 3.

noic acid, thyroid hormone, and vitamin D are known to activate the transcription of specific genes through a receptor-mediated mechanism (17). The receptors for these compounds comprise a superfamily of intracellular proteins that, in contrast to cell-surface receptors, directly modulate transcription by interaction with their target genes. Binding to their respective ligands, which enter the cell by diffusion, increases affinity for specific enhancer-like DNA sequence elements associated with the target genes, resulting in activation or repression of transcription (6, 14, 20, 43–45). Whereas the amino-terminal domain, which is thought to be responsible in part for modulating gene expression (18, 46, 47), exhibits little sequence similarity among the different receptors, the central DNA-binding domain is highly conserved (13, 14, 18–20), and the carboxyl-terminal ligand-binding domain is moderately conserved (18, 19, 21, 22). The nur/77 protein shares all of these common structural features.

A ligand for the nur/77 protein has not been identified. Since both retinoic acid and thyroid hormone bind multiple tissue-specific receptors (12, 15, 48), it is possible that nur/77 is another receptor for one of the known ligands. An additional possibility is that nur/77 may bind an intracellular molecule generated by growth factor action, rather than a ligand that diffuses into the cell. It has also been suggested that thyroid hormone-related and cholesterol-derived compounds, some of which are known to alter gene expression at the transcriptional level (49), may act through binding to as yet unknown receptors (31). One such compound may be the ligand for nur/77.

The appearance of nur/77 mRNA is not restricted to growth factor-stimulated fibroblasts. This mRNA also appears in mouse liver within 1 hr after partial hepatectomy (2) and thus appears to play a role in the proliferative response *in vivo*. A related RNA appears in the rat pheochromocytoma cell line PC12 after exposure to nerve growth factor (NGF)

(ref. 50; M. E. Greenberg and L.F.L., unpublished observation), which induces neuronal differentiation of PC12 cells (51). Recently, a cDNA (NGFI-B) derived from mRNA of NGF-stimulated PC12 cells has been reported to encode a protein closely related to nur/77 (50). Since the nur/77 protein sequence contains only 22 amino acid substitutions when compared to the NGFI-B sequence, it is likely that nur/77 is the murine homologue of NGFI-B. It is striking that NGF and platelet-derived growth factor (or fibroblast growth factor) acting on different cell types to induce differentiation and growth, respectively, activate a subset of identical genes that encode probable transcription factors: the putative ligand-binding nuclear receptor described in this report, the protein encoded by the *fos* gene (33, 52), and another protein with zinc-finger sequences (38–41). The same regulatory proteins thus appear to mediate a variety of cellular responses to external signals, suggesting that some of the target genes of immediate early transcription factors are cell-type-specific.

We thank Richard Hall for excellent technical assistance. This work was supported by a grant from the National Institutes of Health to L.F.L. (R01 CA46565). T.G.H. is supported by a predoctoral training grant from the Public Health Service. L.F.L. is a recipient of the American Cancer Society Junior Faculty Research Award and is a Pew Scholar in the Biomedical Sciences.

1. Rozengurt, E. (1986) *Science* **234**, 161–166.
2. Nathans, D., Lau, L. F., Christy, B., Hartzell, S., Nakabeppu, Y. & Ryder, K. (1988) *Cold Spring Harbor Symp. Quant. Biol.*, in press.
3. Lau, L. F. & Nathans, D. (1985) *EMBO J.* **4**, 3145–3151.
4. Lau, L. F. & Nathans, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1182–1186.
5. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Nature (London)* **318**, 635–641.
6. Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E. & Evans, R. M. (1987) *Science* **237**, 268–275.
7. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J., Argos, P. & Chambon, P. (1986) *Nature (London)* **320**, 134–139.
8. Chang, C., Kokontis, J. & Liao, S. (1988) *Science* **240**, 324–326.
9. Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S. & Wilson, E. M. (1988) *Science* **240**, 327–330.
10. Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlansky, F., Guiochon-Mantel, A., Galibert, F. & Milgrom, E. (1987) *Biochem. Biophys. Res. Commun.* **143**, 740–748.
11. Weinberger, C., Thompson, C. C., Ong, E. S., Lebo, R., Gruol, D. J. & Evans, R. M. (1986) *Nature (London)* **324**, 641–646.
12. Thompson, C. C., Weinberger, C., Lebo, R. & Evans, R. M. (1987) *Science* **237**, 1610–1614.
13. Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. (1987) *Nature (London)* **330**, 444–450.
14. Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. (1987) *Nature (London)* **330**, 624–629.
15. Brand, N., Petkovich, M., Krust, A., Chambon, P., de The, H., Marchio, A., Tiollais, P. & Dejean, A. (1988) *Nature (London)* **332**, 850–853.
16. McDonnell, D. P., Mangelsdorf, D. J., Pike, J. W., Haussler, M. R. & O'Malley, B. W. (1987) *Science* **235**, 1214–1217.
17. Yamamoto, K. R. (1985) *Annu. Rev. Genet.* **19**, 209–252.
18. Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. (1986) *Cell* **46**, 645–652.
19. Gronemeyer, H., Turcotte, B., Quirin-Stricker, C., Bocquel, M. T., Meyer, M. E., Krozowski, Z., Jeltsch, J. M., Lerouge, T., Garnier, J. M. & Chambon, P. (1987) *EMBO J.* **6**, 3985–3994.
20. Green, S. & Chambon, P. (1987) *Nature (London)* **325**, 75–78.
21. Rusconi, S. & Yamamoto, K. R. (1987) *EMBO J.* **6**, 1309–1315.
22. Kumar, V., Green, S., Staub, A. & Chambon, P. (1986) *EMBO J.* **5**, 2231–2236.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
24. Linzer, D. H. & Mordacq, J. C. (1987) *EMBO J.* **6**, 2281–2288.
25. Shaw, G. & Kamen, R. (1986) *Cell* **46**, 659–667.
26. Rogers, S., Wells, R. & Rechsteiner, M. (1986) *Science* **234**, 364–368.
27. Rechsteiner, M., Rogers, S. & Rote, K. (1987) *Trends Biochem. Sci.* **12**, 390–394.
28. Kozak, M. (1986) *Cell* **44**, 283–292.
29. Kozak, M. (1980) *Cell* **22**, 7–8.
30. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3.
31. Evans, R. M. (1988) *Science* **240**, 889–895.
32. Cochran, B. H., Reffel, A. C. & Stiles, C. D. (1983) *Cell* **33**, 939–947.
33. Greenberg, M. E., Hermanowski, A. L. & Ziff, E. B. (1986) *Mol. Cell Biol.* **6**, 1050–1057.
34. Almendral, J. M., Sommer, D., MacDonald-Bravo, H., Burckhardt, J., Perera, J. & Bravo, R. (1988) *Mol. Cell. Biol.* **8**, 2140–2148.
35. Lamph, W. W., Wamsley, P., Sassone-Corsi, P. & Verma, I. M. (1988) *Nature (London)* **334**, 629–631.
36. Ryder, K., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1487–1491.
37. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. & Charnay, P. (1988) *EMBO J.* **7**, 29–35.
38. Christy, B. A., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7857–7861.
39. Milbrandt, J. (1987) *Science* **238**, 797–799.
40. Sukhatme, V. P., Cao, X., Chang, L. C., Tsai-Morris, C., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T., Le Beau, M. M. & Adamson, E. D. (1988) *Cell* **53**, 37–43.
41. Lemaire, P., Revelant, O., Bravo, R. & Charnay, P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4691–4695.
42. Cohen, D. R. & Curran, T. (1988) *Mol Cell. Biol.* **8**, 2063–2069.
43. Klein-Hitpass, L., Schorpp, M., Wagner, U. & Ryffel, G. U. (1986) *Cell* **46**, 1053–1061.
44. Cato, A. C. B., Miksicek, R., Schütz, G., Arnemann, J. & Beato, M. (1986) *EMBO J.* **5**, 2237–2240.
45. Adler, S., Waterman, M. L., He, X. & Rosenfeld, M. G. (1988) *Cell* **52**, 685–695.
46. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. & Chambon, P. (1987) *Cell* **51**, 941–951.
47. Tora, L., Gronemeyer, H., Turcotte, B., Gaub, M. & Chambon, P. (1988) *Nature (London)* **333**, 185–188.
48. Nakai, A., Seino, S., Sakurai, A., Szilak, I., Bell, G. I. & De-Groot, L. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2781–2785.
49. Sudhof, T. C., Russell, D. W., Brown, M. S. & Goldstein, J. L. (1987) *Cell* **48**, 1061–1069.
50. Milbrandt, J. (1988) *Neuron* **1**, 183–188.
51. Greene, L. A. & Tischler, A. S. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2424–2428.
52. Kruijer, W., Schubert, D. & Verma, I. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7330–7334.