Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 327–331, January 1989
Medical Sciences

Characterization and expression of a cDNA encoding the human androgen receptor

(androgen action/dihydrotestosterone/male sexual differentiation)

WAYNE D. TILLEY, MARCO MARCELLI, JEAN D. WILSON, AND MICHAEL J. MCPHAUL

Department of Internal Medicine, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235-8857

Contributed by Jean D. Wilson, October 5, 1988

ABSTRACT **We report the isolation of a cDNA that encodes the complete human androgen receptor. The cDNA predicts a protein of 917 amino acids with a molecular weight of 98,918. Introduction of the cDNA into heterologous mammalian cells caused expression of high levels of a protein that binds dihydrotestosterone with the affinity, specificity, and sedimentation properties characteristic of the native human androgen receptor. Comparisons with the amino acid sequence of previously cloned steroid hormone receptors revealed a high degree of sequence conservation with the progesterone, glucocorticoid, and mineralocorticoid receptors in the putative hormone and DNA-binding domain regions.**

Many members of the steroid and thyroid hormone receptor family have been purified to homogeneity, and the cDNAs for these receptors have been cloned by a variety of techniques (1). In contrast, it has not been possible to purify the androgen receptor to homogeneity, and less is known about its structure than for most intracellular receptors. Utilizing indirect techniques based on the high degree of conservation of nucleotide sequence in the DNA-binding domains of previously cloned steroid and thyroid hormone receptors (1) and the known X-chromosome location of the androgen receptor (2), three groups of investigators have reported isolation of cDNAs that encode portions of the human androgen receptor (3–5). In this paper, we report the isolation of a cDNA encoding the complete human androgen receptor.* We show that introduction of this cDNA into mammalian cells leads to the expression of high levels of a protein with the affinity, specificity, and sedimentation properties expected for the native receptor.

MATERIALS AND METHODS

Cell Culture. Lymphoblastoid cell lines were obtained from the American Type Culture Collection: GM03798, 46,XY; GM03714, 46,XX; GM01202B, 49,XXXXY; GM06061A, 49,XXXXX. Dulbecco's modified phosphate-buffered saline was from Hazleton (Lenexa, KS). Minimal essential medium was from GIBCO.

Isolation of Genomic Clones. A human X-chromosome library was prepared by L. Deaven (Los Alamos National Laboratory) from partially digested human X-chromosome genomic DNA inserted into the bacteriophage vector Charon 35. Nitrocellulose filter lifts and processing of filters were performed as described (6). Prehybridizations and hybridizations were performed under conditions of low stringency [10% formamide/5× SSC (1× SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7.0)/5× Denhardt's solution (1× Denhardt's solution is 200 μg each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin per ml)/0.1% sodium dodecyl sulfate/denatured salmon sperm DNA (100 μg/ml)] for 16 hr at 42°C. Hybridizations were performed with a mixture of DNA restriction fragments encompassing the entire cysteine-rich DNA-binding domain of the estrogen (7, 8), glucocorticoid (9), progesterone (10), and mineralocorticoid receptors (11) and of v-erbA (12). The individual DNA fragments were labeled to a high specific activity (>5 × 10^8 dpm per μg of DNA) by the random hexamer labeling method (13). After hybridization, the filters were washed three times for 30 min each in 2× SSC containing 0.1% sodium dodecyl sulfate at temperatures from 38°C to 55°C. Positive clones were purified by four successive platings and rescreenings. The screening of multiple platings of the purified clones with probes to the individual DNA-binding domains enabled the genomic clones to be grouped into two classes. Class 1 hybridized to the human estrogen receptor, and class 2 hybridized to the human progesterone and glucocorticoid receptors.

Southern Analysis of Genomic Clones. The chromosomal localization of the two classes of clones was established by Southern blotting of DNA prepared from human lymphoblastoid cell lines containing different numbers of X chromosomes. The DNA was digested with the restriction endonuclease *Eco*RI, and 10 μg of each digested DNA was electrophoresed on a 0.8% agarose gel and transferred to a nylon membrane (Biotrans, ICN). Prehybridizations and hybridizations were performed under conditions of high stringency (50% formamide/5× SSC/5× Denhardt's solution/0.1% sodium dodecyl sulfate/100 μg of denatured salmon sperm DNA per ml) for 16 hr at 42°C. DNA fragments from classes 1 and 2 of genomic clone were labeled by the random hexamer method. After hybridization, the membrane was washed twice in 0.1× SSC containing 0.1% sodium dodecyl sulfate for 60 min at 60°C.

Construction and Screening of Prostatic cDNA Libraries. A prostate cDNA library was prepared by using polyadenylylated RNA isolated from an open prostatectomy of a man with prostatic hyperplasia. First- and second-strand syntheses were performed by modifications of the method of Gubler and Hoffman (14) and processed as described (6). After *in vitro* packaging, the λgt10 library had a complexity of 2 × 10^6 independent recombinants. The library was screened at conditions of high stringency with a 1.4-kilobase (kb) restriction fragment from genomic clone AX1 (from class 2) containing the region homologous to the human progesterone and glucocorticoid receptors.

To ensure the isolation of a cDNA containing the complete androgen receptor, a second prostate library was constructed with an oligonucleotide (CTTTTGAAGAAGACCTT) derived from the sequence of the genomic clone AX1 (see Fig. 3). Two micrograms of the specific oligonucleotide was hybridized to 10 μg of poly(A)$^+$ RNA from the same prostate RNA preparation used to make the first cDNA library. The

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.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04150).

primer-extended library (10^6 independent recombinants) was screened with two restriction fragments derived from the largest cDNA clone (A1) shown in Fig. 1.

Sequencing. The cDNA was digested with restriction endonucleases, subcloned into appropriately digested M13 vectors, and sequenced by the dideoxy chain-termination technique (15). The entire sequence was determined on both strands. DNA sequence compilation and analysis were performed using the PC Gene (IntelliGenetics, Mountain View, CA) computer software.

Analysis of Mammalian Cells Transfected with the Androgen Receptor cDNA. The full-length human androgen receptor cDNA clone (A1) was subcloned into the expression vector CMV3 (provided by David W. Russell, University of Texas Southwestern Medical Center at Dallas). This vector allows the expression of cloned segments of DNA in eukaryotic cells under the control of the strong transcriptional promoter derived from the cytomegalovirus. Forty-eight hours after introduction of the cDNA (16) into COS-1 cells (17), the cells were harvested in Dulbecco's modified phosphate buffer, centrifuged at 800 × *g* for 10 min at 4°C, and washed once in Tris saline (50 mM Tris·HCl/150 mM NaCl, pH 7.4). The cell pellet was resuspended in 2–3 vol of TEGM buffer [10 mM Tris·HCl/1 mM EDTA/10% (vol/vol) glycerol/1 mM 2-mercaptoethanol/10 mM sodium molybdate, pH 7.2] and homogenized by repeated aspiration with a plastic syringe through a 25-gauge needle. The homogenate was centrifuged at 250,000 × *g* for 30 min, and 0.2-ml portions of cytosol (0.3 mg of protein) were incubated with various amounts of [^{3}H]dihydrotestosterone in the presence or absence of a 200-fold excess of unlabeled dihydrotestosterone for 5 hr at 4°C.

For examination of the specificity of binding of steroids to the androgen receptor, aliquots of cytosol (as described above) were incubated with 3 nM [^{3}H]dihydrotestosterone either alone or in the presence of a 1-, 10-, or 100-fold excess of unlabeled steroids for 5 hr at 4°C. For both the saturation and competition studies, unbound hormone was removed by addition of 50 μl of a dextran-coated charcoal suspension [5% (wt/vol) charcoal (Mallinckrodt), 0.5% (wt/vol) dextran T-70 (Pharmacia), and 0.1% (wt/vol) gelatin (Sigma) in 10 mM Tris/1 mM EDTA, pH 7.2]. The tubes were incubated for 5 min at 0°C and centrifuged for 10 min at 2000 × *g*. Aliquots of the supernatant were assayed for radioactivity after mixing with 5 ml of scintillation fluid. Protein was measured by the method of Lowry *et al.* (18).

RESULTS

The screening of the X-chromosome genomic library with the mixture of conserved DNA-binding domain fragments (see *Materials and Methods*) at low stringency identified 11 positive clones (of 10^7 total recombinants). Hybridization to the individual probes revealed two groups: class 1 clones hybridized strongly to the human estrogen receptor; class 2 clones hybridized weakly to the human glucocorticoid and progesterone receptors.

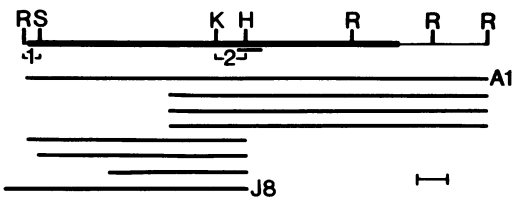

FIG. 1. Map of the androgen receptor cDNA clones. Restriction map of cDNA clones isolated from the oligo(dT)-primed (A series) and primer-extended (J series) cDNA libraries. R, *Eco*RI; S, *Sma* I; K, *Kpn* I; H, *Hin*dIII. Thick line, coding region; central thin bar containing the *Hin*dIII restriction site is the DNA-binding domain. Regions 1 and 2 were used to screen the primer-extended library. (Bar = 200 bp.)

Representative clones of both classes were examined for their X-chromosomal localization by Southern analysis using DNA from lymphoblastoid cell lines containing various numbers of X chromosomes. Hybridization with a representative clone from class 1 (FXE0.7) revealed a pattern consistent with an autosomal location (Fig. 2 *Left*). In contrast, hybridization of the same Southern blot with a probe from class 2 genomic clone AX1 (Fig. 2 *Right*) revealed a clear-cut gradation of signal proportional to the number of X chromosomes present in the parent cell line, consistent with an X-chromosomal location for clone AX1.

Both classes of genomic clones were also characterized by partial sequence analysis. Class 1, which hybridized strongly to the human estrogen receptor, contained a region of >200 base pairs (bp) that was identical to the sequence reported for the human estrogen receptor. Thus, this class of genomic clone was considered to represent a contaminant of the original X-chromosome library with chromosome 6, which is known to contain the gene that encodes the estrogen receptor (19). Sequence analysis of the second class of positive clones, represented by clone AX1, revealed a structure that was unique but highly related to the human progesterone and glucocorticoid receptors, leading us to focus our subsequent work on this class of positive clones.

We then screened the prostate cDNA library with a 1.4-kb restriction fragment derived from genomic clone AX1. This identified four positive clones, as shown in Fig. 1, ranging in size from 1.4 to 3.1 kb. To ensure the isolation of a cDNA clone encoding the entire open reading frame of the receptor, a primer-extended library was constructed from the same prostate mRNA as the original library. Screening this library with restriction fragments derived from cDNA clone A1 (shown in Fig. 1) resulted in the identification of four additional clones. The clones isolated from the primer-extended library had identical structures to that of A1, indicating that only a single class of 5′ terminus existed. Clone J8 was 150 bp longer than A1 at its 5′ end.

The nucleotide and predicted amino acid sequences of the human androgen receptor, as derived from analysis of cDNA clones A1 and J8, are shown in Fig. 3. The cDNA contains a single large open reading frame beginning at nucleotide 163 and extending to nucleotide 2913. The open reading frame predicts a protein of 917 amino acids and a molecular weight of 98,918. The entire open reading frame is contained within clone A1. We believe the ATG located at position 163 to be

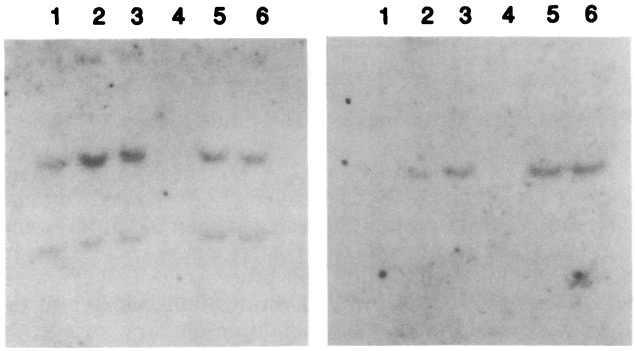

FIG. 2. Genomic clone AX1 is located on the human X chromosome. DNA (10 μg) from lymphoblastoid cell lines containing variable numbers of X chromosomes were digested with *Eco*RI, electrophoresed on a 0.8% agarose gel, and transferred to nitrocellulose. Lanes: 1, 46,XY; 2, 46,XX; 3, 49,XXXXY; 4, unlabeled molecular mass markers; 5 and 6, 49XXXXX. (*Left*) Blot probed with genomic clone FXE0.7 (class 1). (*Right*) Identical blot probed with genomic clone AX1 (class 2).

FIG. 3. Nucleotide and predicted amino acid sequence of the human androgen receptor. The nucleotide sequence is derived from analysis of androgen receptor clones A1 and J8. The complete sequences of clones A1 (residues 119–3061) and J8 (residues 1–1892) are shown. A termination codon in-frame with the translation initiation site is shown with an asterisk. The sequence corresponding to the oligonucleotide used in the construction of the primer extended library is underlined. Nucleotides 1558–3061 correspond to the partial cDNA previously reported (5). It is identical except for the presence in our sequence of GAA at positions 1849–1851, an A at position 2982, and a single C following nucleotide 3001.

the translation initiation site based on the in-frame termination codon located at nucleotide 58 and the fact that the sequence surrounding this ATG agrees with the consensus sequence described by Kozac (20).

The protein predicted to be encoded by the full-length androgen receptor cDNA has structural similarities to previously described steroid receptor sequences. Two regions show a high degree of relatedness to corresponding regions of the human progesterone, glucocorticoid, and mineralocorticoid receptors. First, the region of the cysteine-rich DNA-binding domain (amino acids 557–622) showed 85%, 79%, and 80% conservation of amino acid residues as compared to analogous domains of the progesterone, glucocorticoid, and mineralocorticoid receptors, respectively. The steroid-binding domain (amino acids 669–917) of the human androgen receptor exhibited 54%, 50%, and 51% amino acid identities with the corresponding domains of the progesterone, glucocorticoid, and mineralocorticoid receptors. In contrast, a lesser degree of amino acid conservation was observed in these domains for other members of the steroid and thyroid hormone receptor family. For example, the DNA- and steroid-binding domains of the human estrogen receptor exhibited 59% and 23% identities when compared to the corresponding regions of the human androgen receptor.

Fig. 4 shows an analysis of [^{3}H]dihydrotestosterone binding in cytosol fractions derived from COS-1 cells transiently transfected with the full-length androgen receptor cDNA (A1) subcloned in the expression vector CMV3. The specific binding reaches saturation levels with free dihydrotestosterone concentrations in the range 2–5 nM. Scatchard analysis of the binding data indicates an apparent dissociation constant (K_d) of ≈0.5 nM. The number of specific [^{3}H]dihydrotestosterone binding sites was 357 fmol per mg of cytosol protein. Cytosol from control (nontransfected) COS-1 cells showed no detectable specific binding of [^{3}H]dihydrotestosterone. Competition with different unlabeled steroids for the binding of 3 nM [^{3}H]dihydrotestosterone to cytosol derived from transfected cells (Fig. 5) revealed that the order of competitor effectiveness was methyltrienolone > dihydrotestosterone > testosterone > estradiol = progesterone > trimacinolone acetonide. Similar affinities and specificities

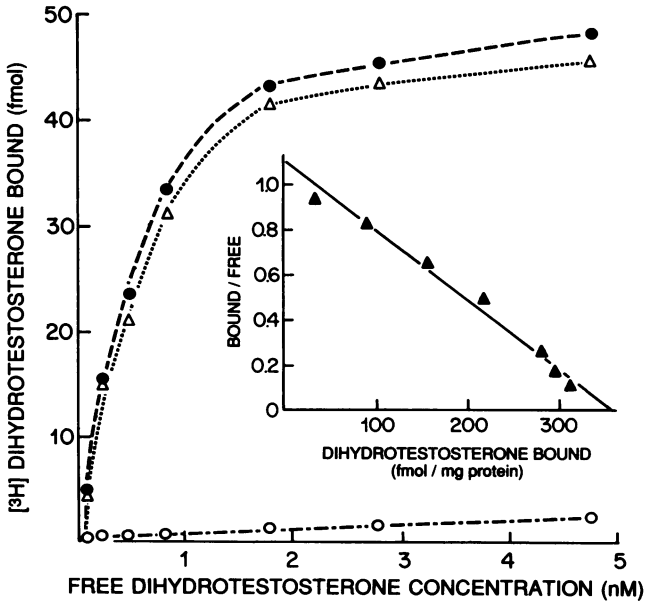

FIG. 4. COS-1 cells transfected with the human androgen receptor express high-affinity saturable dihydrotestosterone binding. Aliquots (0.2 ml containing 0.3 mg of protein) of cytosol from COS-1 cells transfected with the androgen receptor cDNA (A1) were incubated with increasing concentrations of [^{3}H]dihydrotestosterone in the absence or presence of unlabeled dihydrotestosterone. Specific binding (△), total binding (●), and nonspecific binding (○) are shown. Mock-transfected COS-1 cells showed no detectable specific androgen binding. (*Inset*) Scatchard plot of the specific binding data.

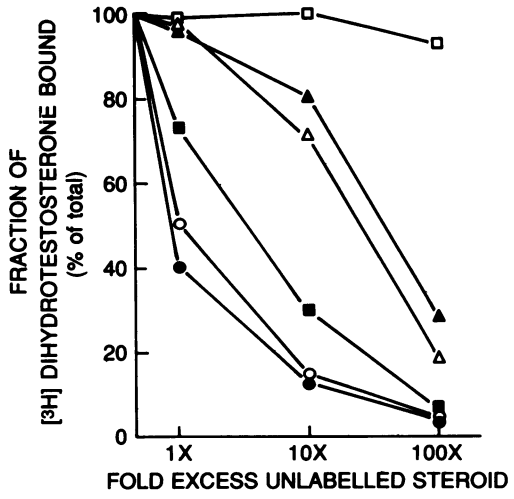

FIG. 5. Competition analysis of COS-1 cells transfected with the human androgen receptor. Aliquots of COS-1 cell cytosol (0.2 ml containing 0.3 mg of protein) were incubated for 5 hr at 4°C with 3 nM [^{3}H]dihydrotestosterone in the presence of various concentrations of unlabeled competitor steroid; methyltrienolone (●), dihydrotestosterone (○), testosterone (■), estradiol (△), progesterone (▲), and triamcinolone acetonide (□).

for dihydrotestosterone binding to the expressed cDNA were observed in monolayer cultures. Finally, the sedimentation coefficient of the androgen receptor expressed in COS-1 cells as determined by density-gradient analysis of low salt cytosol extracts revealed a single peak of radioactivity with an approximate sedimentation coefficient of 8.3S (data not shown).

DISCUSSION

We have isolated and characterized cDNA clones that encode the complete human androgen receptor. Three groups of investigators have reported the isolation of cDNAs encoding a portion of the human androgen receptor (3–5). Only one of these groups reported the nucleotide sequence of their partial cDNA (5). The sequence reported here is identical to that sequence previously reported with three minor differences (noted in Fig. 3).

The predicted amino acid sequence of the human androgen receptor shares structural similarities to members of the steroid hormone receptor family already described (7–12). In particular, the androgen receptor possesses a high degree of sequence conservation in the portion that corresponds to the cysteine-rich DNA-binding domain of other steroid hormone receptors. Less striking conservation is found in the carboxyl-terminal hormone-binding domain. Within these domains, the androgen receptor is most closely related to the progesterone, glucocorticoid, and mineralocorticoid receptors.

Other regions of the predicted protein sequence deserve comment. Three segments of the androgen receptor are composed of homopolymeric elements. One region (amino acids 58–77) contains 20 glutamine residues, while a second element (amino acids 448–470) is composed of 23 glycine residues. A third smaller region (amino acids 371–378) is composed of 8 proline residues. Although the function of these residues is not known, all three homopolymeric regions are in the amino-terminal segment of the molecule, which is the region that is least conserved among the steroid receptors and that is thought to be involved in the modulation of transcriptional activity (1). Interestingly, while the rat (21) and mouse (22) glucocorticoid receptors and the chicken progesterone receptor (23) contain regions of repeated glutamine or glutamic acid residues, such polymeric segments

are absent from the other steroid hormone receptors described to date. Similar homopolymeric regions containing glycine or glutamine residues were identified by computer analysis within the protein sequences of the Notch protein of *Drosophila*, human type II cytoskeletal keratin, and human procollagen α_1 (III)-chain precursor. We presume that these similarities represent common structural features of these molecules.

The introduction of the androgen receptor cDNA into COS-1 cells results in the expression of high levels of specific androgen binding. The affinity and specificity of the androgen receptor detected in these transfected cells are characteristic of those described for the human androgen receptor (24, 25). The sedimentation coefficient of the transfected receptor in low salt sucrose gradients was 8.3S, which is also consistent with that reported for the native human androgen receptor (24). Thus, the androgen receptor cDNA described here encodes a protein that possesses characteristics expected of the native human androgen receptor when expressed in heterologous cells.

A number of quantitative and qualitative defects in the androgen receptor leading to androgen resistance have been described in humans (26). Affected individuals have phenotypes ranging from that of women with primary amenorrhea to that of otherwise normal but infertile men, depending on the degree of impairment of receptor function. The availability of a complete cDNA encoding the human androgen receptor should allow elucidation of the molecular mechanisms of androgen action and, in particular, characterization of the mutations leading to androgen resistance.

Note Added in Proof. The sequence for the human androgen receptor cDNA reported by Chang *et al.* (27) after the submission of this paper is similar to that reported here except that they observed a G at position 48, an A at position 1579, an additional C residue following nucleotides 20 and 3001, and the dinucleotide GC at positions 793 and 794. In addition, the sequence reported here encodes 20 glutamine and 23 glycine residues in the homopolymeric regions, while that reported by Chang *et al.* encodes 17 and 27, respectively. Finally, the sequence of our cDNA following nucleotide 3044 diverges from that reported by Chang *et al.* but agrees with that reported by Trapman *et al.* (5).

W.D.T. and M.M. should be considered equal first authors. We thank Brenda Hennis for excellent assistance in manuscript preparation. This work was supported by grants from the National Institutes of Health (DK03892), from the Perot Family Foundation, from the Robert A. Welch Foundation (I-1090), and from the Medical Life and Health Insurance Medical Research Fund. W.D.T. is a recipient of a C. J. Martin Fellowship from the National Health and Medical Research Council of Australia. M.M. was supported by a Consiglio Nazionale delle Ricerce–North Atlantic Treaty Organization fellowship and a fellowship from Farmitalia, Milan, Italy. M.J.M. is a Culpeper Medical Scholar.

1. Evans, R. M. (1988) *Science* **240,** 889–895.
2. Migeon, B. R., Brown, T. R., Axelman, J. & Migeon, C. J. (1981) *Proc. Natl. Acad. Sci, USA* **78,** 6339–6343.
3. Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S. & Wilson, E. M. (1988) *Science* **240,** 327–330.
4. Chang, C., Kokontis, J. & Liao, S. (1988) *Science* **240,** 324–326.
5. Trapman, J., Klaassen, P., Kuiper, G. G. J. M., van der Korput, J. A. G. M., Faber, P. W., van Rooij, H. C. J., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E. & Brinkmann, A. O. (1988) *Biochem. Biophys. Res. Commun.* **153,** 241–248.
6. McPhaul, M. J. & Berg, P. (1987) *Mol. Cell. Biol.* **7,** 1841–1847.
7. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320,** 134–139.
8. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. & Shine, J. (1986) *Science* **231,** 1150–1154.
9. 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.
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. Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E. & Evans, R. M. (1987) *Science* **237,** 268–275.
12. Debuire, B., Henry, C., Benaissa, M., Biserte, G., Claverie, J. M., Saule, S., Martin, P. & Stehelin, D. (1984) *Science* **224,** 1456–1459.
13. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
14. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
15. Sanger, F., Jackson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
16. Frost, E. & Williams, J. (1978) *Virology* **91,** 39–50.
17. Gluzman, Y. (1981) *Cell* **231,** 175–183.
18. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265–268.
19. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J.-R. & Chambon, P. (1987) *Cell* **51,** 941–951.
20. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
21. Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wikström, A. C., Gustafsson, J.-A. & Yamamoto, K. R. (1986) *Cell* **46,** 389–399.
22. Danielsen, M., Northrup, J. P. & Ringold, G. M. (1986) *EMBO J.* **5,** 2513–2522.
23. Conneely, O. M., Dobson, A. D. W., Tsai, M.-J., Beattie, W. G., Toft, D. O., Huckaby, C. S., Zarucki, T., Schrader, W. T. & O'Malley, B. W. (1987) *Mol. Endocrinol.* **1,** 517–525.
24. Kovacs, W. J., Griffin, J. E., Weaver, D. D., Carlson, B. R. & Wilson, J. D. (1984) *J. Clin. Invest.* **73,** 1095–1104.
25. Breiner, M., Romalo, G. & Schweikert, H.-U. (1986) *Klin. Wochenschr.* **64,** 732–737.
26. Wilson, J. D. & Griffin, J. E. (1985) *Trends Genet.* **1,** 335–339.
27. Chang, C., Kokontis, J. & Liao, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7211–7215.