Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4437–4440, July 1987
Biochemistry

Domain structure of the glucocorticoid receptor protein

(proteolysis/steroid binding/DNA binding/protein sequence)

JAN CARLSTEDT-DUKE*, PER-ERIK STRÖMSTEDT*, ÖRJAN WRANGE†, TOMAS BERGMAN‡, JAN-ÅKE GUSTAFSSON*, AND HANS JÖRNVALL‡

*Department of Medical Nutrition, Karolinska Institute, Huddinge University Hospital, F69, S-141 86 Huddinge, Sweden; and Departments of †Medical Cell Genetics and ‡Chemistry, Karolinska Institute, Box 60400, S-104 01 Stockholm, Sweden

Communicated by Viktor Mutt, March 26, 1987 (received for review December 1, 1986)

ABSTRACT **The purified rat liver glucocorticoid receptor protein was analyzed by limited proteolysis and amino acid sequence determination. The NH_2 terminus appears to be blocked. The steroid-binding domain, defined by a unique tryptic cleavage site, corresponds to the COOH-terminal part of the protein with the domain border in the region of residue 518. The DNA-binding domain, defined by a region with chymotryptic cleavage sites, is immediately adjacent to the steroid-binding domain and reflects another domain border in the region of residues 410–414. The results described at the protein level in this report confirm functional data previously obtained by mutations at the genetic level.**

Biochemical, immunological, and genetic analyses of glucocorticoid receptor protein (GR) imply that the protein folds into three distinct functional domains, which mediate hormone binding (domain A), DNA binding (domain B), and an immunodominant domain (domain C) of unknown function (1–7). At the protein level, these domains can be defined by limited proteolysis of crude or purified preparations of rat liver GR (2, 4, 6–8). Certain regions of the protein are highly susceptible to proteolysis with a variety of proteases such as α-chymotrypsin, trypsin, papain, or lysosomal extract. These regions represent the interdomain segments and are probably hinge regions within the protein. The steroid-binding domain is defined by proteolysis of GR with trypsin, giving rise to a fragment with an estimated size of 27 kDa according to $NaDodSO_4$/polyacrylamide gel electrophoresis (4). Digestion of GR with α-chymotrypsin, lysosomal extract, or very low concentrations of trypsin (7, 8) results in a fragment with an estimated size of 39 kDa (4). This fragment contains both the steroid-binding and DNA-binding domains (7). Further digestion of this fragment with trypsin results in a fragment containing only the steroid-binding domain (7).

Recently, cDNA clones for the human (9), rat (10), and mouse (11) GR have been described that now enable a more exact definition of the domains. In this report, we compare the amino acid sequences of rat GR defining the functional domains at the protein level with the nucleotide sequences obtained from the cDNA clone for the rat GR (10). We have clearly established a unique tryptic site defining the steroid-binding domain and have also strong indications of a region with chymotryptic sites defining the DNA-binding domain.

MATERIALS AND METHODS

Purification of GR. Rat liver GR was purified as described (12, 13). Cytosol, incubated with the synthetic glucocorticoid [^{3}H]triamcinolone acetonide, was passed through phosphocellulose, the receptor was activated by dilution and incubation at 25°C, and it was chromatographed on DNA-cellulose. GR was eluted with 27.5 mM $MgCl_2$ and further purified by chromatography on DEAE-Sepharose, eluted with a linear NaCl gradient. The receptor was detected by analysis for ^{3}H radioactivity. Each purification resulted in a yield of ≈50 μg of GR, starting from eight rat livers. The purified preparations of GR contained the 94-kDa GR, the 72-kDa GR-associated protein that is unrelated to GR (12, 13), as well as very small amounts of proteolytic GR fragments (usually <5% of total protein according to densitometric analysis of the Coomassie blue-stained $NaDodSO_4$/polyacrylamide gel following electrophoresis).

Limited Proteolysis of GR. Purified GR was incubated with α-chymotrypsin (1 μg per μg of GR) or trypsin (0.3 μg per μg of GR) at 10°C for 30 min. Proteins were precipitated with 20% (wt/vol) trichloroacetic acid and analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis. The proteolytically derived fragments were detected by negative staining with 1 M KCl, cut out, and electroblotted onto glass fiber filter discs (14) for direct analysis by gas-phase sequencer degradations.

Cleavage of GR by CNBr. Purified GR was cleaved with CNBr and the resulting peptides were purified by gel-exclusion chromatography (Sephadex G-50 in 30% acetic acid) and subsequent HPLC (μBondapak C_{18}) with a gradient of acetonitrile in 0.1% trifluoroacetic acid. Peptides obtained were submitted to analysis in a sequencer.

RESULTS AND DISCUSSION

Protein Structure. Tests for direct analysis of purified rat liver GR (12, 13) by liquid-phase or gas-phase sequencer degradations were negative, suggesting that the NH_2 terminus is blocked. This was the case for analysis of the 94-kDa GR purified by $NaDodSO_4$/polyacrylamide gel electrophoresis and for analysis of purified preparations of GR, containing both the 94-kDa GR and the 72-kDa GR-associated protein. Different preparations of the receptor were therefore cleaved with trypsin, α-chymotrypsin, and cyanogen bromide. Resulting peptides were submitted to gas-phase sequencer analysis. Six of the CNBr fragments identified the NH_2-terminal structures shown in Fig. 1, which are in complete agreement with corresponding structures indirectly deduced from the cDNA nucleotide sequence for the open reading frame (10), thus confirming the conclusions from the latter by direct analysis at six different regions. In addition, peptides from two other regions starting at positions 120 and 337 were detected but were incompletely analyzed and are not included in Fig. 1, although these segments also support the indirectly deduced structure. Combined, all the sequences range from close to the NH_2 terminus (residues 28–42) to close to the COOH terminus (residues 770–790), inferring that the cDNA clone described (10) represents the intact receptor protein. This was further confirmed by determination of the amino acid composition of the purified

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: GR, glucocorticoid receptor protein.

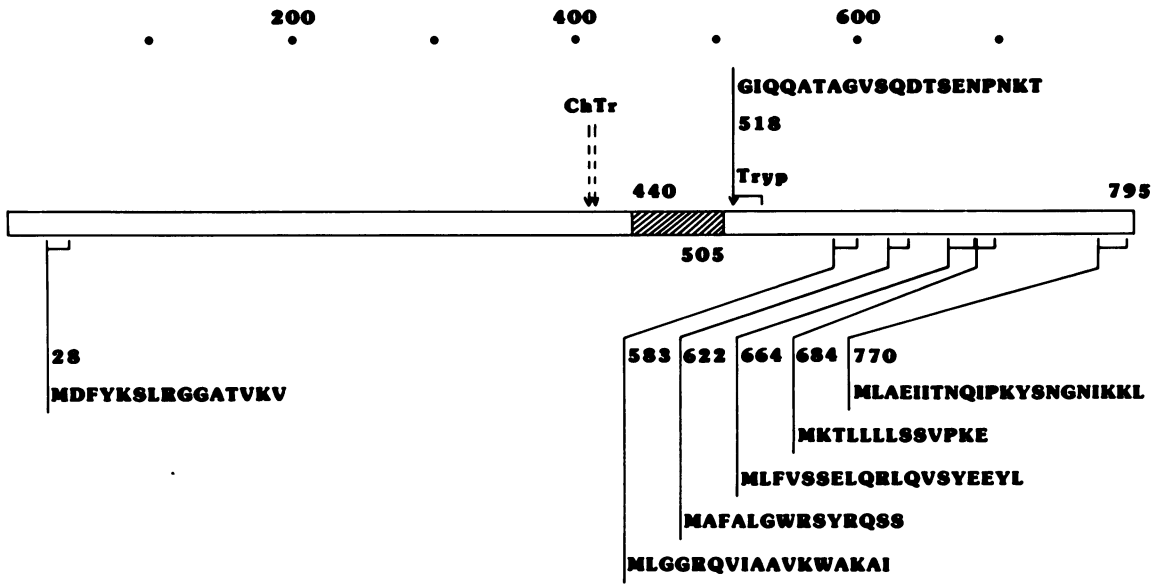

FIG. 1. Amino acid sequences (designated by the single-letter code) confirmed for the purified GR. GR was purified as described (11, 12) and cleaved with α-chymotrypsin (ChTr), trypsin (Tryp), or CNBr, and the resulting peptides were analyzed for amino acid sequence. The structures obtained were compared with the corresponding nucleotide sequence of rat liver GR cDNA (10). Numbers refer to the amino acid residues as defined by the nucleotide sequence for the coding region of the cDNA (total, 795 residues). Dashed arrows for the chymotryptic cleavages indicate multiple and probably nonstoichiometric cleavages, producing a less clearly defined segment than the well-defined cleavage in the case of trypsin. Hatched region corresponds to the putative DNA-binding region (10). The sequences shown in the lower part of the figure correspond to the peptides identified after cleavage with CNBr, which cleaves after the methionine shown at the first position in each case.

receptor protein after hydrolysis (Table 1) and comparison of these values with those indirectly deduced. Finally, the purified activated GR has a size, as determined by electrophoresis (4, 12), that is in complete agreement with that deduced from the cDNA open reading frame. Since the data from the cDNA and the peptide analyses for the purified activated GR are similar in all these respects, extensive proteolytic processing of the protein, posttranslationally or during activation, appears unlikely. However, limited processing occurs. Thus, as stated above, the NH_2 terminus appears to be blocked. From the NH_2-terminal structure (10), the known properties of *N*-acetyl-blocked proteins in general (15), and the rules for NH_2-terminal processing (16), it appears likely that the mature NH_2 terminus is formed by an acetylated residue, which can be the initiator methionine itself. Other minor modifications cannot be completely excluded, especially not in segments involving the terminal residues (1–27 or 791–795), which were not directly demonstrated by sequence analysis of the protein.

Table 1. Amino acid composition of GR determined directly from acid hydrolysis and compared to the composition indirectly deduced from cDNA

	Hydrolysis of purified GR, mol %		Indirectly from cDNA clone	
			No.	mol %
Cys	0.6		20	2.5
Asp	10.1		40	5.0
Asn			34	4.3
				9.3
Thr	5.5		44	5.5
Ser	9.8		88	11.1
Glu	13.8		43	5.4
Gln			62	7.8
				13.2
Pro	5.9		46	5.8
Gly	8.4		57	7.2
Ala	6.0		42	5.3
Val	5.1		39	4.9
Met	2.3		23	2.9
Ile	3.8		33	4.1
Leu	10.0		82	10.3
Tyr	3.1		22	2.8
Phe	3.4		24	3.0
Lys	6.8		51	6.4
His	1.4		9	1.1
Arg	4.3		29	3.6
Trp	ND		7	0.9
		Total	795	

The values obtained for the purified receptor protein are the averages of four preparations but are uncorrected for slow release (Val, Ile) and destruction (Trp, Cys, Ser). Tryptophan was not determined (ND). The data for the cDNA clone are from ref. 10.

Steroid-Binding Domain. Digestion of the receptor protein with trypsin (0.3 μg per μg of GR) yields a 27-kDa proteolytic fragment (Fig. 2). This fragment corresponds to the steroid-binding domain, domain A (6–8). Gas-phase sequencer analysis of this fragment after isolation by $NaDodSO_4$/polyacrylamide gel electrophoresis and electroblotting (14) shows that it starts with the NH_2 terminus at position 518 of the intact molecule. This result establishes that the fragment corresponds to the COOH-terminal region of the protein, with the domain border located close to residue 518 (Fig. 1). The result confirms the conclusions from other results by Hollenberg and collaborators (9, 18), in which they have described two different cDNA clones, of which the product of only one binds steroids, corresponding to human GRs, OB7 and OB10. A large portion of the OB10 open reading frame overlaps with OB7. Fifty codons before the OB7 COOH terminus, OB10 shifts to a nonhomologous sequence that specifies 15 in-frame codons and a 1.4-kilobase 3′ nontranslated segment homologous with rat 3′ nontranslated sequences situated between the two polyadenylylation signals (9, 10) and obviously derived by elimination of part of the mRNA corresponding to this part of the A domain. Expression of the two clones OB7 and OB10 results in two different forms of GR of which only one, corresponding to OB7, binds glucocorticoids (9). Thus, deletion of 50 residues at the COOH terminus of GR destroys the steroid-binding capacity, which is in full concordance with the present localization of the steroid-binding domain

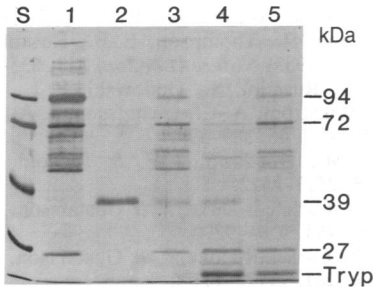

FIG. 2. Analysis of purified and digested GR by $NaDodSO_4$/12% polyacrylamide gel electrophoresis. After electrophoresis, the gel was stained with ammoniacal $AgNO_3$. Purified rat liver GR (lane 1) was digested with α-chymotrypsin (lane 3) or trypsin (lane 5). An aliquot of purified GR was digested with α-chymotrypsin and further purified on DNA-cellulose (4, 17) (lane 2). The purified chymotryptic 39-kDa fragment was further digested with trypsin (lane 4) at a concentration of trypsin equivalent to that used for the sample shown in lane 5 (0.3 μg of trypsin per μg of intact GR). The positions of intact GR (94 kDa), the GR-associated protein (12, 13) (72 kDa), the chymotryptic GR fragment (39 kDa), the tryptic GR fragment (27 kDa), and trypsin (Tryp) are indicated. α-Chymotrypsin migrates with the front (not shown). This particular preparation of purified GR contained a small amount of the proteolytic 27-kDa fragment (lane 1), but this was removed by further purification on DNA-cellulose (lane 2). The standard proteins (lane S) shown are 97.4, 67, 45, 30, and 20.1 kDa, respectively.

(domain A) to the COOH terminus. The size of the domain according to the amino acid sequence is 31,600 Da (residues 518–795).

DNA-Binding Domain. Digestion of the purified GR with α-chymotrypsin (1 μg per μg of GR) results in a 39-kDa fragment (Fig. 2). This band is slightly broader and less distinct than the band corresponding to the tryptic 27-kDa fragment, indicating a heterogeneity in the α-chymotrypsin cleavage, which was confirmed by NH_2-terminal amino acid sequence analysis (Fig. 1). Thus, the results indicated a mixture of NH_2 termini corresponding to several cleavages. This mixture increased the complexity of the interpretations and decreased the yields of each structure, but two segments were identified to contain tyrosine (at position 4) and methionine (at position 5). In the region in question, these properties fit only with cleavages after positions 409 and 413 (10). Starts of the 39-kDa fragment at these positions would imply that the 39-kDa fragment(s) also contain the entire 27-kDa fragment. To confirm this conclusion, the 39-kDa fragment was digested with trypsin, which produced a 27-kDa fragment identical in NH_2-terminal sequence (Fig. 1) and electrophoretic mobility (Fig. 2) to the 27-kDa fragment already described above. Since the tryptic product of the 39-kDa fragment is identical in size to the 27-kDa fragment, it appears likely that there is no proteolysis of the COOH terminus. Consequently, the 39-kDa fragment is concluded to have a ragged NH_2-terminal end with major forms starting close to positions 410–414, corresponding to ≈43 kDa. The fragment is also known to contain both the steroid-binding (A) and DNA-binding (B) domains (6–8) and to correspond to the truncated form of receptor occurring in glucocorticoid-resistant mouse lymphoma cell lines (1, 3, 5). Assuming that the border between domains B and A lies between residues 517 and 518, this would mean that the size of domain B is ≈11,400. The demonstration of the same unique tryptic site for the tryptic product obtained from the 39-kDa fragment, as well as the agreement between the previously calculated size of domain B (4, 6–8) and the size obtained in this study strengthen the present identification of the two domain borders as close to positions 410–414 and close to position 518.

A putative DNA-binding region has previously been ascribed to residues 440–505 (ref 10; hatched region, Fig. 1). This region is rich in cysteine, arginine, and lysine residues and demonstrates a remarkable homology to regions within the human glucocorticoid (9) and estrogen (19, 20) receptors, the mouse glucocorticoid receptor (11), the chicken estrogen receptor (21), and v-*erbA* (9, 10, 19–21). This region is also reminiscent of the nucleic acid binding region of the transcriptional factor TFIIIA (22) as well as of other eukaryotic DNA-binding proteins (23).

The remaining NH_2-terminal part of the protein corresponds to domain C. The exact function of this domain is still unknown, although it appears to play a role in the specific interaction with DNA and in gene regulation. Deletion of this domain reduces the specificity of the DNA-binding by a factor of 260 (17), although binding of hormone and subsequent heat activation are not affected (7). Analysis by RNA blot has shown that this domain appears to be deleted in the glucocorticoid-resistant nt^i S49 mouse lymphoma cells (24, 25). The receptor in these cells has a truncated form with a size and with physicochemical properties corresponding to the α-chymotryptic 39-kDa fragment of the purified receptor protein (1, 3, 5–8).

Correlation with Mutational Data. Recently Giguère *et al.* (26) have identified four different functional regions within the human GR cDNA by insertion mutants. Insertions at residue 550 or COOH-terminal thereof affect the steroid-binding properties of the receptor. The putative DNA-binding region homologous to other steroid receptors and v-*erbA* corresponds to residues 421–481 and insertions in this region affect the transcriptional activity of the mutant protein without affecting the steroid binding. Insertions at residues 120, 204, and 214 (τ_1) reduce the transcriptional activity without any effect on steroid binding. Finally insertions at residues 448 or 490 (τ_2) completely destroy the steroid-dependent transcriptional activity without affecting the binding of the steroid at all. The next position of insertion tested was at residue 515, at which point no effect was seen. In the human GR, the chymotrypsin sites defining the domain B/C border correspond to residues 389 and 393 and the trypsin site defining the domain A/B border corresponds to residue 498. Giguère *et al.* (26) found no effect of insertions at residues 404 or 408, although these would lie within the functional DNA-binding domain presently defined from studies at the protein level. Interestingly, the τ_2 region (residues 488–514) coincides exactly with the domain A/B border, which is the putative hinge region between the two functional domains. The hinge region probably plays a major role in the steroid-dependent activation of the receptor protein to the DNA-binding state. Deletion of 29–229 residues at the COOH-terminal end destroys the steroid-binding activity while enabling steroid-independent binding to specific DNA (27). This indicates that the DNA-binding site is normally protected by the steroid-binding domain. Binding of the steroid results in a conformational change centered on the putative hinge region at the domain border, exposing the DNA-binding site. Thus, the domain structure at the protein level described in this paper provides a valuable complement to the functional data obtained by mutagenesis of GR cDNA. Finally, the direct protein data confirm many regions of the indirectly cDNA-deduced protein sequence.

In summary, amino acid sequence determinations of the glucocorticoid receptor fragments have defined the borders of the functional domains of the protein by direct analysis. The positions of two interdomain regions that are highly susceptible to a variety of proteases have been described. The steroid-binding domain, domain A, comprises the region corresponding approximately to residues 518–795 (31,600 Da). The DNA-binding domain, domain B, is the preceding one, while the large domain C that may contain promoter-

modulatory functions (17, 26) corresponds to the NH_2-terminal fragment. Availability of the cDNA GR clones will enable further analysis of the different domains in variants with site-directed deletions and/or mutations. This will provide valuable information at the nucleotide level complementary to the domain structure now described at the protein level.

Note. Following completion of this manuscript, the work of Danielsen *et al.* (11) was published, in which they cloned two mutant forms of mouse GR. They describe three pertinent single point mutations affecting receptor function. The position of these mutations is concordant with the functional domains described in this paper. A mutation at mouse GR residue 484 Arg→His (equivalent to rat residue 496) destroys the DNA-binding capacity. Mutations at mouse GR residues 546 Glu→Gly and 770 Tyr→Asn (equivalent to rat residues 558 and 782, respectively) destroy the steroid-binding capacity completely (residue 546) or partially (residue 770).

This work was supported by grants from the Swedish Medical Research Council (Projects 2819, 3532, and 6245), the Swedish Cancer Society (Project 1806), and the Alex and Eva Wallström Foundation.

1. Yamamoto, K. R., Gehring, U., Stampfer, M. R. & Sibley, C. H. (1976) *Recent Prog. Horm. Res.* **32,** 3–32.
2. Carlstedt-Duke, J., Okret, S., Wrange, Ö. & Gustafsson, J.-Å. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4260–4264.
3. Okret, S., Stevens, Y.-W., Carlstedt-Duke, J., Wrange, Ö., Gustafsson, J.-Å. & Stevens, J. (1983) *Cancer Res.* **43,** 3127–3131.
4. Wrange, Ö., Okret, S., Radojcic, M., Carlstedt-Duke, J. & Gustafsson, J.-Å. (1984) *J. Biol. Chem.* **259,** 4534–4541.
5. Westphal, H. M., Mugele, K., Beato, M. & Gehring, U. (1984) *EMBO J.* **3,** 1493–1498.
6. Carlstedt-Duke, J. & Gustafsson, J.-Å. (1986) in *Thioredoxin and Glutaredoxin Systems: Structure and Function*, eds. Holmgren, A., Brandén, C.-I., Jörnvall, H. & Sjöberg, B.-M. (Raven, New York), pp. 391–401.
7. Wrange, Ö. & Gustafsson, J.-Å. (1978) *J. Biol. Chem.* **253,** 856–865.
8. Carlstedt-Duke, J., Wrange, Ö., Okret, S., Stevens, J., Stevens, Y.-W. & Gustafsson, J.-Å. (1983) in *Gene Regulation by Steroid Hormones*, eds. Roy, A. K. & Clark, J. H. (Springer, New York), Vol. 2, pp. 151–180.
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. Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. M., Okret, S., Wikström, A.-C., Gustafsson, J.-Å. & Yamamoto, K. R. (1986) *Cell* **46,** 389–399.
11. Danielsen, M., Northrop, J. P. & Ringold, G. M. (1986) *EMBO J.* **5,** 2513–2522.
12. Wrange, Ö., Carlstedt-Duke, J. & Gustafsson, J.-Å. (1979) *J. Biol. Chem.* **254,** 9284–9290.
13. Wrange, Ö., Carlstedt-Duke, J. & Gustafsson, J.-Å. (1986) *J. Biol. Chem.* **261,** 11770–11778.
14. Bergman, T. & Jörnvall, H. (1987) in *Methods in Protein Sequence Analysis*, ed. Walsh, K. A. (Humana, Clifton, NJ), in press.
15. Persson, B., Flinta, C., von Heijne, G. & Jörnvall, H. (1985) *Eur. J. Biochem.* **152,** 523–527.
16. Flinta, C., Persson, B., Jörnvall, H. & von Heijne, G. (1986) *Eur. J. Biochem.* **154,** 193–196.
17. Payvar, F. & Wrange, Ö. (1983) in *Steroid Hormone Receptors: Structure and Function*, eds. Eriksson, H. & Gustafsson, J.-Å. (Elsevier, Amsterdam), pp. 267–284.
18. Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. (1985) *Nature (London)* **318,** 670–672.
19. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. & Shine, J. (1986) *Science* **231,** 1150–1154.
20. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320,** 134–139.
21. Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M. & Chambon, P. (1986) *EMBO J.* **5,** 891–897.
22. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
23. Berg, J. M. (1986) *Science* **232,** 485–487.
24. Miesfeld, R., Okret, S., Wikstrom, A.-C., Wrange, Ö., Gustafsson, J.-Å. & Yamamoto, K. R. (1984) *Nature (London)* **312,** 779–781.
25. Miesfeld, R., Rusconi, S., Okret, S., Wikström, A.-C., Gustafsson, J.-Å. & Yamamoto, K. R. (1985) in *Sequence Specificity in Transcription and Translation*, UCLA Symposium on Molecular and Cellular Biology, New Series, eds. Calender, R. & Gold, L. (Liss, New York), Vol. 30, pp. 535–545.
26. Giguère, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. (1986) *Cell* **46,** 645–652.
27. Godowski, P. J., Rusconi, S., Miesfeld, R. & Yamamoto, K. R. (1987) *Nature (London)* **325,** 365–368.