Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3029–3033, May 1984
Biochemistry

Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus

(steroid hormone receptors/protein–DNA interaction/retrovirus enhancers)

CLAUS SCHEIDEREIT AND MIGUEL BEATO*

Institut für Physiologische Chemie der Philipps-Universität, Deutschhausstrasse 1-2, D-3550 Marburg, Federal Republic of Germany

Communicated by Erwin Bünning, January 30, 1984

ABSTRACT **Glucocorticoid hormones enhance the transcription of mouse mammary tumor virus DNA by mechanisms involving a direct interaction of the hormone receptor with four binding sites in a glucocorticoid regulatory element located between −72 and −192 base pairs upstream of the main transcription initiation site within the proviral long terminal repeat regions. Methylation at the N-7 position of any of three G residues within one of the binding sites prevents binding of the receptor. In addition, in the presence of the receptor, methylation by dimethyl sulfate is reduced at several G residues, indicating sites of contact between the receptor and DNA at these positions. The G residues in the hexanucleotide 5′-T-G-T-T-C-T-3′ / 3′-A-C-A-A-G-A-5′ were protected by the receptor against MH2-specific gene. (*iii*) *myc* is followed by the 3′-terminal *c* region of about 400 nucleotides, which is colinear with that of Rous sarcoma virus except for a substitution near the 5′ end of the long terminal repeat. It is concluded that MH2 contains two genes with oncogenic potential, the Δ*gag–mht* gene, which is closely related to the Δ*gag–raf* transforming gene of MSV 3611, and the *myc* gene, which is related to the transforming gene of MC29. Furthermore, it may be concluded that the cellular proto-*onc* genes, which on sequence transduction become viral *onc* genes, are a small group because among the 19 known *onc* sequences, 5 are shared by different taxonomic groups of viruses of which the *mht/raf* homology is the closest determined so far.**

We have previously shown that the activated glucocorticoid receptor of rat liver recognizes defined DNA sequences upstream of the main promoter in the long terminal repeat (LTR) region of mouse mammary tumor virus (MMTV) (1). Together with gene transfer experiments (2), these binding data define an element of about 120 base pairs (bp) that is responsible both for receptor binding and for glucocorticoid regulation of transcription from the viral promoter (3). In DNase I protection experiments we found that the glucocorticoid regulatory element is composed of two strong and two weak binding sites that yield protected regions of different lengths (1). Common to all four binding sites is the hexanucleotide 5′-T-G-T-T-C-T-3′ / 3′-A-C-A-A-G-A-5′, although further sequence homologies are detected between the two strong binding sites (Fig. 1).

To gain further insight into the DNA features responsible for specific receptor binding, we have studied the influence of methylation at particular G residues on receptor binding, as well as the effect of receptor on the accessibility of individual G residues to methylation by dimethyl sulfate. A quantitative analysis of these data was used to calculate the relative affinity of the receptor for each of the individual binding sites, as well as possible interactions between them. Finally, a three-dimensional model is presented describing the interaction of the receptor with the glucocorticoid regulatory element, and a possible mechanism of transcription activation is discussed.

MATERIALS AND METHODS

Materials. ^{32}P-radiolabeled nucleotides were obtained from Amersham at specific activities of 3000 Ci/mmol (1 Ci = 3.7×10^{10} becquerels). DNA polymerase I (Klenow fragment), phage T4 polynucleotide kinase, and bovine alkaline phosphatase were purchased from Boehringer Mannheim. Restriction enzymes were obtained from New England Biolabs (*Rsa* I), Boehringer Mannheim (*Eco*RI), or Bethesda Research Laboratories (*Sau*3AI, *Bam*HI). Dimethyl sulfate (analytical grade) and piperidine were from EGA Chemie (Steinheim, F.R.G.).

Preparation of DNA, Purification of the Receptor, and Filter Binding Assays. The MMTV-LTR deletion plasmids used for these studies are described elsewhere (2) and were a gift of B. Groner and N. Hynes (Karlsruhe, F.R.G.). The deletion plasmid p-236, containing 236 bp upstream of the main transcription initiation site in the LTR, was labeled at the *Bam*HI site, either at the 3′ end or at the 5′ end, as described (1). After digestion with *Rsa* I or *Sau*3AI, end-labeled LTR fragments of either 438 or 123 bp, respectively, were obtained. The *Eco*RI insert of the deletion plasmid p-137 was labeled at the 3′ end and digested with *Rsa* I to yield a 469-bp LTR fragment.

The 90,000 M_r form of the glucocorticoid receptor from rat liver was isolated in the activated state according to previously described procedures (5, 6). Nitrocellulose filter binding studies were performed as described (6).

Methylation Interference Experiments. DNA fragments labeled at one end were methylated in 200 μl of 50 mM sodium cacodylate/1 mM Na_2EDTA, pH 8.0, by addition of 1 μl of 98.9% dimethyl sulfate and incubation for 3 min at 20°C. The reaction was stopped by adding 50 μl of 1.5 M sodium acetate, pH 7.0/1.0 M 2-mercaptoethanol. After two precipitations with ethanol the DNA was incubated with or without receptor in TGA buffer [50 mM Tris·HCl, pH 7.5/10% (vol/vol) glycerol/1 mM 2-mercaptoethanol/0.1 mM Na_2EDTA/0.1 mg of bovine serum albumin per ml] containing 0.1 M NaCl and 1 mM $MgCl_2$. After 40 min at 25°C the reaction mixtures were filtered through nitrocellulose as described (6). DNA fragments bound to protein as well as free and input DNA fragments were used for strand cleavage reactions, followed by electrophoresis on 6.5% or 8% polyacrylamide sequencing gels (7) and autoradiography at −70°C using intensifying screens.

Methylation Protection Experiments. End-labeled DNA fragments were incubated with 50 to 2000 ng of activated glucocorticoid receptor in 200 μl of TGA buffer containing

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.

Abbreviations: MMTV, mouse mammary tumor virus; LTR, long terminal repeat; bp, base pair(s).

*To whom reprint requests should be addressed.

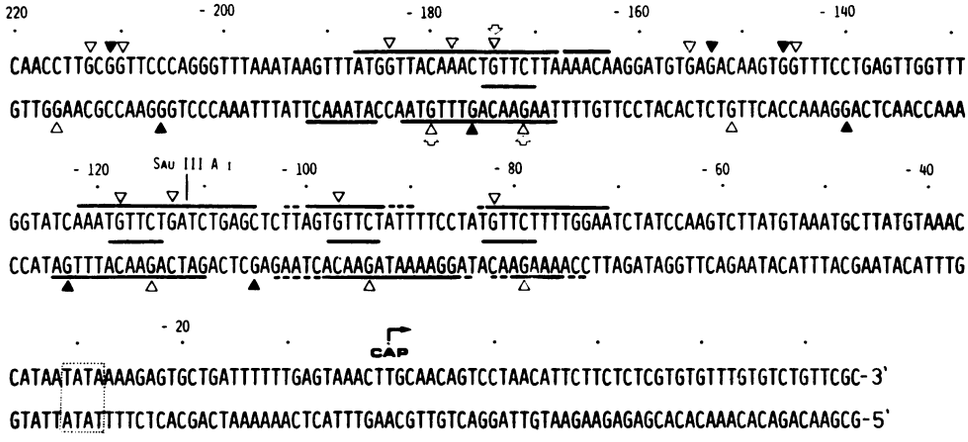

FIG. 1. Nucleotide sequence around the glucocorticoid regulatory element in MMTV-LTR. The nucleotide sequence shown was taken from published reports and confirmed in these experiments (1, 4). The sequences in each strand protected by the glucocorticoid receptor against DNase I digestion are underlined with thick lines. G residues at which methylation prevents receptor binding are indicated by open arrows. Open triangles denote those residues that are protected from methylation in the presence of the receptor, whereas dark triangles indicate G residues whose methylation by dimethyl sulfate is enhanced in the presence of the receptor. The hexanucleotides 5'-T-G-T-T-C-T-3' are marked by solid lines in between the strands. Numbers here and throughout this paper refer to positions upstream of the "cap" site.

60 or 100 mM NaCl and 1 mM $MgCl_2$ at 25°C for 45 min. After chilling on ice for 1 min, 1 μl of native calf thymus DNA (1 mg/ml) and 98.9% dimethyl sulfate to a final concentration of 50 mM were added simultaneously and the incubation was continued at 20°C for 4 min. The reaction was stopped by adding 50 μl of 1.5 M sodium acetate, pH 7.0/1.0 M 2-mercaptoethanol/100 μg of tRNA per ml, and the DNA was precipitated twice with ethanol. In some experiments acid release reactions were performed prior to strand cleavage and electrophoresis on sequencing gels (7, 8). For quantitative evaluation the autoradiograms were scanned with a Desaga microdensitometer and evaluated with a Hewlett–Packard 3390A integrator.

RESULTS

Methylation Interference Experiments. In this type of experiment we analyzed the influence of methylation at the N-7 positions of particular G residues on receptor binding to the glucocorticoid regulatory element. For that purpose end-labeled DNA fragments are partially methylated by dimethyl sulfate and incubated with receptor under conditions that yield about 30–80% binding (6). After incubation, the protein-bound DNA fragments are separated from free fragments by filtration through nitrocellulose filters (6), subjected to strand cleavage reaction at the modified bases, and analyzed in sequencing gels (9). If methylation at a particular G residue interferes with receptor binding, the corresponding band in the autoradiogram is overrepresented in the population of free DNA fragments and underrepresented in the population of protein-bound fragments.

Initial experiments were performed with a DNA fragment containing the intact glucocorticoid regulatory element, and no clear-cut results were obtained (data not shown). We reasoned that this could reflect independent binding of the receptor to the individual binding sites; thus DNA molecules in which an essential G residue at a particular binding site was methylated could still be found in the population of protein-bound fragments due to binding of the receptor to another intact binding site. Similar negative results were obtained with a fragment containing the three receptor binding sites between −72 and −124 and extending up to position −137 upstream of the initiation site. If, however, a fragment was used containing only the strong binding site between −163 and −192, convincing results were obtained (Fig. 2). Methylation of any of the G residues at position −174 in the sense strand and at positions −171 and −180 in the antisense strand interferes with receptor binding. Thus bands at these positions are underrepresented in the population of receptor-bound DNA fragments and overrepresented in the population of free fragments (Fig. 1). Therefore, interaction of the

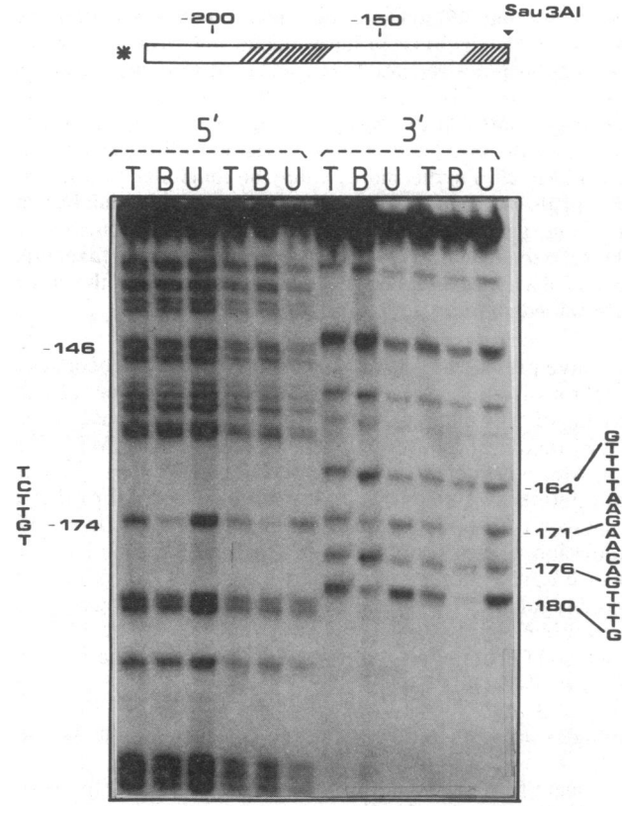

FIG. 2. Methylation interference. Autoradiograms of sequencing gels. The 5'- or 3'-end-labeled 123-bp *Bam*HI/*Sau*3AI fragment from p-236 (2) was partially methylated and incubated with 400 ng (lanes 2, 3, 8, and 9) or 600 ng (lanes 5, 6, 11, and 12) of activated glucocorticoid receptor in 100 μl of TGA buffer. After nitrocellulose filtration, strand cleavage, and electrophoresis in sequencing gels (7), the protein-bound fragments (B) and the unbound fragments (U) were compared with the input DNA (T). Relevant positions are indicated and numbered according to their distance from the cap site. Each set of three lanes corresponds to an experiment. Binding values were 50% (left lanes) and 30% (right lanes).

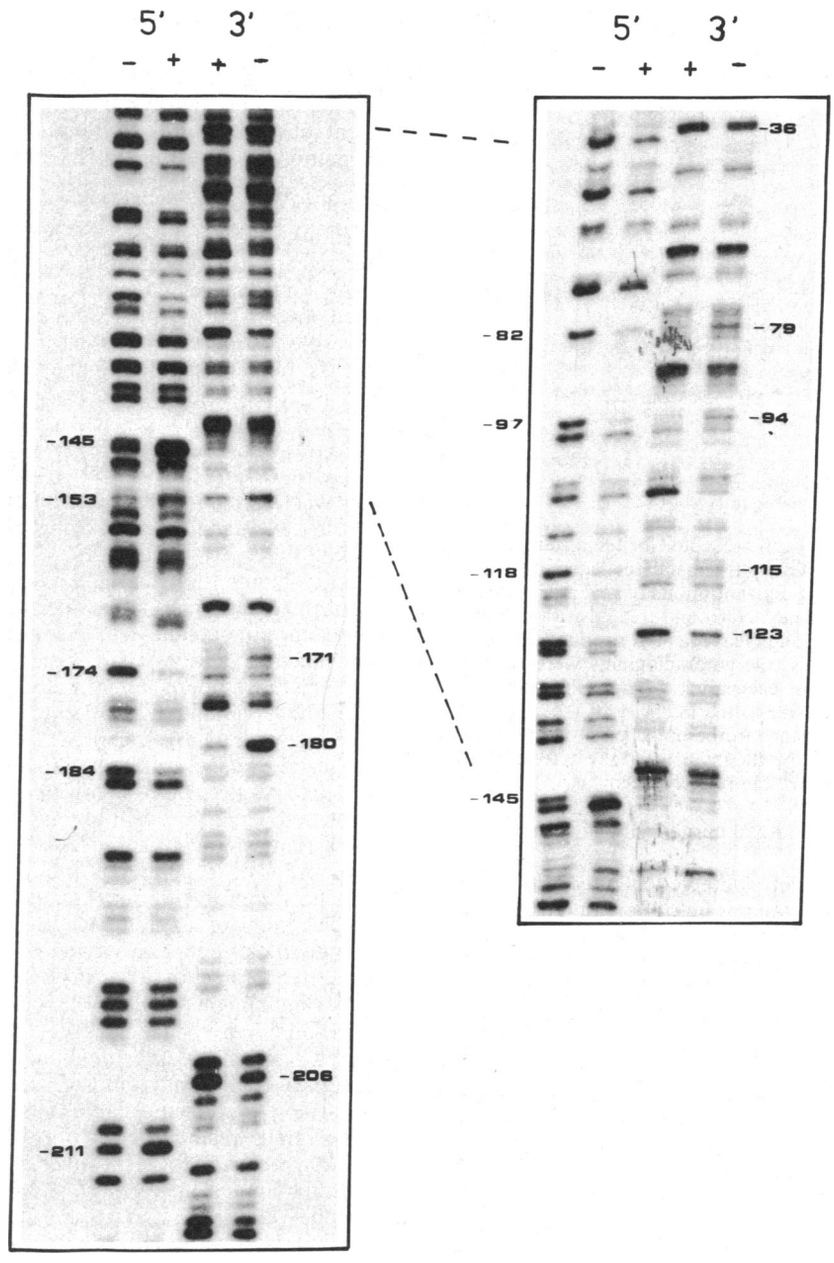

FIG. 3. Methylation protection. Autoradiograms of sequencing gels. The 5'- or 3'-end-labeled *Bam*HI/*Rsa* I fragment from p-236 (2) was incubated with (+) or without (−) 2 μg of activated glucocorticoid receptor prior to methylation and cleavage at the purine residues. The left and right sets of tracks represent different electrophoretic runs of the same samples.

receptor with any of the G residues at these positions is essential for specific binding and can be prevented by methylation at the N-7 position. No other difference in the extent of methylation is found between protein-bound and free DNA fragments, suggesting that no other essential binding sites are located within the DNA fragment used.

Methylation Protection Experiments. These experiments are based on the observation that specific binding of regulatory proteins to DNA prevents methylation of purine residues that are in direct contact with the protein (10). Thus, after the strand cleavage reaction the intensity of the corresponding bands in autoradiograms of sequencing gels is reduced. The influence of receptor binding on methylation by dimethyl sulfate at purine residues is shown in Fig. 3. The most prominent changes in the methylation pattern are detected in the DNase I-protected regions, and they consist mainly of an inhibition of methylation of particular G residues in the presence of the receptor. The extent of methylation protection varies at different sites. For instance, the G residues at position −174 in the sense strand and at −180 in the antisense strand are extensively protected by the receptor. Other affected residues are also labeled in Fig. 3 and summarized in Fig. 1. In all four binding sites the G residues of the hexanucleotide $\begin{matrix}5'\text{-T-G-T-T-C-T-}3'\\3'\text{-A-C-A-A-G-A-}5'\end{matrix}$ are hypomethylated in the presence of the receptor, indicating a direct contact between the protein and the DNA within this conserved sequence element.

In addition, other G residues within the protected regions and in their neighborhood are more methylated in the presence of the receptor. Particularly evident are changes at positions −211 and −146 in the sense strand, and at positions −206, −140, −123, and −105 in the antisense strand (Fig. 3, summarized in Fig. 1). The G methylation pattern is considerably altered between positions −140 and −155, a region that does not show a clear pattern in DNase I protection ex-

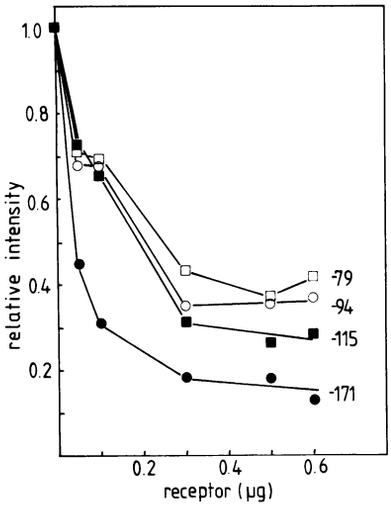

FIG. 4. Relative affinities of the receptor for the different binding sites. The 3'-end-labeled 438-bp fragment (see legend to Fig. 3) was incubated with increasing concentrations of receptor in 300 μl of TGA buffer containing 60 mM NaCl and treated with dimethyl sulfate. After methylation, strand cleavage at G residues, and electrophoresis in sequencing gels, the autoradiograms were scanned with a microdensitometer. For each track the intensities of the G residues were calculated relative to the intensity of the doublet at −36/−37, which does not change in the presence of the receptor. The values were then divided by the intensity of the corresponding G band in the input lane to give the relative intensity.

periments (1) and was unchanged in methylation interference experiments (see above).

No clear-cut changes were detected in the methylation pattern of A residues, with the possible exception of the A at position −178 in the sense strand (Figs. 3 and 1). Since, however, this A is flanked by two other A residues, the observed change in the presence of receptor may not be due to direct contact but may reflect an alteration in base stacking, which is known to influence the accessibility of the central A in a triplet (7).

Relative Affinities of the Receptor for the Different Binding Sites. To estimate the relative affinities of the receptor for the individual binding sites, we performed nitrocellulose filter binding assays with fragments containing the intact glucocorticoid regulatory element and with subfragments thereof. These results showed that the affinity of the receptor for the upstream binding site is higher than its affinity for the downstream subsites (data not shown). In an attempt to quantitate the relative affinities of the receptor for the four different binding sites within the intact glucocorticoid regulatory element, we have measured the influence of receptor concentration on the extent of methylation at positions −79, −94, −115, and −171 in the antisense strand. The selected G residues correspond to equivalent positions within the hexanucleotide $\frac{5'\text{-T-G-T-T-C-T-}3'}{3'\text{-A-C-A-A-G-A-}5'}$ that appears to be equally essential for binding of the receptor to each of the subsites. The results show that the receptor binds to all four subsites with affinities within the same order of magnitude, although clear differences are observed (Fig. 4). The upstream binding site represented here by the G at position −171 is bound with about twice the affinity of the site at −115, and the other two binding sites at −94 and −79 are bound with slightly lower affinities. Even at the lowest receptor concentration, however, we have not found a situation in which only the G at −171 is protected.

DISCUSSION

The results reported above support our previous findings with filter binding and DNase I protection experiments, demonstrating a sequence-specific binding of the glucocorticoid receptor to a regulatory element within the LTR of MMTV (1). In addition, these results allow a more precise analysis of the interaction between receptor and DNA and yield initial insight into the features of the binding sequence relevant for recognition by the receptor. The fact that the G residues in the conserved hexanucleotide $\frac{5'\text{-T-G-T-T-C-T-}3'}{3'\text{-A-C-A-A-G-A-}5'}$ are protected by the receptor against methylation by dimethyl sulfate suggests a direct contact between the protein and the DNA double helix at these sites and confirms the significance of this short sequence for receptor recognition. Similar results have been obtained in receptor binding studies with the human metallothionine IIA gene, which also contains a glucocorticoid regulatory element upstream of its promoter. Within this element we found the hexanucleotide 5'-T-G-T-C-C-T-3', 3'-A-C-A-G-G-A-5'', in which the methylation of two G residues was affected by receptor binding (11).

In a LTR fragment containing only the upstream strong binding site, methylation at the N-7 position of three G residues, including those in the hexanucleotide, prevents binding of the receptor. It is therefore probable that the receptor approaches the double helix from the major groove, and that sequence recognition involves direct contacts with G·C base pairs that are hindered after methylation by dimethyl sulfate.

The observation that no methylation interference could be detected with DNA fragments containing more than one binding site suggests that binding of the receptor to individual sites can take place independently, since methylation at one site does not prevent binding to another site. The relative affinities of the receptor for the different sites are of the same order of magnitude, although the upstream site binds somehow tighter. Our data do not allow us to distinguish between binding of separate receptor entities to each site with a high degree of cooperativity and, alternatively, binding of a single receptor entity, for instance, a tetramer, to all four sites simultaneously but with slight differences in affinity.

In addition to the changes located within the DNase I-protected regions, receptor binding alters the methylation pattern in two adjacent regions centered at positions −210 and −150 (Fig. 1). It is unusual to find protection against methylation in a G residue that is not directly in contact with the binding protein. Since, however, no DNase I protection was detected in these regions (1), and since methylation of the G residues in these regions does not interfere with receptor binding, we tend to interpret the observed changes in methylation pattern as a consequence of indirect structural alterations of the double helix, subsequent to receptor binding at the adjacent sites. Alternatively, cryptic binding sites could be located in these regions that are not detected in filter binding studies or DNase protection experiments but are apparent at the higher receptor concentrations used for methylation protection experiments (10, 12). This view is supported by weak sequence homologies of the region around −150 and the upstream binding site. Interestingly, at −145/−146 we find an altered methylation of two G residues that are located in the heptamer 5'-G-T-G-G-T-T-T-3', a sequence strikingly homologous to the enhancer "core" of DNA tumor viruses (13). It is conceivable that alterations in the helix conformation at this site result in a better recognition of the core analogue by those factors of the transcriptional machinery that usually interact with enhancers. In fact, there is experimental evidence that DNA fragments containing the glucocorticoid regulatory element can act as hormone-dependent enhancers in gene transfer experiments (14).

The data reported here, together with previous results of DNase I protection experiments, allow us to formulate a model for the interaction of the glucocorticoid receptor and the DNA double helix (Fig. 5). In the upstream binding site

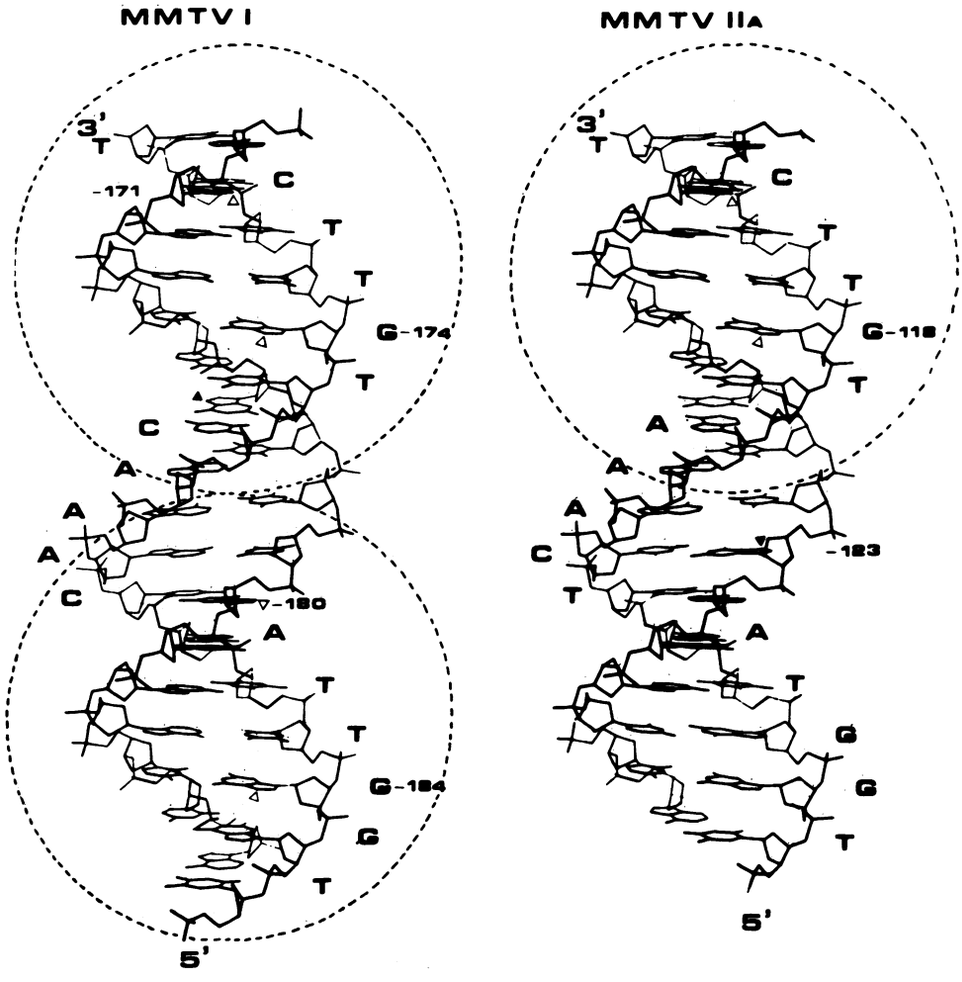

FIG. 5. Computer pictures of two binding sites within the glucocorticoid regulatory element. The nucleotide sequence between −186 and −170 (*Left*) and that between −129 and −114 (*Right*) are represented as double helixes. The purine positions that are protected against methylation are indicated by open triangles, and those that are hypermethylated in the presence of the receptor are marked by dark triangles. The bound receptor molecules are suggested by the dashed circles.

the protected G residues are separated by 10 (±1) base pairs and, therefore, are located on the same face of the helix at two consecutive turns. Since the N-7 position of the G residues is located in the major groove, we can envisage a contact of the receptor with two pairs of G residues through the major groove as depicted in Fig. 5. The limited symmetry of the binding site calls for a similar symmetry of the binding form of the receptor, which could be accomplished by dimerization. A virtually identical type of interaction can be postulated for binding of the receptor to the glucocorticoid regulatory element in the human metallothionein IIA gene (11). The other weaker binding sites in the LTR of MMTV do not exhibit a similar type of symmetry with pairs of G residues located at equivalent positions in two consecutive turns of the helix (Fig. 5). Therefore, they could represent binding sites for a monomeric form of the receptor, an interpretation supported by the shorter length of the corresponding DNase-protected regions (Fig. 1). Alternatively, all four binding sites may interact with a single tetrameric form of the receptor, and the differences in affinities and protection pattern may reflect the lack of an overall symmetry at the DNA level. Further insight into the molecular details of the interaction and its functional significance could be obtained from protein cross-linking experiments as well as from binding studies and gene transfer experiments using point mutations within individual binding sites.

We thank Heinz Bosshard (European Molecular Biology Laboratory, Heidelberg) for the computer pictures of the receptor binding sites and Hannes M. Westphal and Uwe Vaupel (Marburg, F.R.G.) for the receptor preparations. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

1. Scheidereit, C., Geisse, S., Westphal, H. M. & Beato, M. (1983) *Nature (London)* **304**, 749–752.
2. Hynes, N. E., van Oyen, A. J. J., Kennedy, N., Herrlich, P., Ponta, H. & Groner, B. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3637–3641.
3. Groner, B., Ponta, H., Beato, M. & Hynes, N. E. (1983) *Mol. Cell. Endocrinol.* **32**, 101–116.
4. Fasel, N., Pearson, K., Buetti, D. & Diggelmann, H. (1982) *EMBO J.* **1**, 3–7.
5. Wrange, Ö., Carlstedt-Duke, J. C. & Gustafsson, J.-A. (1979) *J. Biol. Chem.* **254**, 9284–9290.
6. Geisse, S., Scheidereit, C., Westphal, H. M., Hynes, N. E., Groner, B. & Beato, M. (1982) *EMBO J.* **1**, 1613–1619.
7. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
8. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
9. Siebenlist, U. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 122–126.
10. Ogata, R. T. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5851–5854.
11. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M. & Beato, M. (1984) *Nature (London)*, in press.
12. Ross, W. & Landy, A. (1983) *Cell* **33**, 261–272.
13. Weiher, H., König, M. & Gruss, P. (1983) *Science* **219**, 626–631.
14. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. (1983) *Cell* **33**, 489–499.