UCSF
UC San Francisco Electronic Theses and Dissertations

Title
A detailed analysis of the DNA-binding affinity and sequence specificity of the glucocorticoid receptor DNA-binding domain

Permalink
https://escholarship.org/uc/item/5dh0j8bg

Author
La Baer, Joshua,

Publication Date
1989

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

copyright 1989

by

Joshua La Baer

Dedicated to

my mother for giving me the confidence and unconditional love that only she can give

my father for giving me the wherewithal and the self-discipline that I so often needed to finish such a protracted education

my sister for her lifelong friendship and for teaching me the art (and the joy) of silliness

and Robyn, my rock, for her unfaltering love, understanding, emotional support and patience.

Acknowledgements

I wish to thank Cynthia Wollberger for her help with the computer graphics, Bruce Malcolm for his expertise in the preparation of the oligonucleotides, Dennis Sakai for providing the plasmid pOTCO, Beth Dombroski, Marc Gartenberg, and Donald Crothers for sharing manuscripts before publication, and especially Leonard Freedman for providing the purified glucocorticoid receptor derivatives used in this study.

I am grateful to Susan Jones and Steve Yoshinaga for tolerating me so generously while we cohabited the same laboratory bay, Jay Thomas for mirth and frivolity, members of the "Annex" for indulging my love of jazz, and all of my colleagues in the Yamamoto laboratory for many useful suggestions and enjoyable discussions.

I would also like to like to offer thanks to David Cox and Ira Herskowitz for advice and guidance, and special thanks to Donald Crothers for many useful discussions and helpful ideas in the development of this project. Peter Walter and Bruce Alberts were an immense help as thesis committee members offering useful suggestions, important crtiticisms, and alternative approaches to daunting problems.

Finally, I will be eternally indebted to Keith Yamamoto for his demanding standards, advice, understanding, encouragement, compulsiveness, suggestions, ideas, corrections, friendship, and, above all, his inspiring dedication to teaching and science.

List o

List o

Abstrac

Introdu

Results

DNA

Hyd

Two

The

Bind

Rela

Coop

Satu

OP-Cu

Effec

Table of Contents

List of Tables 3

List of Figures 4

Abstract 5

Introduction 7

Results 15

DNAase I Footprinting 16

Hydroxyl Radical Footprinting 21

Two Molecules of the Receptor Derivative Bind at the RBS 24

The Receptor Derivative Makes Close Major Groove Contacts at Both Occupied Sites 29

Binding Affinity of the Receptor Derivative for an RBS-Containing DNA Fragment 32

Relative Contribution of Sites a and b Toward Binding Affinity 35

Cooperativity of Protein Binding 42

Saturation Mutagenesis of the RBS 43

OP-Cu Footprinting of Mutants 48

Effects of RBS Substitution Mutations *in vivo* . . 54

Discussion 58

The Receptor Derivative Displays a DNA Binding Pattern Similar to Helix-Turn-Helix Proteins 58

Recognition of the Binding Site Depends on 4 Critical Base Pairs 61

Cooperativity Aids Binding at Site b 63

Recognition Sequences and Context Both Contribute to Protein Binding 67

Methods and Materials 72

End Labeling and Purification of DNA Fragments . . 72

DNAase I Footprinting 73

Hydroxyl Radical Footprinting 74

Electrophoresis in Nondenaturing Polyacrylamide Gels 75

1,10 Phenanthroline Copper Footprinting 76

Methylation Interference 77

Absolute Affinity Constant Determination 78

Relative Affinity Constant Determination 80

Saturation Mutagenesis 82

Plasmid Sequencing 83

Cell Culture and Transfections 86

Extract Preparation and CAT Assays 87

Bibliography 89

List of Tables

Table 1. Affinity Constant Quotients for Several RBS Mutants 41

Table 2. Distribution of Mutations in the Saturation Mutagenesis 45

Table 3. The Effects of Point Mutations in the RBS on the Affinity of T7X556 47

Table 4. Dexamethasone-Induced CAT Expression in Transient Transfection Assays by RBS Dimers . . . 57

Table 5. Glucocorticoid Receptor Binding Sequences . . 64

List of Figures

Figure 1. Construction of RBS Derivatives 17

Figure 2. DNAase I and Hydroxyl Radical Footprinting Analysis of T7X556 Binding to RBSwt-256 18

Figure 3. Summary of T7X556 Footprinting of the RBS Using Different DNA-cleavage Agents 20

Figure 4. Glucocorticoid Receptor Derivatives Retard the Mobility of RBSwt-256 in a Nondenaturing Polyacrylamide Gel 26

Figure 5. 1,10 Phenanthroline-Copper Ion Complex Footprinting Analysis of T7X556 Binding to RBSwt-256 28

Figure 6. Methylation Interference of T7X556 Binding to RBSwt-256 31

Figure 7. Determination of Equilibrium Binding Constants K_1 and K_2 for T7X556 Binding to RBSwt-256 34

Figure 8. Assay to Determine the Relative Affinity of T7X556 for Different DNA Fragments 37

Figure 9. Footprinting Analysis of T7X556 Binding to RBS Substitution Mutants 50

Figure 10. Mutations That Ablate Site b or That Balance the Two Consensus Sequences Do Not Shift the Footprints 52

Figure 11. Model for Sequence-Specific Binding by T7X556 68

A Detailed Analysis of the DNA-Binding Affinity and Sequence Specificity of the Glucocorticoid Receptor DNA-Binding Domain

Joshua La Baer

Abstract

The glucocorticoid receptor transduces hormonal signals to the transcriptional apparatus by binding to specific regulatory DNA-sequences located near responsive promoters. Receptor derivatives that contain the recptor DNA-binding domain can protect these regulatory sequences, termed glucocorticoid response elements (GREs), from cleavage by nucleases or by chemicals in vitro. In these experiments, I mapped the protein contacts of one such receptor derivative, T7X556, with a receptor binding site (RBS) by testing the protection afforded by the protein to four different DNA-cleavage agents. Even though the RBS lacked sequence symmetry, two molecules of T7X556 bound symmetrically to it. Each bound to one face of the DNA double helix, centered over the major groove, where it makes close contacts. One molecule bound with high affinity to a hexanucleotide sequence that matches the consensus sequence for receptor binding, and the other bound with lower affinity to a cryptic pseudosite adjacent to the first. Using a quantitative comparison assay to study protein binding, I

determ
sites
inform
the bi
demons
cooper
for fu

In
nucleot
the RBS
nucleot
specifi
at thes
were fu
vivo in
protein
sequence
even whe
exactly
specific
binding
Presumab
required

determined the relative affinities of these two binding sites and compared them to nonspecific sites. This information, along with the absolute affinity constants for the binding of each protein molecule, allowed me to demonstrate that protein binding is improved ~150 fold by cooperativity. Moreover, both binding sites are necessary for function in vivo in a transient transfection assay.

In order to determine the relative contribution of each nucleotide in a binding site towards binding, I saturated the RBS with point mutations. I found that four of these nucleotides, 5' -G-ACA- 3', are likely to make strong specific contacts with the protein since certain mutations at these positions severely reduced binding. These results were further substantiated in a footprinting assay and in vivo in a transient transfection assay. Paradoxically, the protein did not bind to a nearly identical hexanucleotide sequence only 5 bp away from the high affinity binding site, even when the nonbinding sequence was mutated to match exactly the high affinity site. Thus, in addition to the specific recognition sequences, sequences outside the binding site must also contribute significantly to binding. Presumably, such sequences create an appropriate "context" required for the protein to complex with the binding site.

Introduction

Glucocorticoid response elements (GREs) are specific DNA sequences that confer steroid-dependent enhancement of transcription upon linked promoters in responsive cells (Chandler et al., 1983). GREs are selectively recognized in vivo (Miesfeld et al., 1986) and in vitro (Payvar et al., 1981, 1983; Scheidereit et al., 1983) by the glucocorticoid receptor, an intracellular protein that transduces the hormonal signal to the transcriptional apparatus (for review, see Yamamoto, 1985). In the absence of hormone, the receptor resides primarily in the cytosol; the binding of hormone facilitates the alteration of the protein, a poorly understood process termed "transformation," which results in the translocation of the hormone-receptor complex to the nucleus (Picard and Yamamoto, 1987) where it binds to the GREs and induces changes in the nearby chromatin (Zaret and Yamamoto, 1984).

Manipulation and expression of the gene encoding the 795 amino acid rat glucocorticoid receptor revealed that the determinants for several protein functions, including hormone binding, DNA binding, nuclear localization, and transcriptional regulation, are organized into identifiable domains (Rusconi and Yamamoto, 1987; Miesfeld et al., 1987, 1988; Picard and Yamamoto, 1987), reflecting in part the organization of the genomic sequence (M. Jacobson, personal

communication). In the appropriate context, each functional domain can act independently; thus, fusion of the transcriptional enhancement determinants located in the N-terminal third of receptor to the DNA-binding domain of the E. coli LexA repressor permits lex operator-specific transcriptional enhancement (Godowski et al., 1988), and fusion of the hormone-binding determinants, located in the C-terminal third of the receptor, confer steroid regulation on the unrelated adenovirus E1A gene product (Picard et al., 1988).

In cells that lack endogenous receptor, a 150 amino acid receptor fragment comprising amino acids 407-556 mediates GRE-specific transcriptional enhancement, demonstrating that the DNA-binding determinants are contained within this genetic segment (Miesfeld et al., 1987). Deletion studies have further delimited the DNA-binding domain to amino acids 440-525 (Rusconi and Yamamoto, 1987; Miesfeld et al., 1987). In fact, an E. coli-expressed polypeptide including only amino acids 440-525 suffices for DNA-binding in vitro (L. Freedman, personal communication). The amino acid sequence of this region displays considerable homology to the DNA-binding domains of other ligand-activated intracellular receptors, emphasizing the importance of this region to the function of these proteins and lending cogent support to the notion of a nuclear receptor superfamily (for review, see Evans, 1988).

Alignm
these
protei
for th
et al.
contai
coordi
two cy
analys
recept
throug
that m
(Freed

Th
is unkn
and rec
several
to DNA
substru
TFIIIA,
(Rhodes
TFIIIA a
zinc-bin
or even
two of t
with his

Alignment of the different receptor sequences reveals that these proteins share homology with a class of DNA-binding proteins containing the "zinc finger" motif first described for the Xenopus 5S rRNA transcription factor TFIIIA (Miller et al., 1985). All of the receptor sequences studied so far contain two "finger" homologies, each predicted to coordinate a zinc atom with four cysteines instead of the two cysteines and two histidines found in TFIIIA. Physical analysis of the DNA-binding domain of the rat glucocorticoid receptor confirmed that it coordinates two zinc atoms, each through a tetrahedral arrangement of four cysteines, and that metal coordination is essential for DNA binding (Freedman et al., 1988).

The mechanism of DNA binding by glucocorticoid receptor is unknown. Indeed, the "zinc finger" motif was identified and recognized as a DNA-binding domain only within the last several years, and there is little specific data pertaining to DNA binding for any of the proteins that contain this substructure. Klug and Rhodes have proposed two models for TFIIIA, which contains nine repeated Zn-binding domains (Rhodes and Klug, 1986; Fairall et al., 1986). (Whether TFIIIA and steroid receptors, which have significantly fewer zinc-binding domains, share a common DNA-binding mechanism or even a common structure is also not known. Thus, when two of the putative zinc-coordinating cysteines are replaced with histidines in the estrogen receptor, the protein loses

its a
first
groov
the p
succe
groov
"finge
major
"finge
groove

A
protei
mechani
exhaust
structu
as dime
They usu
recogniz
groove a
The heli
consist
helix,"
bonds wit
α-helix t
sugar-pho
acid-DNA

its activity in vivo; Green and Chambon, 1987.) In the first model the successive "fingers" follow along the major groove around the double helix, while in the second model the protein lies on one face of the double helix while successive "fingers" point alternately into the major grooves on either side. Both models predict that a single "finger" associates with each half-turn of the helix in the major groove; however, others have suggested that multiple "fingers" may bind at a single exposed region of the major groove (Berg, 1988; Vrana et al., 1988).

At present, the helix-turn-helix class of DNA-binding proteins has provided us with the greatest detail on the mechanism of sequence-specific DNA binding - the result of exhaustive genetic analysis as well as crystallographic structure determination. Typically, these proteins bind DNA as dimers at fully or partially symmetric operator sites. They usually bind DNA on one face of the double helix, recognizing the operator through contacts in the major groove at two successive half-turns of the double helix. The helix-turn-helix domains of the monomeric subunits consist of an α-helical region termed the "recognition helix," which lies in the major groove and forms hydrogen bonds with the bases there, connected by a turn to a second α-helix that lies across the major groove and contacts the sugar-phosphate backbone. As expected, the specific amino acid-DNA contacts and the orientation of the recognition

helix in the major groove vary from protein to protein (for reviews, see Pabo and Sauer, 1984; and Schleif, 1988).

Detailed analyses of operator sites have proved invaluable in the study of the helix-turn-helix DNA-binding proteins. The identification by genetic selection of severe and moderate operator mutants implicated contacted bases, and the subsequent recovery of suppressor mutations in the protein aided the identification of the corresponding amino acid contacts (Youderian et al., 1982; Ebright et al., 1984). The travail of a similar genetic selection in animal cells has delayed the availability of a comparably detailed analysis of GREs; nevertheless, they have been partially characterized. Footprinting studies using isolated glucocorticoid receptor and GRE-containing DNA fragments initially revealed protected sequences varying in length from 15 to 44 bp that lacked extensive homology. All of the footprints did exhibit at least one occurrence of a sequence related to the degenerate consensus 5'-AGA(A/T)CA-3' (Scheidereit et al., 1983; Payvar et al., 1983), whose importance was verified by mutagenesis (DeFranco and Yamamoto, 1986; O. Wrange and D. DeFranco, unpublished results). The subsequent observation that some GRE-containing fragments have partially symmetric permutations of this degenerate consensus (resembling 5'-AGA(A/T)CANNNTG(A/T)TCT-3') implied that the protein can bind DNA in a head to head fashion and led to the suggestion

that the consensus be expanded correspondingly (Scheidereit et al., 1986; Jantzen et al., 1987). Similar dyad symmetry has also been observed for other hormone response elements (HREs; Klein-Hitpass et al., 1988) that are themselves closely related to GREs in sequence (Martinez et al., 1987; Klock et al., 1987).

While these consensus sequences have helped identify potential receptor binding sites, their considerable degeneracy has limited their usefulness as predictive tools. Thus, a single partial palindrome can suffice to mediate GRE activity in vivo sometimes (Jantzen et al., 1987; Strähle et al., 1987), but not always (Jantzen et al., 1987; Martinez et al., 1987), though in both cases it appears to match the consensus. This reflects a limitation of consensus sequences in general: they do not result from systematic and quantitative comparisons, but instead are usually derived from modest data sets that are further limited by evolutionary constraints, such as non-random base pair usage and potentially incomplete divergence from common progenitor sequences. Thus, a consensus sequence represents only a summary of functionally-related sequences that typically reside in dissimilar contexts.

Indeed, in addition to the consensus sequence itself, it is important to consider the context of that sequence. Unrecognized features of the surrounding sequences may determine the degree to which a given consensus can

functio
the act
Althoug
than th
know wh
noncont
observe
1988).
in the
GRE fun
needed.

In
by the
interac
fragmen
contacts
variety
the avid
I used e
to separ
protein'
the bind
affinity
site, I c
itself.
each nucl

function. In the extreme, a particular context might negate the activity of an otherwise sufficient binding site. Although the effects of context are more difficult to study than those of consensus sequences, (because it is harder to know where to look), some significant effects of noncontacted or distantly located sequences have been observed (Koudelka et al., 1987; Gartenberg and Crothers, 1988). Therefore, to identify the specific contacted bases in the GRE, as well as assess the importance of context to GRE function, a systematic and comprehensive approach was needed.

In order to better understand sequence-specific binding by the glucocorticoid receptor, I studied in detail the interaction between a single binding site and a DNA-binding fragment of the receptor. Initially, I mapped the protein contacts on the binding site by footprinting the DNA using a variety of DNA cleavage agents. I then sought to determine the avidity and selectivity of the protein-DNA interaction. I used electrophoresis in a nondenaturing polyacrylamide gel to separate bound from free DNA in order to ascertain the protein's absolute affinity for a DNA fragment containing the binding site. By then comparing the protein's relative affinity for the DNA fragment with and without the binding site, I computed the intrinsic affinity of the binding site itself. To examine the relative binding contribution of each nucleotide, I saturated the binding site with

substitution mutations and compared the affinity of the receptor for the mutant and wild type fragments. Finally, I tested the predictions of the affinity analysis by footprinting a subset of the mutants and assaying their function *in vivo*.

Results

I initiated my study of the interactions between the glucocorticoid receptor and its cognate sites on DNA by mapping the bases protected by bound protein after subjecting the protein-DNA complex to a variety of cleavage agents. To facilitate my study, I employed a derivative of the receptor that is readily expressed and purified from E. coli (Freedman et al., 1988). This derivative, which encompasses the entire DNA binding domain (amino acids 407-556, denoted T7X556), retains the ability, albeit with reduced magnitude compared to the intact receptor, to activate GRE-linked promoters in mammalian cells either when expressed in vivo from an appropriate expression vector (Miesfeld et al., 1987) or when introduced as purified protein by lipofection (L. Freedman, personal communication). On a variety of DNA fragments tested, the purified protein binds to the same GRE sequences as the full length protein (data not shown). I reasoned that the reduced size of T7X556 (ca. 19kD), as compared with the intact receptor (ca. 90kD), would allow me to map the protein-DNA contacts with finer precision. Indeed, in DNAase I footprinting experiments, the former protects about 7 fewer bases per footprint (data not shown).

With the intention of later performing a mutational analysis, I desired the shortest possible DNA sequence that

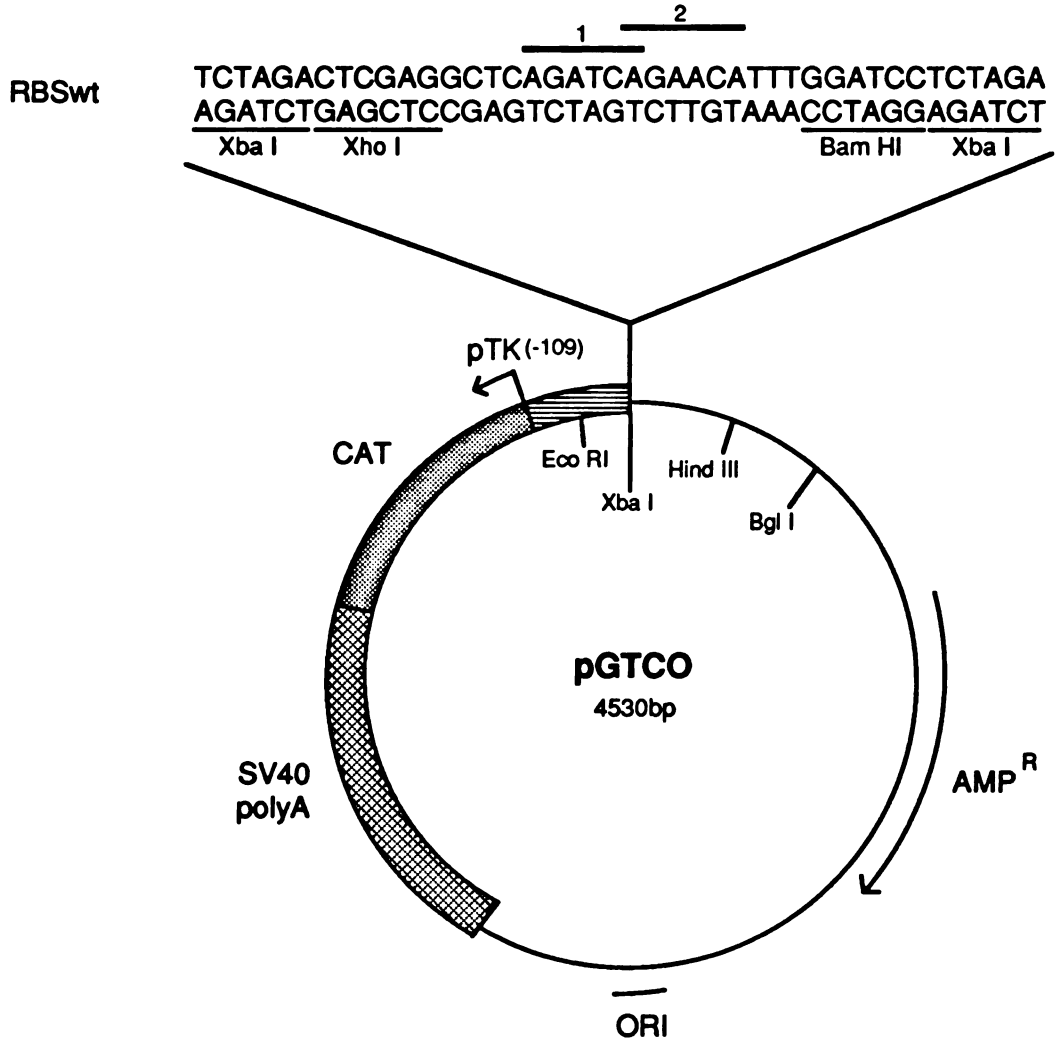
could provide a strong footprint. The binding site that I chose derives from one of several well-defined footprints that had been identified in the 5' LTR of murine mammary tumor virus (MTV, bases -122/-105 relative to the transcriptional start; Payvar et al., 1983). I found this sequence particularly intriguing because it contains two consensus sequences (see Figure 1), either or both of which could contribute to binding. Unlike some identified binding sites (Jantzen et al., 1987; Martinez et al., 1987; Klock et al., 1987), this sequence lacked dyad symmetry; indeed, the two consensus sequences, which overlap at one base, are both oriented in the same 5' - 3' direction. This receptor binding site (RBS) was chemically synthesized along with the flanking restriction sites (to allow both unidirectional and bidirectional cloning) and inserted next to a chloramphenicol acetyltransferase (CAT) gene that is driven by the TK promoter (-109; McKnight et al., 1981).

DNAase I Footprinting

When a 256 bp DNA fragment containing the RBS (RBSwt-256, Figure 1), was subjected to cleavage by DNAase I after incubation with T7X556, I observed 22 protected phosphates on each strand (see Figures 2a and 3). The center of the footprint lies near the T-A base pair at position 17 and at both ends the top strand is shifted three bases 3' relative to the bottom strand. The footprint extends past the MTV-derived sequences, completely covering the Bam HI site and

Figure 1. Construction of RBS Derivatives

Sequences from the 5' LTR of MTV, (-122/-105 relative to the transcriptional start), were chemically synthesized along with the indicated flanking restriction sites and linked to the TK promoter driving the CAT gene. Two overlapping sequences that match the observed consensus for glucocorticoid receptor binding are indicated by overlining. The sequences of the clustered substitution mutant RBS490, (substituted bases are indicated by filled circles), and of the RBS dimers used in the in vivo experiments are also shown. Specific fragments used in these experiments include: RBSx-256, Eco RI/Bgl I; ΔRBS-220, Eco RI/Bgl I without the Xba I insert; RBSx-102, Eco RI/Hind III.



RBS490

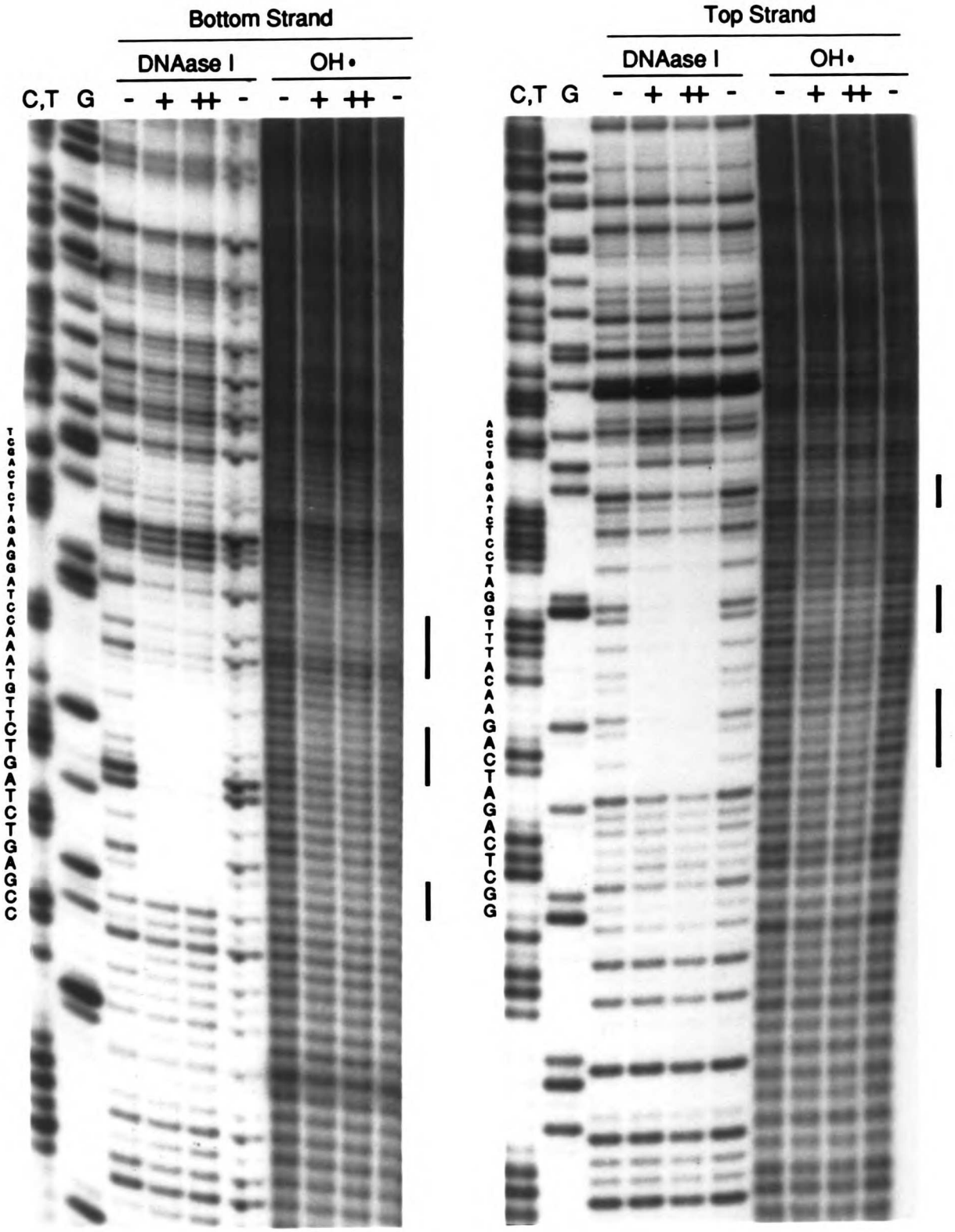
```
                                                ●●●●●
TCTAGACTCGAGGCTCAGATCAGAACATTTCTCAGCTCTAGA
AGATCTGAGCTCCGAGTCTAGTCTTGTAAAGAGTCGAGATCT
Xba I   Xho I                              Xba I
```

RBS-DIMER

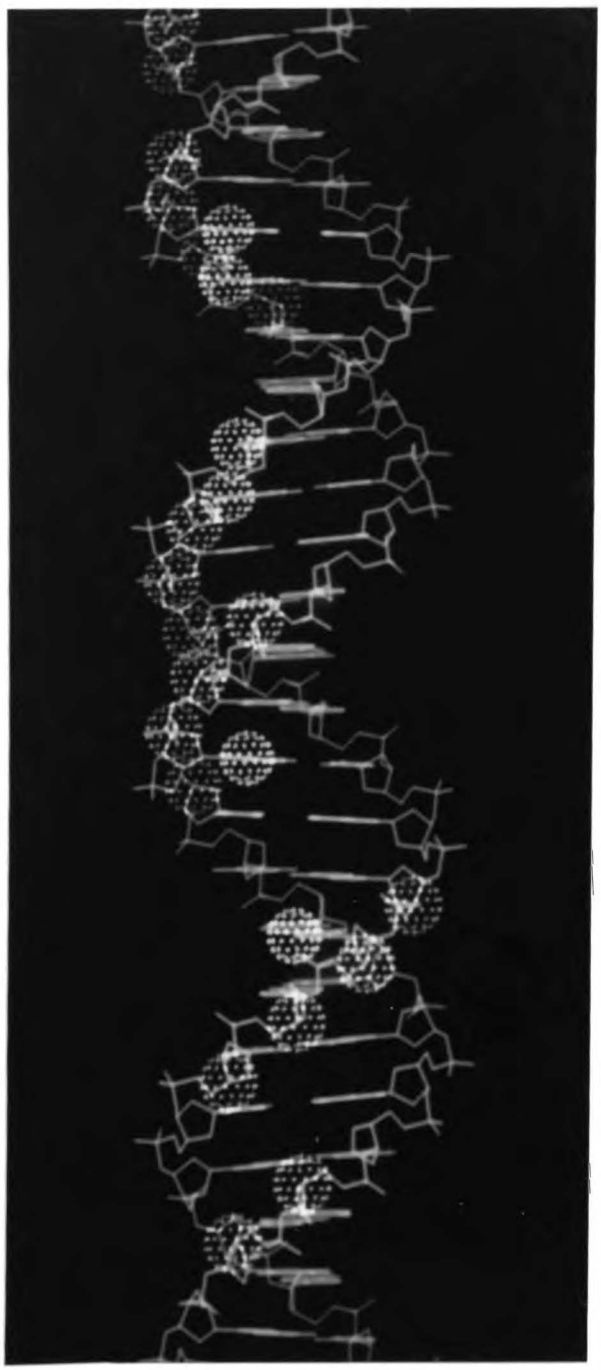
```
TCTAGAGGATCCAAA...AGCCTCGAGGCT...TTTGGATCCTCTAGA
AGATCTCCTAGGTTT...TCGGAGCTCCGA...AAACCTAGGAGATCT
Xba I   Bam HI          Xho I           Bam HI  Xba I
```

Figure 2. DNAase I and Hydroxyl Radical Footprinting Analysis of T7X556 Binding to RBSwt-256

a) An RBS-containing 256 bp fragment (Eco RI/Bgl I) was 3' end labeled (bottom strand) or 5' end labeled (top strand) at the Eco RI site and footprinted with 0 ng (-), 400 ng (+) or 600 ng (++) of T7X556 using the indicated DNA-cleavage agent. Bases that are reproducibly protected from hydroxyl radical cleavage are indicated by solid bars. Lanes G and C,T are probe DNA submitted to guanine-specific or cytosine- and thymidine-specific cleavage, respectively.


Bottom Strand
DNAase I
OH•
C,T G - + ++ - - + ++ -
TGACTCTAGAGGATCCAAATGTTCTGATCTGAGCC
Top Strand
DNAase I
OH•
C,T G - + ++ - - + ++ -
AGCTGAGATCTCCTAGGTTTACAAGACTAGACTCGG

b) The RBS is displayed here using computer graphics to illustrate the protection from hydroxyl radical cleavage (red spheres indicate the 4' carbon of the protected ribose group) and methylation interference data (green spheres indicate the N-7 of guanines that inhibit binding when methylated). Two different rotations around the central axis are shown to demonstrate the two protected faces of the double helix.



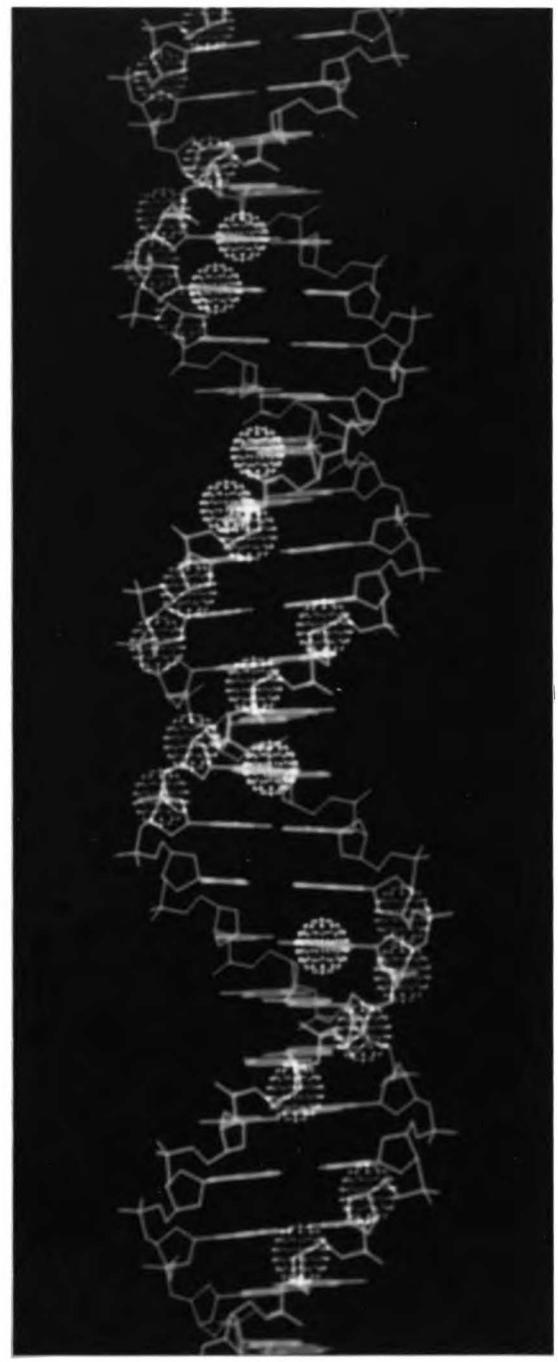
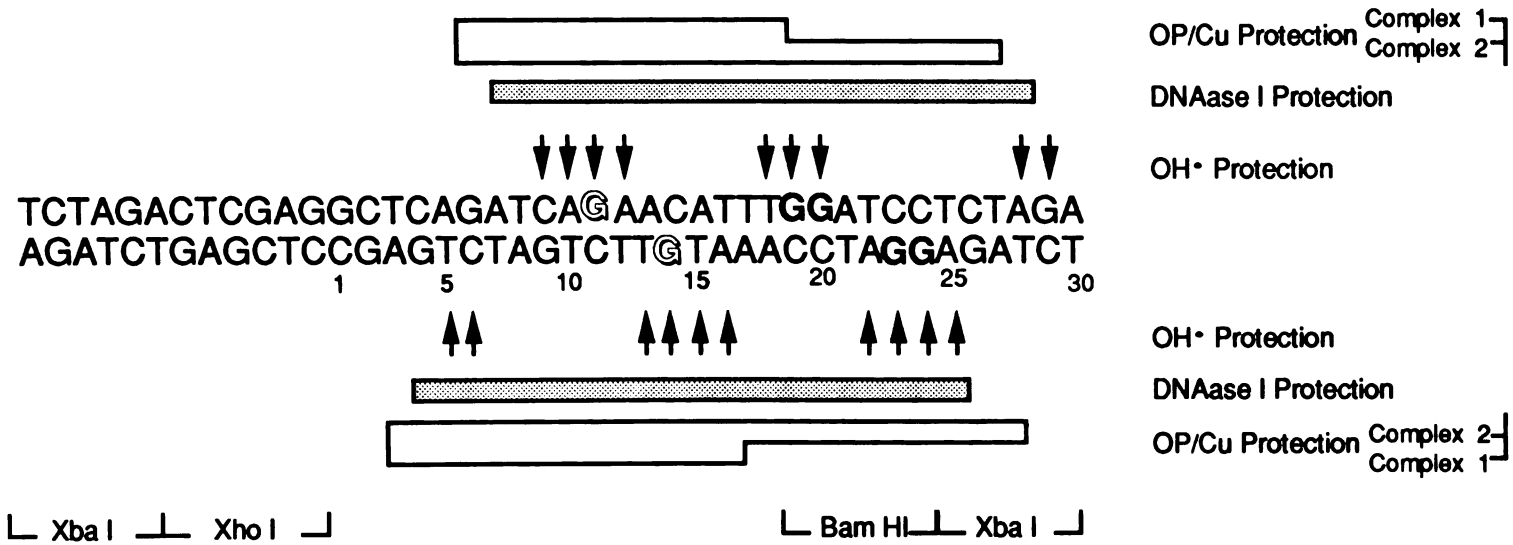


Figure 3. Summary of T7X556 Footprinting of the RBS Using Different DNA-cleavage Agents

Bases protected from DNAase I and hydroxyl radical cleavage are indicated by the filled boxes and the arrows, respectively. Bases protected from 1,10 phenanthroline-copper ion complex (OP-Cu) are indicated by open boxes with the thinner region indicating bases protected in complex 2 only. DNA fragments methylated at the N-7 position of guanines represented in outline are absent from both complexes, while those methylated at guanines represented in bold are absent from complex 2 only. Throughout this manuscript, the base pairs of the RBS are referred to by number starting with the first MTV-derived base pair (as indicated in the figure). Unless otherwise specified, substitution mutants are referred to by number and the new base in the top strand, e.g. 11C is a mutation of guanine to cytosine at position 11.

Summary of Footprint Data



reaching partly over the Xba I site; however, the protection over these non-MTV sequences is not complete. The sharp footprint borders on both strands near the Xho I side of the footprint provide useful information about the positioning of T7X556 on the double helix.

DNAase I has an extended DNA-binding surface with contacts along one face of the double helix (Suck and Oefner, 1986; Suck et al., 1988). In addition to its contacts at the site of cleavage, it also contacts the uncut strand 4-5 bp away in either direction. In the 3' direction from the cleavage site, this occludes the major groove for 5 bp, in the 5' direction it blocks the minor groove for 4 bp. As depicted in Figure 3, the nearest unprotected cleavage in the top strand of the RBS occurred at G6, indicating that DNAase I can occupy the major groove for 5 bp in the 3' direction from this position, i.e. between G6 (top strand) and C11 (bottom strand). By this argument, the protection at A7 implies that T7X556 begins occupying the major groove near C11. Using similar reasoning for the bottom strand, the nearest unprotected cleavage at G4 suggests that DNAase I can occupy the minor groove up until T8 (top strand) where T7X556 presumably begins to cover the minor groove. Hence, the DNAase I data suggest that T7X556 has a boundary in the minor groove near T8 (top strand) and a boundary in the major groove near C11 (bottom strand).

Hydroxyl Radical Footprinting

To obtain more detailed information about protein-DNA contacts I footprinted the RBS using hydroxyl radical cleavage. This technique exploits the highly reactive hydroxyl radical to cleave the DNA fragment by abstracting a hydrogen atom from the deoxyribose sugars along its backbone (Tullius and Dombroski, 1986). Unlike most cleavage reagents, the hydroxyl radical, which is the size of a water molecule and which demonstrates almost no sequence-dependence, is small enough to be able to cut exposed sites on the "backside" of a DNA helix to which a protein is bound (Tullius et al., 1988). The free radical was generated using the Fenton reaction, in which iron(II) reduces hydrogen peroxide to give hydroxyl radical.

In my initial experiments with hydroxyl radical cleavage, I did not observe any footprinting by T7X556 on the RBS even under binding conditions identical to those used in DNAase I footprinting. After testing each of the hydroxyl radical reagents in DNAase I footprint reactions, I determined that quantities of hydrogen peroxide normally used to achieve appropriate DNA cleavage (0.03% final) inhibited DNA-binding by T7X556. This apparent hypersensitivity to hydrogen peroxide was also observed with the transcription factor TFIIIA (Tullius et al., 1988). Since both of these DNA-binding proteins contain the "zinc finger" structural motif, I speculate that the hydrogen peroxide may disrupt this structure, perhaps by oxidizing

the cysteine residues that coordinate the zinc atom. Nevertheless, by reducing the concentration of the hydrogen peroxide in the reaction, and by initiating the reactions without premixing the reagents (in order to maximize exposure of the protein-DNA complex to the short-lived free radicals; see Materials and Methods), I achieved weak but reproducible footprinting.

T7X556 protected three separate clusters of bases on each of the strands as shown in Figure 2a. Despite the lack of sequence symmetry, it can be seen in the computer graphic representation of these protections (Figure 2b) that the footprint is symmetric about an axis near the TA base pair at 17. While each half of the footprint lies on one face of the double helix, the two halves of the footprint are rotated about one third of a turn around the double helix relative to each other. In each half of the footprint, T7X556 covers the minor groove at the outside of the footprint, the major groove moving inward, and the minor groove again at the center. The protein boundary mapped by hydroxyl radical footprinting agrees nicely with that predicted by the DNAase I footprinting. Thus, the protein begins protecting the top strand at T8 over the minor groove and passes closely over the base pair at position 11 in the major groove. Notably, neither solution footprint assay provided evidence for protection of the consensus sequence numbered 1 in Figure 1. Indeed, the unprotected cutting of

the top strand at A5 is inconsistent with T7X556 binding over that consensus since DNAase I will occlude the major groove over this region.

Two Molecules of the Receptor Derivative Bind at the RBS

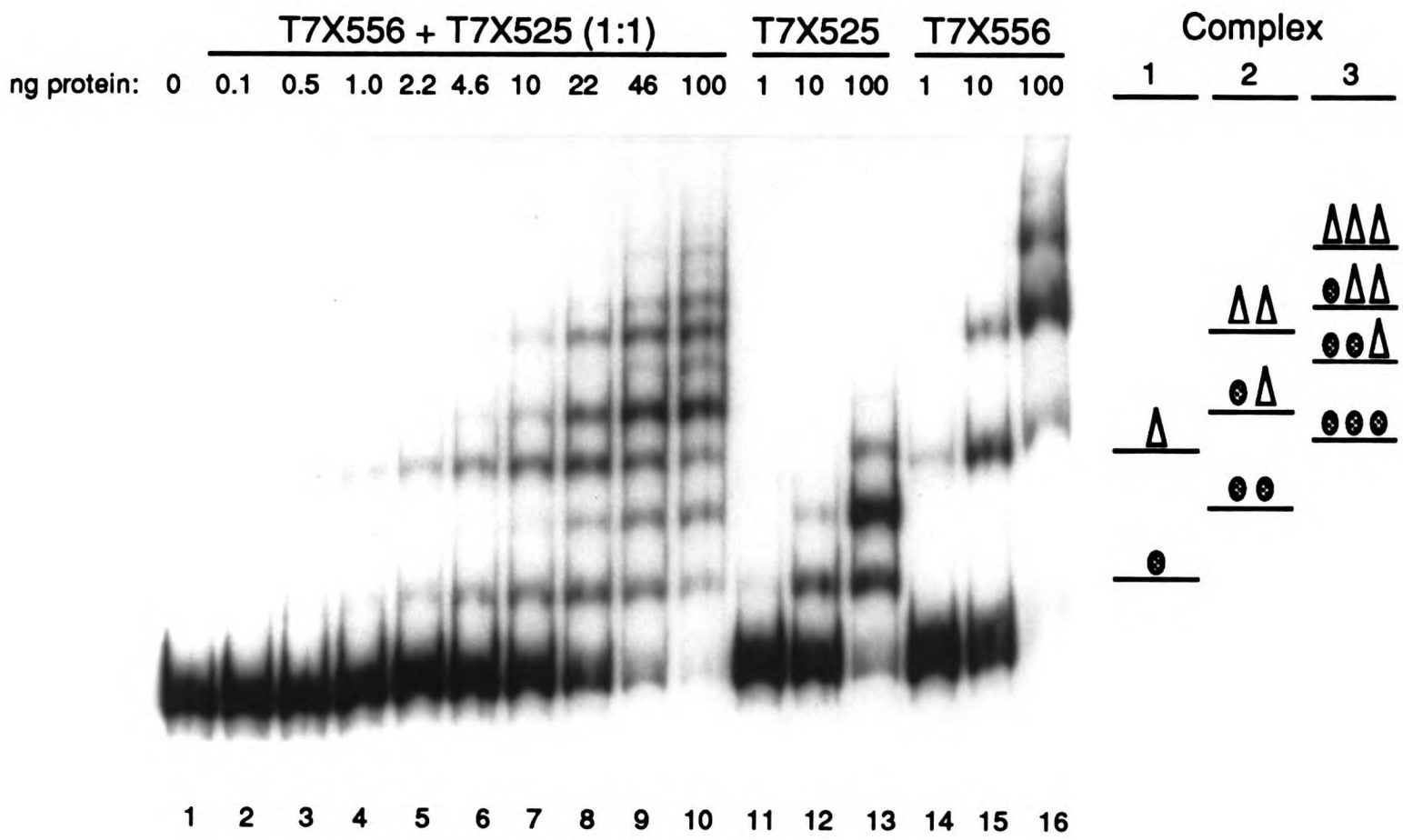
The symmetry of the footprint, particularly considering the protection of two slightly offset faces of the helix, suggested the possibility that this footprint represented two molecules of T7X556 bound at the RBS. Others have suggested the possibility that steroid receptors bind to their cognate sites in a head to head fashion (Scheidereit et al., 1986; Jantzen et al., 1987; Martinez et al., 1987); however, in those experiments the suggestion was prompted by the observation that the binding sites displayed dyad symmetry with each half containing an apparent consensus sequence. While we might expect two proteins to bind to such symmetric sites, it would be interesting if this also occurred at a site which lacked obvious symmetry. To determine the number of protein molecules bound at the RBS and identify their boundaries, I employed a technique that allows footprinting of resolved protein-DNA complexes <u>in situ</u> in a nondenaturing polyacrylamide gel using the nuclease activity of 1,10-phenanthroline-copper ion complex (OP-Cu; Kuwabara and Sigman, 1987). To accomplish this, I first needed to establish the identity of the resolved complexes in the nondenaturing gel.

Incubation of RBSwt-256 with increasing concentrations

of T7X556 followed by electrophoresis on a nondenaturing polyacrylamide gel produced a series of sequentially retarded protein-DNA complexes (Figure 4, lanes 14-16). I demonstrated that these retarded species represented one, two, three, etc. protein molecules bound per fragment by using a 1:1 mixture of T7X556 and a slightly smaller (15kD vs. 19kD) DNA-binding derivative of the receptor, T7X525, in the shift assay. T7X525 also retards the mobility of the DNA fragment into several discrete species, which have mobilities distinct from those produced by T7X556 (Figure 4, lanes 11-13). At low concentrations of the 1:1 protein mixture, only two retarded species were observed, each corresponding to the least-retarded species observed with T7X556 or T7X525 alone (compare lane 4 with 11 and 14) and thus representing one protein molecule bound per fragment (referred to as complex 1). With increasing concentration three additional species appeared, the fastest and slowest of which correspond to two molecules of either T7X525 or T7X556 per fragment (referred to as complex 2), respectively (compare lane 7 with 12 and 15). The novel intermediate species represents RBSwt-256 bound to one molecule each of the two proteins and, as expected, was present in roughly twice the abundance of the other two complex 2 species. The absence of an intermediate species at the lower concentration verified that the two complex 1 species indeed represent one protein molecule per fragment, since at least

Figure 4. Glucocorticoid Receptor Derivatives Retard the Mobility of RBSwt-256 in a Nondenaturing Polyacrylamide Gel

A 256 bp end-labeled fragment containing the RBS was incubated with the indicated amounts of either T7X556 (lanes 14-16) or T7X525 (lanes 11-13) or a 1:1 (w:w) mixture of both (lanes 2-10) and then resolved on a nondenaturing polyacrylamide gel. The identities of the different protein-DNA complexes are schematized on the right where T7X525 is represented by filled ovals and T7X556 is represented by open triangles.


T7X556 + T7X525 (1:1)
T7X525
T7X556
Complex
ng protein: 0 0.1 0.5 1.0 2.2 4.6 10 22 46 100 1 10 100 1 10 100
1 2 3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

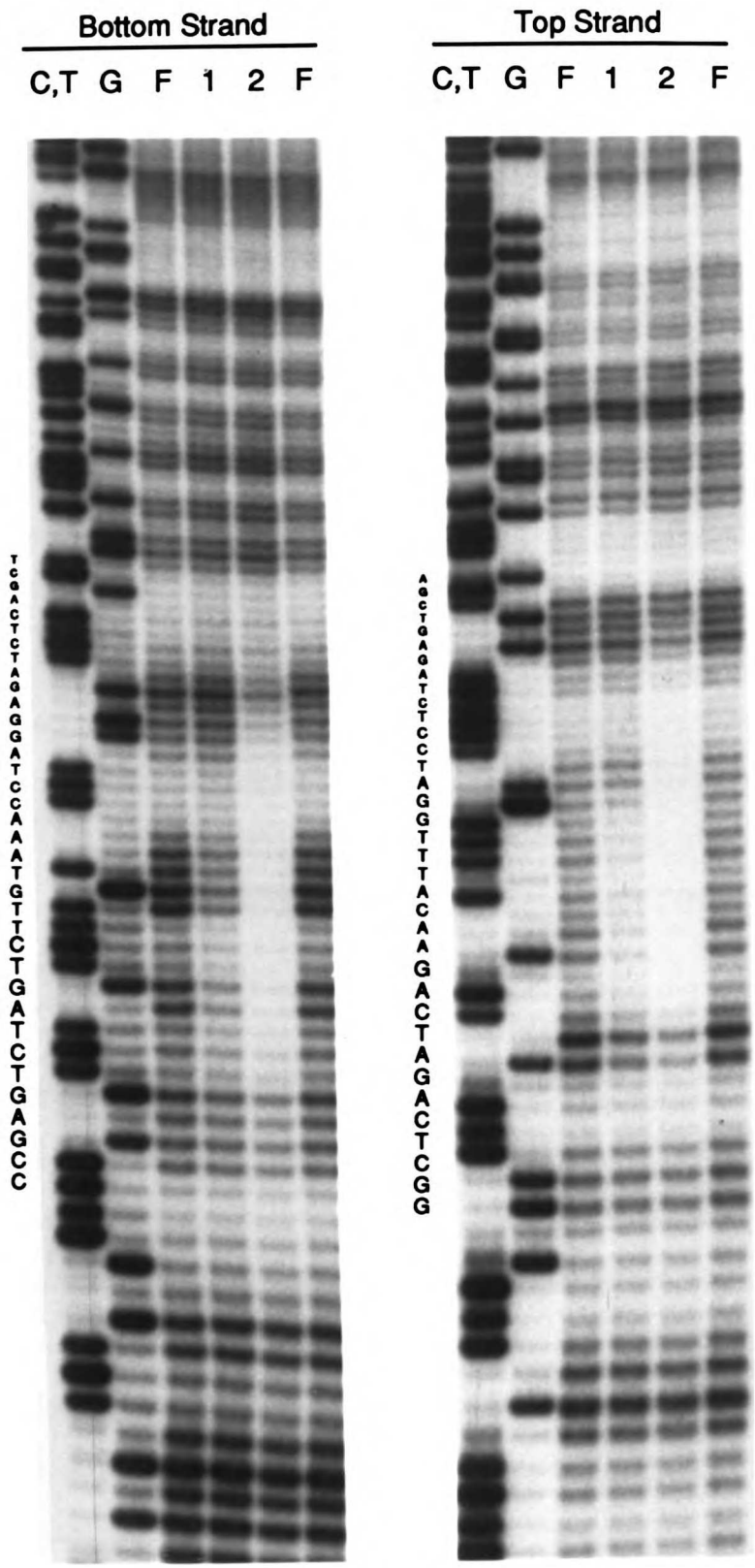
one intermediate species should be present if the protein/DNA ratio were greater than one. (I could have obtained similar results if both proteins existed as nonexchangeable multimers in solution; however, hydrodynamic studies, including gel filtration and sucrose gradients, indicate that T7X556 is a monomer in solution; L. Freedman, personal communication.)

To locate the protein molecules in each of the specific complexes, I footprinted them in situ in a nondenaturing polyacrylamide gel using OP-Cu. After immersing the gel in OP-Cu solution, the DNA was isolated from the various complexes or from the unshifted band and resolved on a sequencing gel. Compared to unbound DNA, complex 1 displayed partial protection of both strands at a 15 bp region centered over the second consensus sequence (see Figure 5). Complex 2 demonstrated protection of a larger footprint, which extended 3' to half cover the Xba I site on both strands, and which exhibited much stronger protection than the complex 1 footprint. The borders of this extended footprint coincide well with those of the DNAase I and hydroxyl radical footprints and hence confirm the notion that these footprints represent two molecules of T7X556 on the DNA. Footprinting of complexes with three and four proteins did not show any additional protection, nor any enhanced protection at existing footprints (data not shown), indicating that the third and fourth proteins are

Figure 5. 1,10 Phenanthroline-Copper Ion Complex Footprinting Analysis of T7X556 Binding to RBSwt-256

An RBS-containing 256 bp fragment (Eco RI/Bgl I) was 3' end labeled (bottom strand) or 5' end labeled (top strand) at the Eco RI site, incubated with T7X556, and resolved on a nondenaturing polyacrylamide gel. The resulting complexes 1 and 2 and the free DNA (F) were then footprinted <u>in situ</u> in the gel with 1,10 phenanthroline-copper ion complex (OP-Cu), eluted from the gel, and resolved on a sequencing gel. Lanes G and C,T are probe DNA submitted to guanine-specific or cytosine- and thymidine-specific cleavage, respectively.

OP-Cu Footprinting



distributed randomly on the rest of the DNA fragment. Thus, RBSwt-256 possesses two specific binding sites for T7X556, one with higher affinity that lies over the consensus site marked 2 in Figure 1 (site a) and an unanticipated site that lies outside the MTV-derived sequences over the Bam HI site (site b).

The Receptor Derivative Makes Close Major Groove Contacts at Both Occupied Sites

From the above data, I conclude that each half of the symmetric footprint corresponds to a molecule of T7X556. Inspection of the computer graphic representation of the hydroxyl radical footprint (Figure 3) reveals each protein molecule centered over the major groove at a position that, for site a, coincides precisely with a consensus sequence. I thought it likely that the protein makes close contacts with these bases in the major groove and tested this supposition using the methylation interference assay. In this technique, RBSwt-256 was partially methylated at the N-7 positions of guanine residues, which are exposed in the major groove, and less efficiently at the N-3 positions of adenine residues, which are exposed in the minor groove. After incubating the modified DNA with the protein, the resulting complexes were resolved by electrophoresis, the DNA was isolated and cleaved at the methylated bases, and then resolved on a sequencing gel. When a methyl group at a particular position inhibits protein binding, presumably by

preventing a close contact, molecules modified there will not be retarded in the nondenaturing polyacrylamide gel. DNA from the shifted complex will then lack molecules cleaved at that position and the corresponding band on the sequencing gel will be diminished.

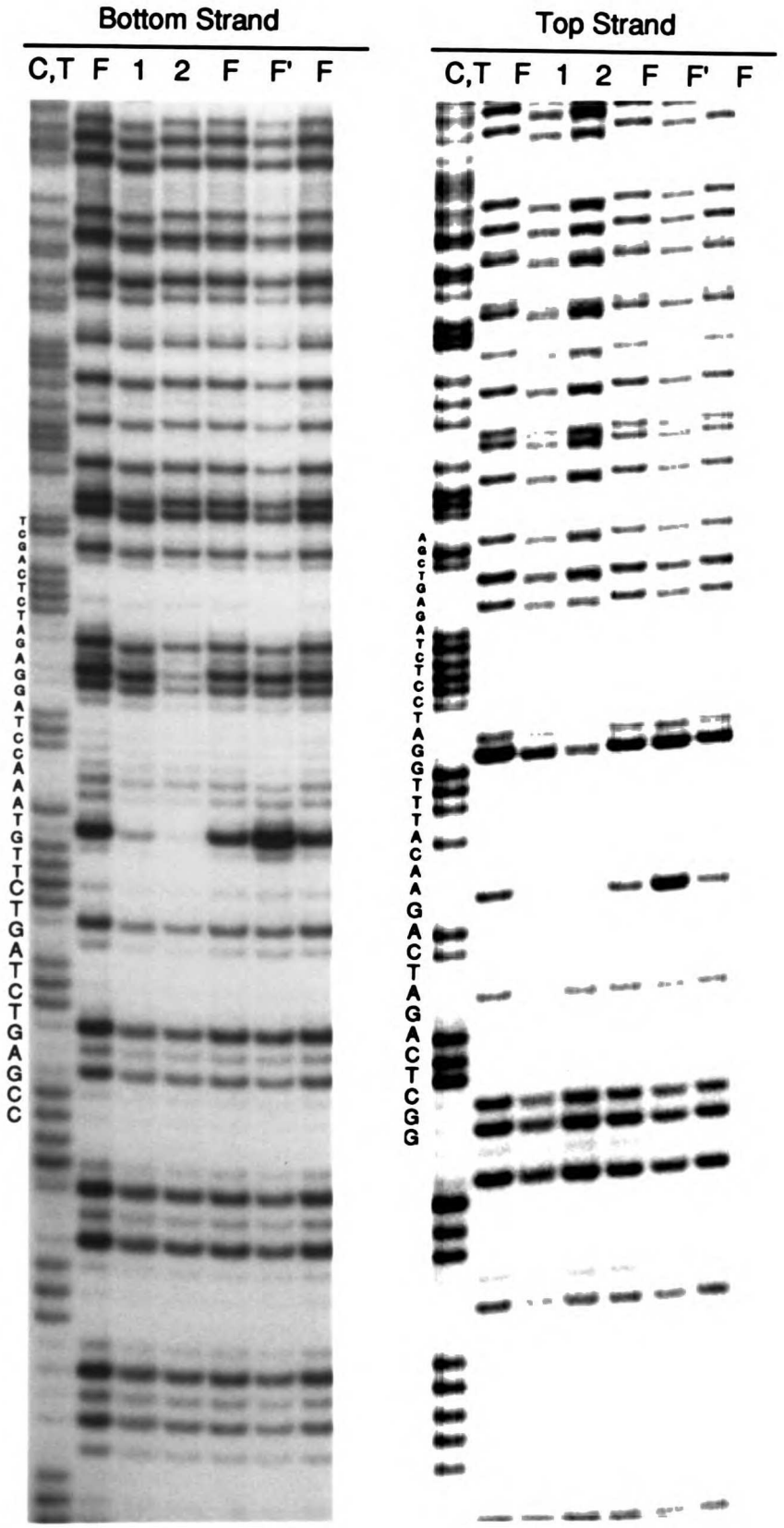
As expected, methylation of both of the G residues in the bound consensus sequence (site a), G11 (top strand) and G14 (bottom strand), prevented close major groove contacts in both complex 1 and 2 (Figure 6). In complex 2, four additional G residue contacts were made, all within the Bam HI site (site b); however, methylation at G19 did not seem to inhibit the binding of the second protein molecule as much as at the other three G residues. This latter observation is interesting since it is consistent with the interpretation that at this position the protein appears to lie more over the minor groove than the major groove (Figure 2a). I did not see any evidence for close contacts anywhere within the minor groove, nor did I see any evidence that suggested close contacts with either G residue in the consensus sequence labeled 1 in Figure 1.

The results of the methylation interference assay confirm and extend the results of the previous experiments. Together the data demonstrate that two T7X556 molecules bind contiguously at the RBS, the first at a high affinity site over a consensus sequence and the second at a Bam HI site that had been introduced to facilitate cloning. Each

Figure 6. Methylation Interference of T7X556 Binding to RBSwt-256

An RBS-containing 256 bp fragment (Eco RI/Bgl I) was partially methylated and then 3' end labeled (bottom strand) or 5' end labeled (top strand) at the Eco RI site, incubated with T7X556, and resolved on a nondenaturing polyacrylamide gel. The DNA was eluted from the resulting complexes 1 and 2 and the free DNA (F), cleaved at the modified bases with piperidine, and resolved on a sequencing gel. Lanes indicated by F' represent the probe DNA left in the free band when most of the probe was shifted to complex 2; thus, it should be enriched with molecules methylated at positions that inhibit binding. Lane C,T is probe DNA submitted to cytosine- and thymidine-specific cleavage.

Methylation Interference



protein binds to one face of the double helix covering the minor groove at either end of its footprint while centered over the major groove where it has close contacts. Mysteriously, T7X556 appeared to ignore completely one of the two consensus sequences in the RBS, while instead footprinting a site with less homology. The significance of this observation will be discussed below.

Binding Affinity of the Receptor Derivative for an RBS-Containing DNA Fragment

I sought a better understanding of the contributions made by each of the demonstrated binding sites in the RBS (a and b) towards the binding affinity of the protein for the DNA fragment. To accomplish this, I first determined the overall affinity of the protein for RBSwt-256. Since the results above imply the following equilibrium,

$$\text{RBSwt-256} + \text{T7X556} \overset{K_1}{=} \text{Complex 1} + \text{T7X556} \overset{K_2}{=} \text{Complex 2} \qquad (1)$$

I wanted to determine both binding constants. The ability to separate and identify the various protein-DNA complexes on a nondenaturing polyacrylamide gel aided me in this regard. I opted, however, not to follow the standard approach of mixing a known quantity of labeled DNA fragment with increasing concentrations of purified protein and then quantifying the resulting complexes to determine the binding constants. This approach can be misleading because of the

difficulty in determining the actual concentration of free protein in the dilute binding reactions and because of uncertainty about what fraction of the purified protein is active in DNA binding. To circumvent these problems, I adapted an approach described by Liu-Johnson et al. (1986). In this method, the total concentration of protein and DNA is varied, while the ratio of total protein added to total DNA is kept constant. Under these circumstances and if the experimental conditions are maintained such that all observed complexes contain three or fewer protein molecules, the following expressions are valid (see Materials and Methods):

$$\frac{[\alpha - (r_1 + 2r_2 + 3r_3)](1 - r_T)}{r_1} = \frac{1}{K_1} \cdot \frac{1}{D_T} \qquad (2)$$

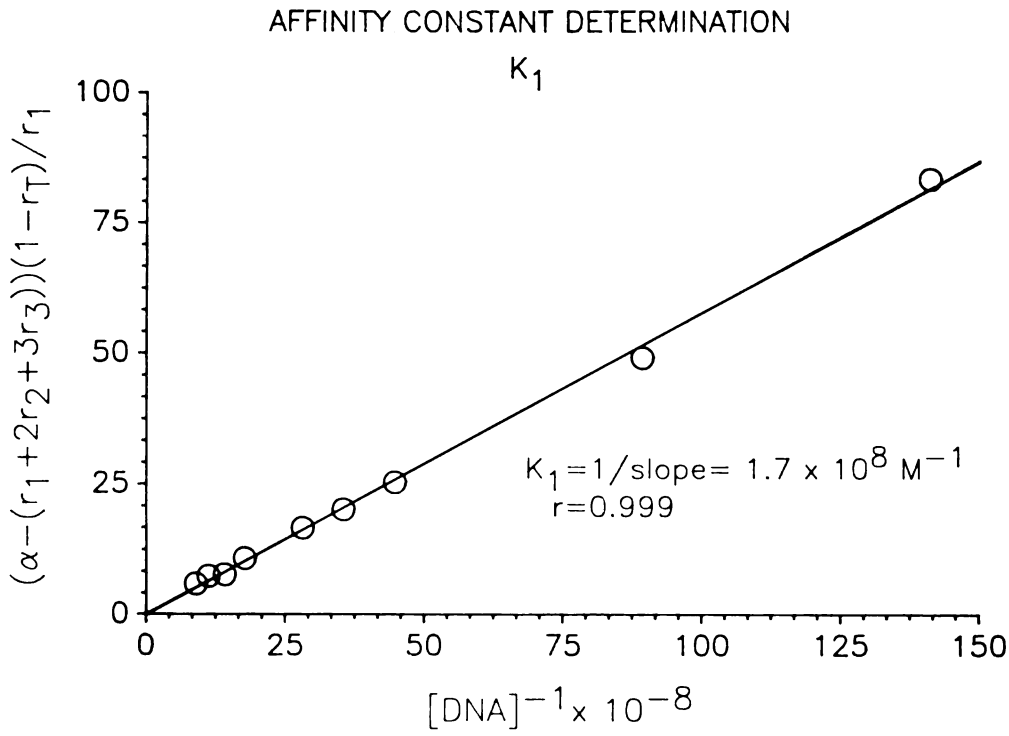
$$\frac{[\alpha - (r_1 + 2r_2 + 3r_3)]r_1}{r_2} = \frac{1}{K_2} \cdot \frac{1}{D_T} \qquad (3)$$

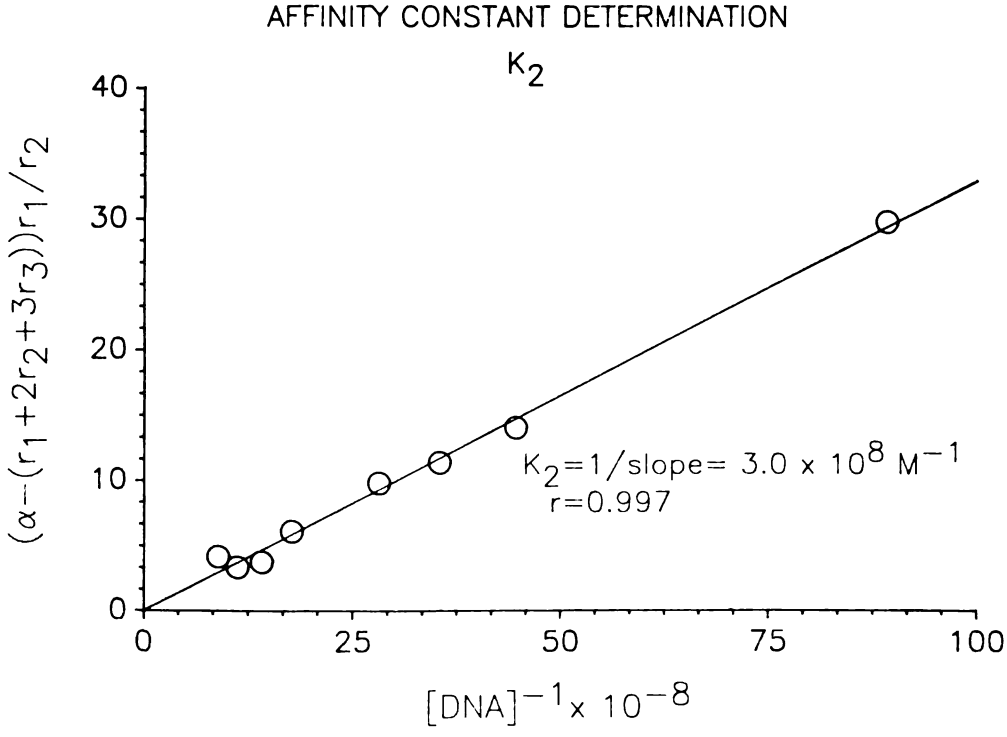
where r_1, r_2, and r_3 are the fraction of radioactivity in the bands for complex 1, 2, and 3, respectively, r_T is the sum of $r_1 + r_2 + r_3$, D_T is total concentration of DNA added, and α is the (unknown but constant) ratio of active protein to total DNA. If the left half of each equation above is plotted versus $1/D_T$ and the value of α is adjusted so that the line passes through the origin, then the slope of the line is the inverse of the association constant.

I determined the values of both constants in three independent experiments and present representative plots

Figure 7. Determination of Equilibrium Binding Constants K_1 and K_2 for T7X556 Binding to RBSwt-256

Equilibrium constants K_1 and K_2 represent the binding constants of the first and second T7X556 molecule to bind to RBSwt-256. Three independent determinations of each constant were performed according to the method described in the text. Shown above are representative plots of one of these determinations. The mean ± standard error of the mean for the three analyses were $K_1 = 1.2 \times 10^8\ M^{-1} \pm 0.6 \times 10^8\ M^{-1}$ and $K_2 = 4.3 \times 10^8\ M^{-1} \pm 1.3 \times 10^8\ M^{-1}$.


AFFINITY CONSTANT DETERMINATION
K1
(α−(r1+2r2+3r3))(1−rT)/r1
100
75
50
25
0
0
25
50
75
100
125
150
K1=1/slope= 1.7 x 10^8 M^−1
r=0.999
[DNA]^−1 x 10^−8


AFFINITY CONSTANT DETERMINATION
K2
(α−(r1+2r2+3r3))r1/r2
40
30
20
10
0
0
25
50
75
100
K2=1/slope= 3.0 x 10^8 M^−1
r=0.997
[DNA]^−1 x 10^−8

from one of these in Figure 7. The average α for these data was 8.7 ± 3.4 (mean ± sd). The fraction of protein that is active can be calculated by determining the product of α and the ratio of total DNA to total protein, which is readily determined. From this, I estimate that the protein preparation used in these experiments was about 95% active. The calculated mean ± standard error of the mean (SEM) for K_1 was $K_1 = 1.2 \times 10^8\ M^{-1} \pm 0.6 \times 10^8\ M^{-1}$ and that for K_2 was $K_2 = 4.3 \times 10^8\ M^{-1} \pm 1.3 \times 10^8\ M^{-1}$. I found it intriguing that in every one of my determinations the value of K_2 was greater than K_1. For independent binding at the two DNA sites, the higher affinity site will be preferentially occupied by the first protein molecule bound, so that the affinity of the second protein should be less than that of the first. That the affinity actually increases for the second protein molecule implies that the two proteins interact cooperatively, thereby contributing to the free energy of binding for the ternary complex.

Relative Contribution of Sites a and b Toward Binding Affinity

The constants K_1 and K_2 are macroscopic affinity constants that describe the affinity of T7X556 for the whole RBSwt-256 molecule. Alternatively, they can be regarded as the sum of a series of microscopic affinity constants, k_i, that describe the affinity of the protein for each potential binding site along the DNA fragment. The footprinting

experiments demonstrated that RBSwt-256 contains only two specific binding sites, a and b, both of which are located within the RBS. Thus, as shown in the Materials and Methods, for RBSwt-256 or any subfragment that contains the RBS, K_1 and K_2 can be exressed as

$$K_1 = nk_{ns} + k_a + k_b \quad (4)$$

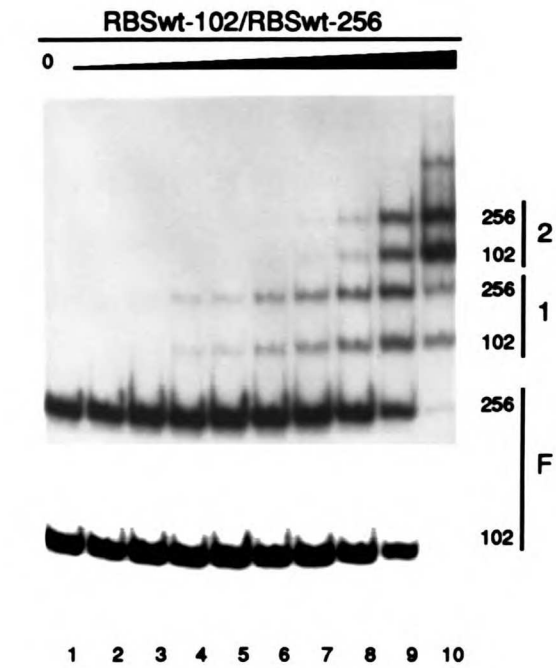
$$K_2 = (n-30)k_{ns} + Y_a k_a + Y_b k_b \quad (5)$$

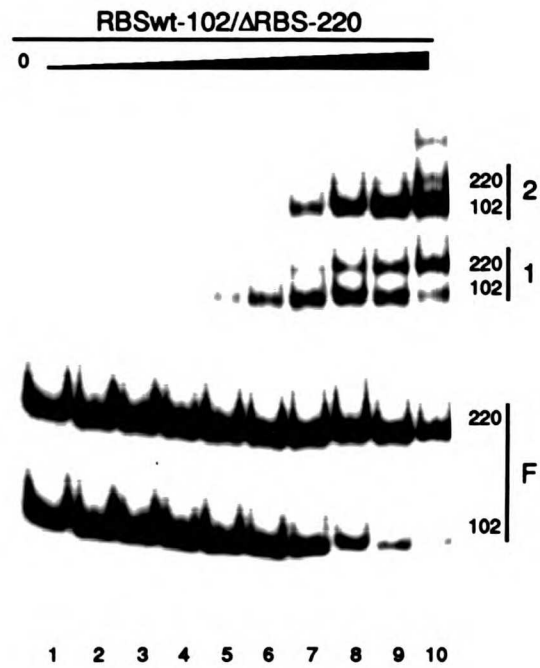
where $Y_a = (nk_{ns}+k_b k_{ab})/K_1$, $Y_b = (nk_{ns}+k_a k_{ab})/K_1$, n represents the number of nonspecific sites on the fragment, k_a, k_b, and k_{ns}, represent the microscopic affinity constants for site a, site b, and a single nonspecific site, respectively, 30 is the number of base pairs unavailable for binding because of the first protein bound, and k_{ab} is a constant describing the cooperative interaction of two proteins occupying site a and site b. To determine the contribution made by the RBS towards the binding affinity of the protein for RBSwt-256, I needed to compute the relative values of the microscopic affinity constants k_a, k_b, and k_{ns} as well as the value of the cooperativity constant, k_{ab}.

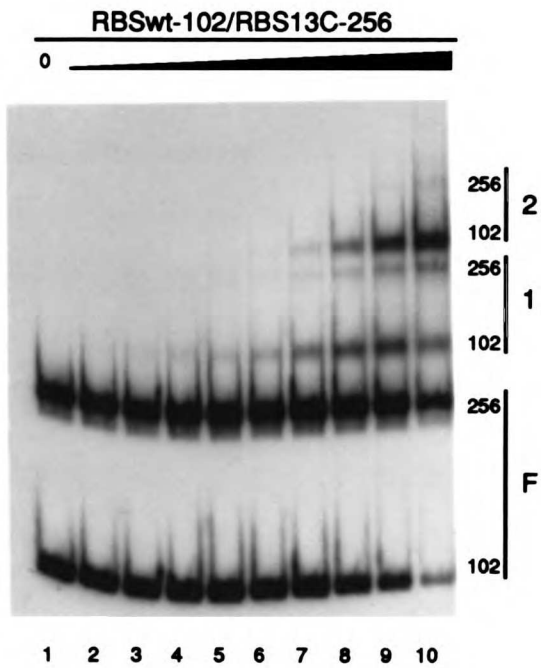
The relative magnitudes of these microscopic affinity constants can be computed by ascertaining the macroscopic affinity constant ratio of RBS-containing molecules versus molecules that have had one or both of the specific sites ablated by mutation or deletion. I performed such comparisons by incubating a mixture of two labeled DNA fragments of interest, which were of different size, with

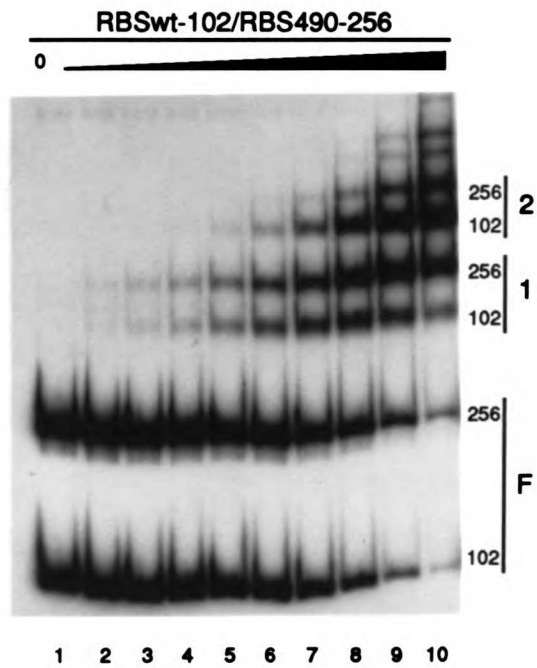
Figure 8. Assay to Determine the Relative Affinity of T7X556 for Different DNA Fragments

Increasing concentrations of T7X556 were incubated with two end-labeled DNA fragments of different sizes and resolved on nondenaturing polyacrylamide gels. The fragment size in base pairs is indicated on the right of each band, as well as whether the band represents complex 1 or 2 or free DNA (F). RBSwt-256 contains the wild type RBS, ΔRBS-220 is the same fragment except that the RBS is deleted, RBS13C-256 is the same fragment except that a cytosine is substituted at position 13 making it a site a mutant, and RBS490-256 contains 5 substitutions at postions 19-23 making it a site b mutant. All were compared to RBSwt-102, which contains the wild type RBS (see Figure 1 for details). The relative affinities, Q_1 and Q_2, were calculated as described in the text and are shown in Table 1. The concentration of T7X556 in ng/10 μl for lanes 1-10 were 0, 0.22, 0.46, 1.0, 2.2, 4.6, 10, 22, 46, 100, respectively, except for the comparison of RBSwt-102/ΔRBS-220 where the concentrations were 0, 0.46, 1.0, 2.2, 4.6, 10, 22, 46, 100, 220, respectively.


RBSwt-102/RBSwt-256
0
256
102
2
256
102
1
256
F
102
1 2 3 4 5 6 7 8 9 10


RBSwt-102/ΔRBS-220
0
220
102
2
220
102
1
220
F
102
1 2 3 4 5 6 7 8 9 10


RBSwt-102/RBS13C-256
0
256
102
2
256
102
1
256
F
102
1 2 3 4 5 6 7 8 9 10


RBSwt-102/RBS490-256
0
256
102
2
256
102
1
256
F
102
1 2 3 4 5 6 7 8 9 10

increasing concentrations of protein. I then separated the resulting protein-DNA complexes on a nondenaturing polyacrylamide gel. The fragment sizes and gel conditions were adjusted to allow resolution of all of the resulting complexes (Figure 8).

When I compared two RBS-containing fragments, RBSwt-256 and a shortened version, RBSwt-102 (Figure 1), I found that roughly the same fraction of each fragment was retarded at each protein concentration (Figure 8). This indicated that the additional 300 nonspecific sites on the longer fragment did not substantially increase the protein affinity for that fragment. In contrast, when I compared RBSwt-102 to a fragment from which the RBS had been deleted, ΔRBS-220 (legend of Figure 1), I found that RBSwt-102 was retarded at much lower protein concentrations, demonstrating the importance of the RBS for protein affinity (Figure 8).

To quantitate the relative affinity of a single protein molecule binding to each of two different DNA fragments X and Y, the term Q_1 can be defined such that Q_{1xy} represents the quotient of the affinity constant K_1 for each of the fragments, denoted as K_{1x} and K_{1y}. Similarly, Q_{2xy} represents the quotient of the affinity constants K_{2x} and K_{2y} for the binding of a second protein molecule. Thus, by measuring the radioactivity in the relevant bands, the ratios of the macroscopic affinity constants can be directly calculated:

$$\frac{K_{1x}}{K_{1y}} = Q_{1xy} = \frac{\text{complex } 1_x/\text{free}_x}{\text{complex } 1_y/\text{free}_y} \quad (6)$$

$$\frac{K_{2x}}{K_{2y}} = Q_{2xy} = \frac{\text{complex } 2_x/\text{complex } 1_x}{\text{complex } 2_y/\text{complex } 1_y} \quad (7)$$

By using the described electrophoretic assay and computing the ratios in this manner, the results were not influenced by several potential sources of experimental error such as different DNA-labeling efficiencies, unequal quantities of the two DNA fragments, and imprecise protein concentrations.

Quantification of the comparison between RBSwt-102 and ΔRBS-220 indicated that a molecule of T7X556 favors the RBS-containing fragment by 5.4 fold (Table 1). Since ΔRBS-220 lacks an RBS, K_1 for that molecule can be represented as

$$K_1 = nk_{ns} \quad (8)$$

where n and k_{ns} are as defined above. If the protein requires at least 15 bp at the ends of the fragment for binding, n can be estimated to be n = 2(length - 30) = 380. Since equation (4) represents K_1 for RBSwt-102, we can substitute it and equation (8) into equation (6) along with the appropriate values of n to obtain

$$\frac{140k_{ns} + (k_a + k_b)}{380k_{ns}} = 5.4$$

which can be reduced to

$$k_a + k_b = 1900k_{ns} \quad (9)$$

By combining equations (9) and (4), I estimate that the RBS is responsible for about 81% of the affinity of T7X556 for

RBSwt-256 and about 93% for RBSwt-102. Thus, the predicted K_1 ratio for the shorter fragment over the longer fragment should be 0.87, which is in close agreement with the empirical result (Table 1).

The OP-Cu footprinting and methylation interference experiments demonstrated that T7X556 prefers site a over site b. To express this quantitatively, I employed two mutants in the RBS. RBS13C has a single base substitution in site a (at position 13), while RBS490 has 5 substitution mutations in site b (Figure 1). Compared with RBSwt-102, RBS13C-256 had a markedly decreased affinity for the protein; in fact, it was effectively as low as that for ΔRBS-220 when measured in the gel electrophoretic assay (Figure 8 and Table 1). As expected, this site a mutation also abolished footprinting at the RBS (see Figure 9), as well as its function <u>in vivo</u> (see Table 4). In contrast, RBS490-256 formed complex 1 almost as well as RBSwt-102; however, the second complex formed with about 7 fold lower affinity in the mutant case (Figure 8 and Table 1). OP-Cu treatment of the appropriate complexes verified that RBS490 specifically ablates site b, since normal footprinting was observed over site a in complex 1, but no additional protection over site b was observed in complex 2 (see Figure 10). (Presumably, the second protein molecule is randomly dispersed over the remaining nonspecific sites.)

The K_1 values of these two mutants can be compared to

Table 1. Affinity Constant Quotients for Several RBS Mutants

The relative affinity of T7X556 for the DNA fragments indicated was assayed as described in the text and in Figure 8. The values of $Q_1 = K_{1x}/K_{1y}$ and $Q_2 = K_{2x}/K_{2y}$ were computed according to equations (6) and (7). Each determination represents the mean value for 3-6 lanes, depending on whether the relevant bands were easily discernable, except for Q_2 for RBSwt-102/ΔRBS-220, where only one lane could be used. For convenience, the value of $Q_T = Q_1Q_2$ is also shown.

Table 1

Affinity Constant Quotients for Several Mutants

	$Q_1 \pm$ s.d.	$Q_2 \pm$ s.d.	Q_T
RBSwt-102 / RBSwt-256	0.8 ± 0.1	1.1 ± 0.3	0.9
RBSwt-102 / ΔRBS-220	5.4 ± 1.0	5.3 —	28.6
RBSwt-102 / RBS13C-256	5.0 ± 0.4	2.7 ± 0.2	13.5
RBSwt-102 / RBS490-256	1.0 ± 0.1	7.1 ± 0.7	7.1
RBS13C-102 / ΔRBS-220	0.5 ± 0.1	2.3 ± 0.1	1.2
RBS490-102 / ΔRBS-220	4.8 ± 0.7	0.5 ± 0.04	2.4

each other by dividing their Q_1 values listed in Table 1, since both were assayed against the same reference fragment, RBSwt-102. By combining this quotient with equation (4) we obtain

$$\frac{K_{1\text{-}490}}{K_{1\text{-}13C}} = \frac{450k_{ns} + k_a}{450k_{ns} + k_b} = 4.8 \qquad (10)$$

After substituting equation (9) and rearranging

$$k_a = 58k_b \qquad (11a)$$

and hence

$$k_a = 1880k_{ns} \ (11b) \quad \text{and} \quad k_b = 32k_{ns} \ (11c)$$

These values confirm the observation that T7X556 has a much higher intrinsic affinity for site a than for site b, which itself is a better binding site than the nonspecific sites. These values were further substantiated by cross-checking them with the results from two entirely independent experiments. By combining the results of the comparisons RBS490-102/ΔRBS-220 and RBS13C-102/ΔRBS-220 with equations (4) and (8) I calculated that $k_a = 1700k_{ns}$ and $k_b = 38k_{ns}$, which agrees well with the results above.

Cooperativity of Protein Binding

As mentioned previously, I inferred cooperative binding from the increased affinity of T7X556 for fragments that already have one bound protein. I used the relationships expressed in equation (11) along with the data from the absolute affinity constant determination to estimate the magnitude of this effect. By substituting these values into

equations (4) and (5), I obtained the following values for all of the microscopic affinity constants

$k_{ns} \approx 5.1 \times 10^4 M^{-1}$

$k_a \approx 0.95 \times 10^8 M^{-1}$

$k_b \approx 1.6 \times 10^6 M^{-1}$

$k_{ab} \approx 1.5 \times 10^2$

Thus, T7X556 protects site b in footprinting experiments because its binding to that site is improved two orders of magnitude by its cooperativity with the T7X556 molecule bound at the high affinity site a. As expected, the binding at site a is also improved by the cooperativity, which explains the increased protection at that site in complex 2 compared to complex 1 in the OP-Cu footprints.

Saturation Mutagenesis of the RBS

To explore in detail the sequence preferences of the glucocorticoid receptor, I saturated the RBS with substitution mutations. This allowed me to identify those sequences that are optimal, acceptable, and unacceptable, and to express this quantitatively. The mutants were obtained by chemically synthesizing the RBS using nucleoside phosphoramidite precursors that had been "doped" with each of the other three at a concentration that maximizes single base substitutions (Derbyshire et al., 1986). The second strand of the oligonucleotide was templated directly from the first using the Klenow enzyme according to the self-primed oligonucleotide synthesis method of Oliphant and

Struhl (1988). Mutation and analysis of the DNA clones obtained from my first oligomer yielded 39 of the possible 54 mutations; details about the mutation frequencies are shown in Table 2. I then synthesized a second oligomer, mutagenizing only those positions still absent from my pool of mutant clones. This furnished me with five additional mutants. Finally, five additional mutants were either cloned out of "double insert" clones or synthesized directly. In total, I obtained 49 of the 54 possible independent mutations in the 18 base pair sequence examined.

I initially analyzed the mutants using the relative affinity assay already described. The mobility shift of each of the 49 mutants, in the context of the RBSx-256 fragment, was compared to that for the RBSwt-102 fragment by electrophoresis in nondenaturing polyacrylamide gels. I then computed the quotients of the macroscopic affinity constants, Q_1 and Q_2, as well as the product of the quotients, Q_T. The results of these experiments are shown in Table 3, where they are expressed as wt/mutant. The columns in this table represent positions in the RBS, and the rows indicate which bases occupy those positions; for example, when the G at position 11 is mutated to a T, the affinity of the first of the two molecules of T7X556 that binds to the RBS is reduced 4.2 fold relative to the wild type fragment.

The values in Table 3 describe the effects of the

Table 2. Distribution of Mutations in the Saturation Mutagenesis

The distribution of mutations observed for each of the two "doped" oligomers is shown along with that predicted by the binomial equation. Multiple inserts, restriction site mutations, deletions, and insertions were all considered unsuitable inserts. The significantly larger number of unsuitable inserts for oligomer 1 reflects its construction with two Xba I ends, which allowed a higher fraction of multiple inserts than oligomer 2, which had Bam HI and Xho I ends. The phosphoramidite precursors were "doped" at 18 positions for oligomer 1, and 5 positions for oligomer 2, thus, the number of possible discrete mutations for each is 54 and 15, respectively.

Table 2

Distribution of Mutations

		Oligomer 1	Oligomer 2	Predicted
Unsuitable Inserts		87	7	
Suitable Inserts		201 (100%)	136 (100%)	
Number of Mutatiions	0	92 (46%)	85 (63%)	44%
	1	73 (36%)	36 (26%)	36%
	2	29 (14%)	13 (9.6%)	14%
	3	6 (3%)	2 (1.5%)	4%
	4	1 (0.5%)	-	0.9%
Number of Discrete Mutations		39/54	11/15	

mutations on the protein's affinity for the entire DNA fragment, not just the effects on binding to the specific site of interest. Since, in the large fragment, the sum of nonspecific and site b binding constitutes about 1/5 of the overall affinity (equations 4 and 9), even complete ablation of site a can only result in a five fold decreased affinity. Thus, I derived the following expression, which relates Q_1 to the altered affinity at site a (see Materials and Methods):

$$q_a = k_{awt}/k_{amut} = Q_1/(1.1 - 0.26Q_1) \qquad (12)$$

where awt and amut represent wild type and mutant, respectively. The function q_a, which describes the change in the intrinsic affinity at the mutant site a, behaves as expected. When mutations cause only a several fold change in the q_a, and hence, the overall fragment affinity is still dominated by site a, the effect is almost directly reflected in Q_1; whereas, once site a becomes significantly impaired and therefore contributes less to the overall affinity, the curve becomes very steep, and large changes in q_a (site a affinity) are accompanied by only small changes in Q_1 (fragment affinity).

The majority of mutations in the RBS resulted in Q_1 values between 1 and 2, which translates to relatively modest effects on the mutant site a binding (~1-3 fold) using equation (12). For the most part, the remainder of the mutations fell into two classes: those with Q_1 values

Table 3. The Effects of Point Mutations in the RBS on the Affinity of T7X556

Each point mutation in the RBS, in the context of the RBSx-256 fragment, was compared to RBSwt-102 in the relative affinity assay described in the text. In this table the values $Q_1 = K_{1wt}/K_{1mut}$ and $Q_2 = K_{2wt}/K_{2mut}$ (see text), are expressed as mean ± s.d. for 4-6 and 3-5 lanes, respectively. Columns represent positions in the RBS, with the wild type sequence indicated above and the numbered position indicated below; rows represent the base that occupies that position, indicated to the left. Filled boxes indicate substitutions that produce wild type sequence, and n.d. is not determined. To facilitate interpretation, dark shading indicates mutations that caused severe (>500 fold) reductions in binding at site a, while light shading indicates mutations that caused moderate (ca. 3-10 fold) reductions in binding at site a. The data for all of the mutants were first quantified by scanning autoradiographs by densitometer. The experiment was repeated for any mutant whose Q_T was greater than 2.0, including all those shaded and RBS7C, 7G, 8G, 9T, 10T, 17A, 17C and 17G, and quantified by excising gel slices and counting Cerenkov radiation in a scintillation counter. In general, the two methods agreed to within ± 10% and always to within ± 30% (except for RBS10C, ± 50%). When both methods were used, the counted value is shown.

	G	C	T	C	A	G	A	T	C	A	G	A	A	C	A	T	T	T	
A	1.0 ±0.1	n.d.	0.94 ±0.1	1.3 ±0.3		1.0 ± 0.1		1.4 ±0.2	1.1 ±0.1		2.5 ±0.3			5.6 ±0.7		n.d.	1.3 ± 0.1	1.5 ± 0.2	Q_1
	1.2 ±0.1		1.2 ±0.1	1.3 ±0.2		1.3 ± 0.3		1.3 ±0.1	1.4 ±0.2		1.6 ±0.3			4.6 ±0.5			1.3 ± 0.2	0.75 ± 0.1	Q_2
	1.2		**1.1**	**1.4**		**1.3**		**1.8**	**1.5**		**4.0**			**25.8**			**1.7**	**1.1**	Q_T
C	n.d.		1.3 ± 0.1		1.0 ±0.1	1.3 ± 0.2	1.7 ± 0.2	n.d.		2.6 ± 0.4	5.5 ±0.6	2.1 ±0.4	5.0 ± 0.4		5.1 ± 0.7	2.2 ±0.5	1.2 ± 0.1	1.7 ± 0.3	Q_1
			1.1 ± 0.2		1.2 ±0.1	1.0 ± 0.1	1.0 ± 0.1			1.5 ± 0.1	2.1 ±0.4	1.7 ±0.6	2.7 ± 0.2		2.8 ±0.8	1.1 ±0.1	1.1 ± 0.2	1.0 ± 0.1	Q_2
			1.4		**1.2**	**1.3**	**1.7**			**3.9**	**11.6**	**3.6**	**13.5**		**14.3**	**2.4**	**1.3**	**1.7**	Q_T
G		1.0 ±0.1	1.1 ± 0.1	1.0 ± 0.1	1.3 ±0.1		1.4 ± 0.2	1.9 ±0.2	0.89 ±0.2	0.88 ± 0.1		1.0 ±0.1	5.0 ±0.5	4.6 ±0.4	2.9 ± 0.5	3.0 ±0.3	1.8 ± 0.3	1.1 ± 0.1	Q_1
		1.2 ±0.2	1.2 ± 0.1	1.1 ± 0.1	0.90 ±0.1		1.2 ± 0.1	1.1 ±0.2	1.3 ±0.1	1.3 ± 0.1		1.2 ±0.2	4.6 ± 0.4	4.4 ±2.4	2.2 ± 0.2	1.6 ±0.1	1.1 ± 0.1	0.79 ± 0.1	Q_2
		1.2	**1.3**	**1.1**	**1.2**		**1.7**	**2.1**	**1.2**	**1.2**		**1.2**	**23.0**	**20.2**	**6.4**	**4.8**	**2.0**	**0.87**	Q_T
T	n.d.	0.98 ±0.2		1.1 ± 0.1	1.1 ±0.1	0.99 ± 0.1	1.1 ± 0.1		0.57 ±0.1	1.5 ± 0.1	4.2 ±0.6	1.2 ±0.1	2.4 ±0.2	5.9 ±0.5	3.6 ±0.3				Q_1
		1.1 ±0.1		1.1 ± 0.1	1.1 ±0.1	1.2 ± 0.2	1.2 ± 0.1		1.3 ±0.1	1.3 ± 0.1	2.9 ±0.2	1.3 ±0.3	1.8 ±0.1	4.9 ±0.7	2.0 ±0.1				Q_2
		1.1		**1.2**	**1.2**	**1.2**	**1.3**		**0.74**	**2.0**	**12.2**	**1.6**	**4.3**	**28.9**	**7.2**				Q_T
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	

between 2 and 3 that caused moderate decreases in the mutant site a affinity (~3-10 fold) and those with Q_1 values greater than 4.2 that represent severe mutations (>500 fold). One mutation, RBS15T, displayed a Q_1 of 3.6 (~20 fold) and, surprisingly, only one mutation, RBS9T, resulted in an increased affinity (1.7 fold). All of the mutations that caused significant changes in affinity were located within positions 10-16, at the bound consensus sequence. At least one of the substitutions at each of the positions 11, 13, 14, and 15 completely abolished binding at site a, further implicating these particular bases as important contact points for the protein. The protein also seems to prefer certain bases at positions 10, 16, and perhaps 12, though the effects of mutation at these positions is more moderate. Consistent with previous data, substitutions at positions within the consensus sequence that is not protected by the protein in footprinting experiments (site 1 in Figure 1) caused little or no effect on protein affinity for the RBS. As expected, changes in K_2 generally mirrored the effects observed with K_1, i.e., all positions with increased Q_1 values also had increases in Q_2; however, some substitutions affected K_2 by an unusually large amount. (see positions 13 and 14).

OP-Cu Footprinting of Mutants

I was curious about how well T7X556 could footprint some of the substitution mutations of the RBS, and whether

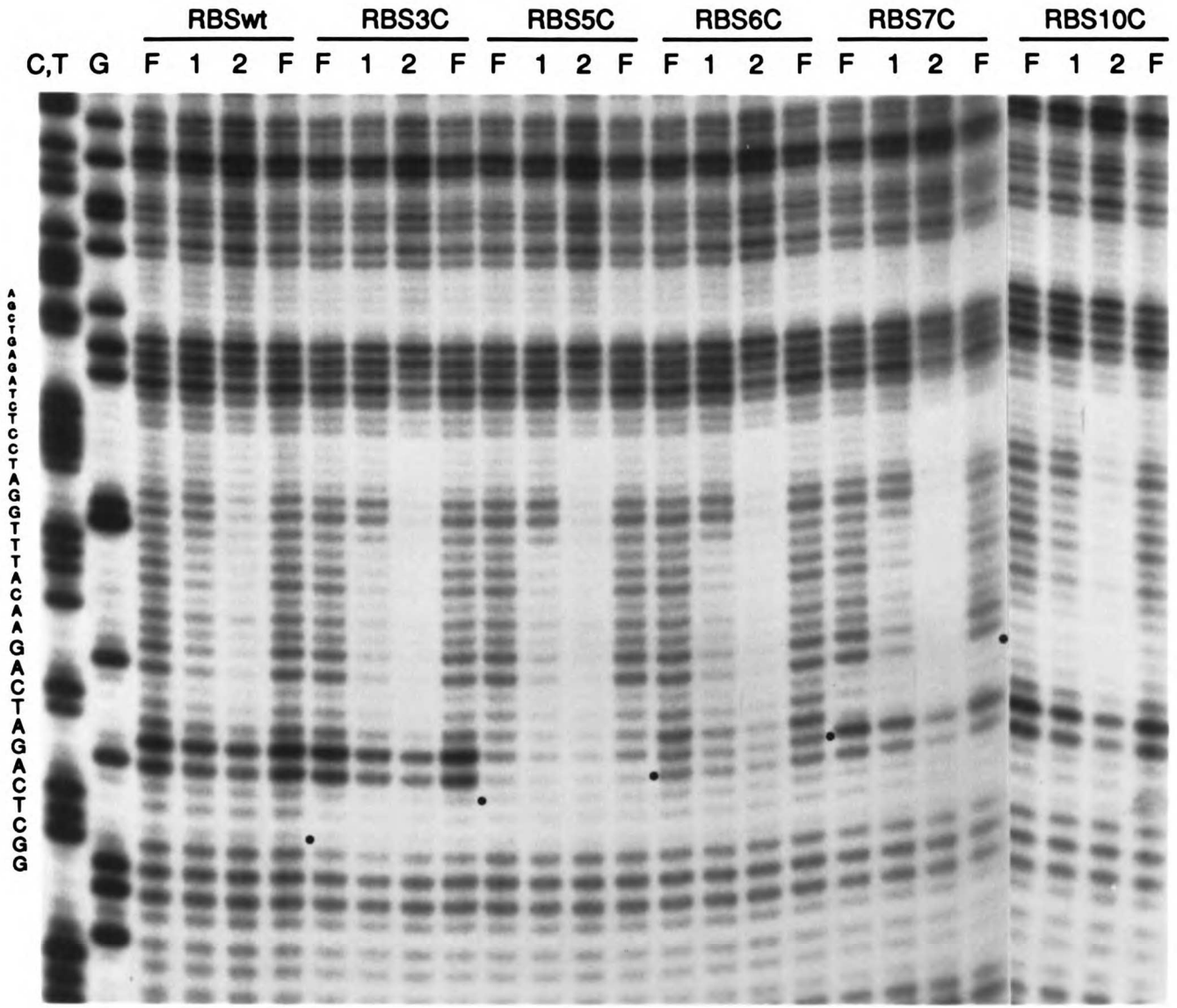
this ability correlated with the relative binding affinities. I selected a series of mutant RBSs corresponding to the second row of Table 3, (all substitutions of C), and footprinted them in the context of the RBSx-256 fragment using the OP-Cu technique described above (Figure 9). For several of the substitutions, (RBS5C and RBS6C, for example), I observed an altered cleavage pattern near the mutation even in the free DNA lanes, presumably due to differential cleavage preferences of the OP-Cu reagent (Veal and Rill, 1988).

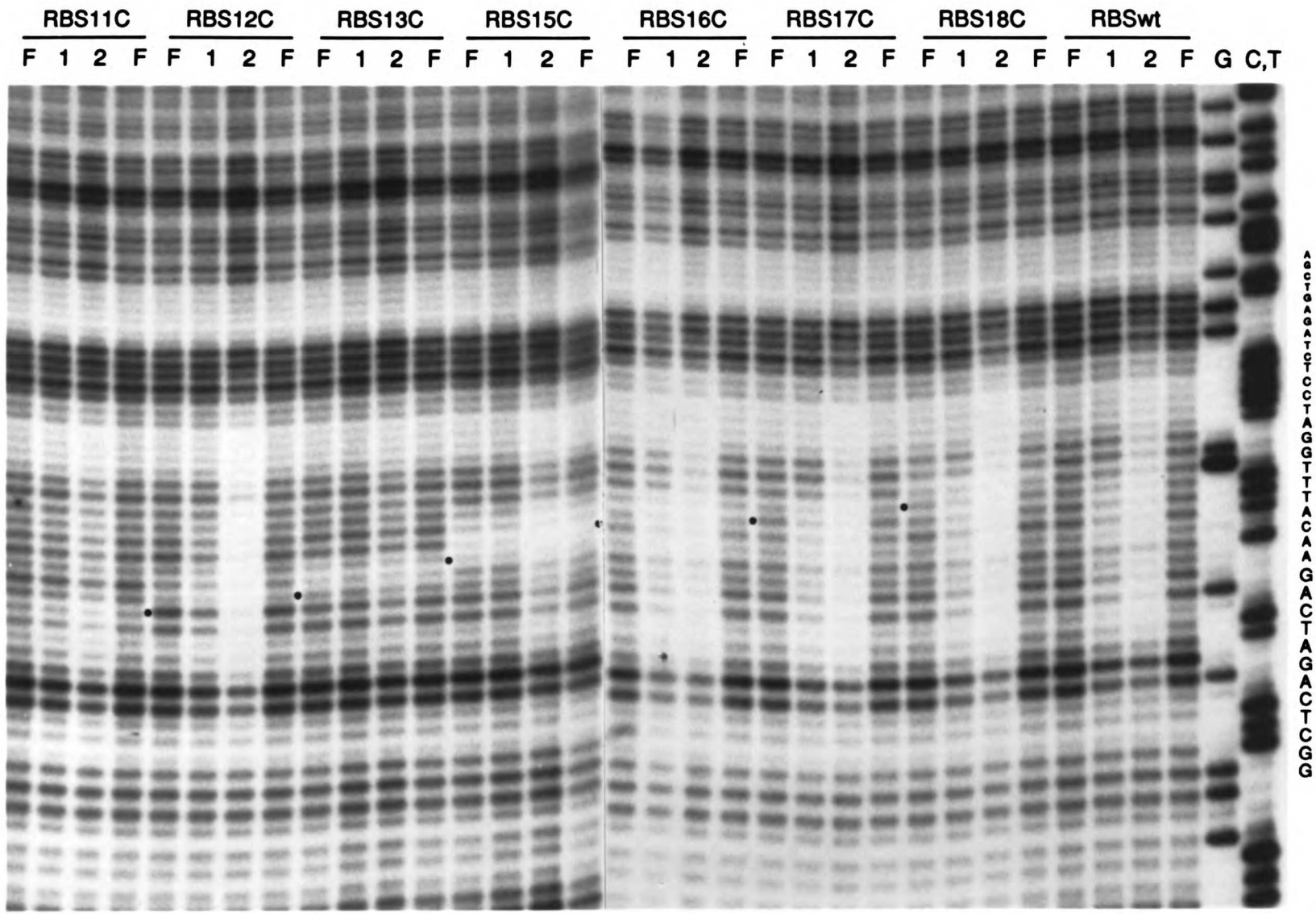
I found that the footprinting results were in good agreement with the data from the relative affinity assay. Mutations that caused a severe reduction in binding affinity - RBS11C, RBS13C, and RBS15C - also displayed severely reduced footprinting. For all three mutants, no obvious protection was observed in complex 1; however, in complex 2, RBS11C, and to a greater extent, RBS15C displayed some faint protection. The rest of the mutants, even those that exhibited moderately reduced binding affinity, displayed normal footprints over the RBS. In every case where I observed footprinting, the location and borders of the footprints, both in complex 1 and 2, were the same as those for the wild type RBS.

Why does T7X556 ignore the consensus sequence labeled 1 in Figure 1? Although the relative affinity data from the substitution mutants suggested that a T at the fourth

Figure 9. Footprinting Analysis of T7X556 Binding to RBS Substitution Mutants

DNA fragments (Eco RI/Bgl I) containing different point substitution mutations in the RBS were each 5' end-labeled at the Eco RI site and footprinted with OP-Cu as described earlier. DNA from complexes 1 and 2 or free DNA (F) are indicated at the top of each lane. The black dot preceding each group of four lanes indicates the base that has been mutated in that group. Lanes G and C,T are probe DNA submitted to guanine-specific or cytosine- and thymidine-specific cleavage, respectively.


RBSwt
RBS3C
RBS5C
RBS6C
RBS7C
RBS10C
C,T G F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F
A G C T G A G A T C T C C T A G G T T T A C A A G A C T A G A C T C G G

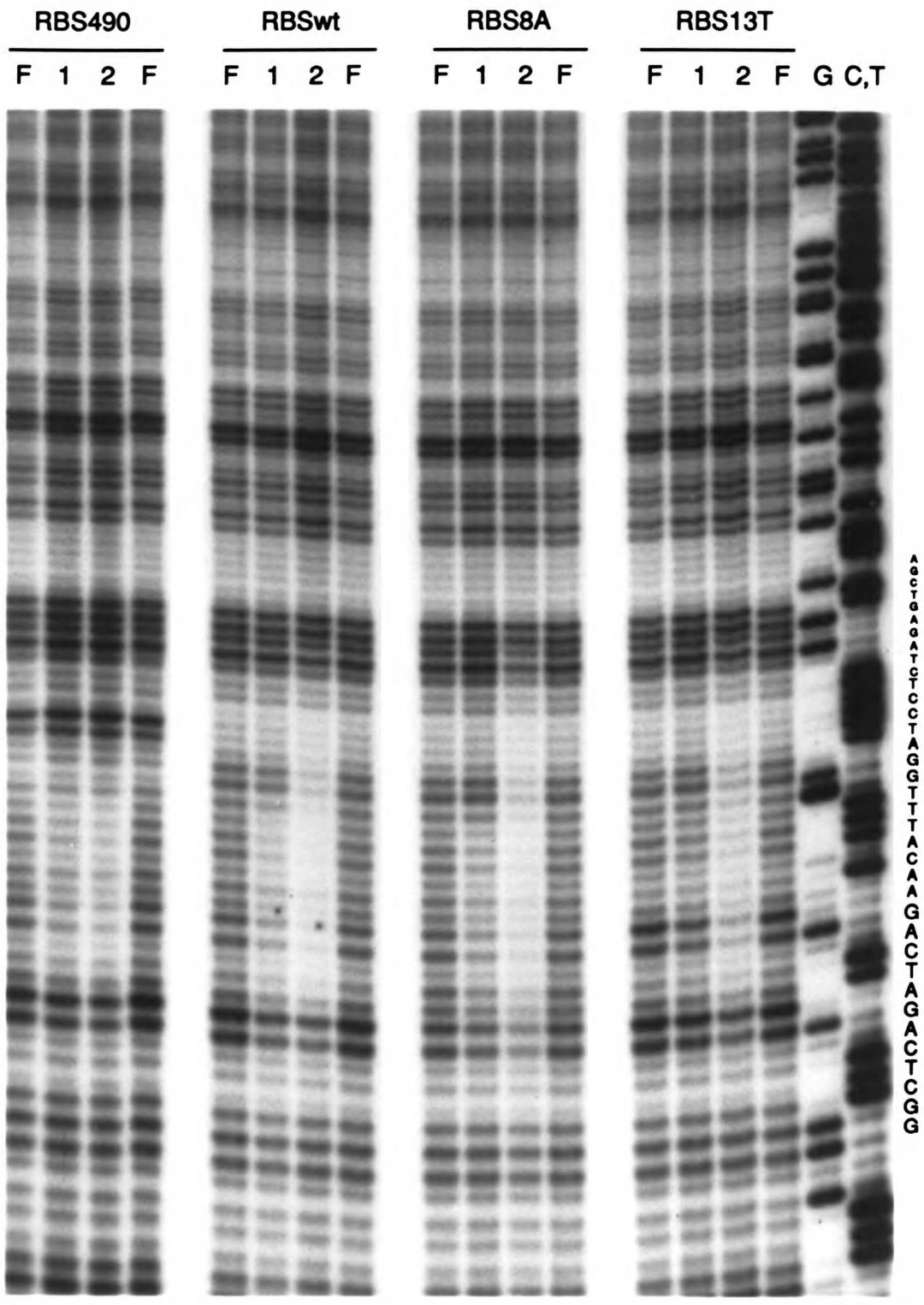

RBS11C
RBS12C
RBS13C
RBS15C
RBS16C
RBS17C
RBS18C
RBSwt
F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F F 1 2 F
G
C,T
AGCTGAGATGTCCTAGGTTTACAAGACTAGACTCGG

position of the second consensus sequence causes only a moderate reduction in affinity (see Table 3, RBS13T), it was a formal possibility that the first consensus sequence was not recognized because of the T in its fourth position. If this were true, then mutants RBS8A and RBS13T should be equalized with respect to the two consensus sequences, since, in the case of the former, both would have the favored sequence, and in the latter, both would have the inactive sequence. By this view, we would expect an extension of the normal footprint at site a to include nucleotides 5-10 on RBS8A, and the absence of any footprint on RBS13T. However, OP-Cu footprints (figure 10) revealed only a reduced, though not absent, protection of site a in complex 1 for RBS13T-256, and protection in complex 2 similar to that for RBSwt-256. Thus, a T in the fourth position of the consensus sequence reduces but does not prevent specific binding at the RBS. The footprints on RBS8A-256 were similar in both quality and location to those of RBSwt-256, and I did not observe any extension of the footprint over the first consensus sequence in either complex. Therefore, even when both consensus sequences are identical, the protein specifically chooses one over the other!

A second reason why the protein might favor the consensus sequence marked 2 in Figure 1 is its proximity to site b. This seemed unlikely since the protein exhibited

Figure 10. Mutations That Ablate Site b or That Balance the Two Consensus Sequences Do Not Shift the Footprints

DNA fragments (Eco RI/Bgl I) containing RBS490, which has 5 substitutions in site b, or substitutions that convert the first consensus to the second (RBS8A) or the second to the first (RBS13T) or RBSwt were each 5' end-labeled at the Eco RI site and footprinted with OP-Cu as described earlier. DNA from complexes 1 and 2 or free DNA (F) are indicated at the top of each lane. Lanes G and C,T are probe DNA submitted to guanine-specific or cytosine- and thymidine-specific cleavage, respectively.


RBS490
RBSwt
RBS8A
RBS13T
F 1 2 F
F 1 2 F
F 1 2 F
F 1 2 F G C,T
AGCTGAGATCTCCTAGGTTTACAAGACTAGACTCGG

its bias even when forming complex 1, where there can be no cooperativity. Nevertheless, the proximity of the two sites could direct the formation of some altered structure in the DNA, such as the opening of the major groove (Dickerson, 1983; Calladine and Drew, 1986). To test this, I footprinted RBS490-256, a mutant which has five substitutions in site b (Figure 1) and a severely reduced value for K_2 compared with a wild type fragment (Table 1). As expected, T7X556 did not protect site b in either complex 1 or in complex 2 (where the second protein was apparently randomly dispersed), verifying that site b had been eliminated (Figure 10). If juxtaposition with site b is important, then RBS490-256 lacks the incentive for the protein to seek the second consensus, and the protein should either redistribute to both consensus sequences, with the concomitant extension of the footprint, or there should be a loss of footprinting altogether. Instead, I found a footprint only over site a for both complex 1 and 2. In both complexes, the borders of the footprint and the degree of protection roughly match those of the wild type complex 1 footprint, thus demonstrating that the protein still prefers the same consensus site, even in the absence of site b.

How the protein chooses one of the consensus sequences remains unanswered; the results above clearly do not support the simple hypothesis that a perfect consensus sequence is sufficient for T7X556 binding. Rather, I conclude that, in

addition to the specific sequence requirements within the binding site, binding also depends on the context of the binding site. Thus, unknown properties of sequences outside of the consensus must influence the protein's decision about whether a potential site is acceptable or not; in the case of the fragments that I chose to test, the context effect is in fact very strong.

Effects of RBS Substitution Mutations in vivo

All of the affinity experiments described so far were performed in vitro using a 150 amino acid derivative of the glucocorticoid receptor. I wanted to know how these in vitro binding data compared to the ability of the full length receptor to functionally interact with the mutants in vivo. As indicated earlier (see Figure 1), I cloned the RBS and all of its derivatives so that they were positioned adjacent to a TK promoter-driven CAT gene; thus, I could test the GRE activity of each RBS by cotransfecting these reporter plasmids with an expression plasmid carrying a full length cDNA of the receptor gene into CV-1 cells, which contain little or no active receptor.

When I tested the GRE activity of the wild type RBS cloned in each orientation, I did not observe any hormone-dependent activation of CAT enzyme activity (data not shown). (While I did not perform an exhaustive study, I varied the amounts and ratios of the two plasmids transfected and tested different plasmid preparations and

different isolates of CV-1 cells; in no case was a single RBS active.) In contrast, a dimerized version of the RBS, which I produced by linking two RBSs at the Xho I site (Figure 1), conferred substantial induction of CAT activity (~20 fold) in the presence of hormone.

I therefore dimerized several of the substitution mutants in order to test their function in vivo. The mutants were dimerized so that each half of the dimer contained the mutation. A summary of the results of four independent transient assays of these mutants can be found in Table 4. Consistent with my previous data, those mutations that severely affected binding affinity and footprinting also markedly affected GRE activity. The mutants RBS11C and RBS13C displayed no GRE activity at all, and while RBS15C did show reproducible induction, its magnitude was greatly reduced relative to wild type. Furthermore, most of the mutations that showed little or no effect on binding affinity also had little effect on in vivo function (RBS10G, RBS12G, RBS12T, for example).

In general, the results for the mutants that I tested in vivo corroborated those in vitro; however, for several of the mutants (e.g., RBS6C, RBS7C and RBS18C), the in vivo activity was significantly lower than would be predicted by the affinity and footprinting data. Thus, sequences outside the hexanucleotide binding site appear to contribute to biological function, suggesting that context is important

for more than just the affinity of the protein for its binding site.

Finally, these in vivo experiments highlighted the importance of site b for function of the GRE. Mutations that abolish binding at that site, as in the dimerized RBS490, demonstrated no induction.

Table 4. Dexamethasone-Induced CAT Expression in Transient Transfection Assays by RBS Dimers

Several dimerized RBS mutants, cloned upstream of a TK-promoter driven CAT gene (see text and Figure 1), were tested for GRE activity by cotransfection into CV-1 cells with a plasmid expressing the full length rat glucocorticoid receptor. After 16 hours of incubation with the two plasmids, the cells were washed, and fresh medium, with or without 10^{-7} M dexamethasone, was added. The cells were harvested after 24 hours, and the CAT activity was determined. The results shown represent a summary of 4 independent transfection experiments and are expressed as mean ± SEM. An Analysis of Variance (ANOVA) on the data for basal expression, which is expressed as % conversion to acetylated chloramphenicol, demonstrated that the mutations had no significant effect on CAT expression without hormone ($F=0.51$). In contrast, an ANOVA on the fold induction observed in the presence of hormone indicated significant ($p<0.01$) differences among the data ($F=5.17$). To determine which mutants were significantly different from wild type, a Bonferroni *t* test was employed, and those mutants that failed the test are indicated by their significance level (for details on these statistics, see Glantz, 1981).

Table 4

Induction of CAT Expression in Transient Transfection Assays

	Basal Expression % Conversion	Fold Induction	
RBS5C	4.5 ± 1.0	11.7 ± 1.5	
RBS6C	3.8 ± 1.2	4.8 ± 0.4	$p<0.05$
RBS7C	3.6 ± 1.7	6.6 ± 0.9	$p<0.05$
RBS10C	3.8 ± 1.4	7.8 ± 2.0	$p<0.05$
RBS11C	3.6 ± 1.7	0.8 ± 0.1	$p<0.001$
RBS12C	3.6 ± 2.4	9.9 ± 1.9	
RBS13C	2.8 ± 1.4	0.7 ± 0.1	$p<0.001$
RBS15C	1.2 ± 0.2	2.6 ± 0.4	$p<0.001$
RBS16C	1.3 ± 0.3	13.4 ± 2.7	
RBS17C	2.0 ± 1.1	17.6 ± 3.4	
RBS18C	4.0 ± 2.1	6.8 ± 1.9	$p<0.05$
RBS10G	3.5 ± 1.9	19.7 ± 6.8	
RBS12G	3.8 ± 1.8	17.4 ± 4.9	
RBS12T	3.6 ± 1.5	13.4 ± 4.9	
RBS490	1.3 ± 0.2	1.3 ± 0.3	$p<0.001$
RBSwt	3.7 ± 1.8	18.0 ± 4.7	
ΔRBS	4.5 ± 1.9	0.2 ± 0.1	$p<0.001$
	$\bar{X}$ ± SEM	$\bar{X}$ ± SEM	

Discussion

The Receptor Derivative Displays a DNA Binding Pattern Similar to Helix-Turn-Helix Proteins

Glucocorticoid receptor, partially purified from rat liver, has previously been used to identify several high-affinity binding sites in the upstream LTR of MTV (Scheidereit et al., 1983; Payvar et al., 1983). I have isolated one of these binding sites and studied in detail its binding to a DNA-binding derivative of the receptor, purified from E. coli. I observed that two molecules of T7X556 bind symmetrically to a single RBS, despite its lack of sequence symmetry. This result extends the recent observations that two steroid receptor derivatives bind to sequences with rotational symmetry (Kumar et al., 1988; Tsai et al., 1988). Thus, while certain symmetric sequences may bind with high affinity, the protein does not require a symmetric binding site to bind symmetrically to the DNA. (The relevance of sequence symmetry will be discussed in more detail below.) Moreover, by utilizing molecules that contained mutations in separate halves of the binding site and an assay that allowed me to quantify the fraction of protein-DNA complexes with both one and two protein molecules, I was able to compute the intrinsic affinity of each half of the binding site. Thus, one protein molecule binds with high affinity ($\sim 10^8$ M^{-1}) over site a, which

coincides with a consensus sequence derived for glucocorticoid receptor binding, the second binds with significantly lower affinity (~10^6 M^{-1}) over the adjacent site b, which corresponds to a restriction site originally added to facilitate cloning.

The hydroxyl radical footprints demonstrated that each molecule of T7X556 protects only one face of the double helix. Moreover, the opposite face of the double helix is unoccupied since DNAase I can bind there while making its unprotected proximate cleavages. This contrasts observations made of another "zinc-finger" protein, TFIIIA, where the protein appears to follow around the DNA in the major groove through at least one turn of the double helix at each end of the protein (Vrana et al., 1988). TFIIIA, however, contains nine zinc fingers, while receptor has only two, thus confounding direct comparisons between the proteins. The helix face protected by the T7X556 molecule at site b is rotated by about one third of a turn around the double helix relative to that at site a. This presumably results from the three base pairs (positions 16-18) that separate the two binding sites. (I have not tested the effect of changing the distance between these two sites.)

At each binding site, T7X556 makes intimate contacts within the major groove. From the methylation interference exhibited at site b, I infer that close contacts can occur over at least six base pairs, though not much more since I

did not observe interference at positions 9 and 26 (bottom strand). The extent of these major groove interactions and the absence of any minor groove interactions is most consistent with the protein-bound RBS existing in the B form. That this is also true of the unbound RBS is supported by its sensitivity to DNAase I and OP-Cu (Suck and Oefner, 1986; Veal and Rill, 1988). I found no evidence for major groove contacts by a second "finger" at an adjacent half-turn as would be predicted by the "zinc finger" binding models proposed by Klug and Rhodes (Rhodes and Klug, 1986; Fairall et al., 1986). Indeed, recent evidence for TFIIIA also suggests that multiple "fingers" may interact with only short regions of major groove (Vrana et al., 1988).

In many ways the binding of T7X556 to the RBS is reminiscent of the binding of helix-turn-helix proteins for their cognate operators. In both cases two protein molecules bind symmetrically to two successive major grooves on the same side of the double helix, making additional backbone contacts on the minor groove that separates them. In each case, the close contacts in the major grooves determine the specificity. To extend the analogy, helix-turn-helix proteins and receptor each depend on at least two substructural units for binding, two α-helical regions in the former (Pabo and Sauer, 1984; Schleif, 1988), and two "zinc fingers" in the latter (Freedman et al., 1988; Severne et al., 1988; Green et al., 1988). Indeed, theoretical

considerations (Berg, 1988), as well as some physical evidence (Párraga et al., 1988), indicate that "zinc fingers" are partially α-helical and may bind in the major groove. Furthermore, recent experiments in our laboratory implicate one of these putative α helices as critical for DNA binding (M. Schena, personal communication). Thus, one of the receptor's "fingers" might lie in the major groove "recognizing" the exposed bases, while the other lies outside the groove, either stabilizing the first and/or contributing to the binding energy through interactions with the backbone. This is consistent with the recent study by Green et al. (1988) which showed that amino acids within or near only one of the "zinc fingers" were needed to determine binding specificity. Though the similarities with helix-turn-helix proteins are tantalizing, direct physical evidence for any structural model of receptor binding awaits further experimentation.

Recognition of the Binding Site Depends on 4 Critical Base Pairs

I observed that the RBS mutants with decreased affinity fell into two discrete classes: those that resulted in moderately reduced affinity (3-10 fold) and those that caused much more severe effects (>500 fold). In thermodynamic terms, reductions in affinity result from the loss of interactions that contribute to the free energy of binding. From the thermodynamic definition of the

equilibrium constant, K,

$$K = e^{-\Delta G/RT}$$

under standard conditions, the ΔG caused by the moderate mutations can be estimated to be around 0.7-1.4 kcal/mole and that caused by the severe mutations to be >4 kcal/mole. Thus, it seems reasonable that the severe mutations represent the loss of several hydrogen bonds (generally ~1 kcal/mole each in solution), or a combination of hydrogen bonds and Van der Waals interactions (~1 kcal/mole). Presumably, the moderate mutations represent binding sites allowing compensatory rearrangements that preserve most of the binding energy.

In this study, I found four positions where a single substitution prevented site-specific binding; that is, at least one of the mutations at positions 11, 13, 14, or 15 caused severely reduced affinity at site a. It is likely that these base pairs, which correspond to the 2nd, 4th, 5th and 6th positions in the consensus sequence, respectively, directly contact the protein. While binding is most favored when the sequence there is -G-ACA, certain substitutions are tolerated (e.g., RBS11A and RBS13T), presumably because of alternative bond formation; however, all substitutions at these positions significantly reduced binding. Not surprisingly, these bases are highly conserved among many reported GREs (Scheidereit et al., 1983, 1984; Payvar et al., 1983; Jantzen et al. 1987; Miksicek et al., 1986); for

example, among the five receptor footprints previously identified in the upstream LTR of MTV (Payvar et al., 1983), four have at least one copy of the sequence -G-ACA, and one, which is the weakest footprint, has the sequence -A-ACA (see Table 5).

While the footprinting experiments indicate that the protein is also in contact with other bases in the RBS, the mutational analysis suggests that they are not as important for site-specific binding, i.e. change at these positions is better tolerated. Nevertheless, the protein does display certain preferences for positions other than the core sequence. Substitution of a C at positions 10 and 12 (1st and 3rd consensus position, respectively) had moderate effects on the binding affinity and in vivo function of the RBS, yet little or no effect was observed when G or T were substituted at these positions. Such mild preferences might result from the absence of a protein contact (e.g., hydrogen bond or Van der Waals interaction) with the less favored base that is provided by the other more favored bases; alternatively, the unfavored base might somehow inhibit a protein contact which would otherwise occur on a nearby base.

Cooperativity Aids Binding at Site b

It is useful to consider site b in the context of the protein's sequence preferences. If we align the core sequence of site b, GGATCC, with that of site a, AGAACA, and

Table 5. Glucocorticoid Receptor Binding Sequences

Sequences from the five footprints identified in the upstream LTR of MTV (Payvar et al., 1983) that resemble the optimal binding sequence for glucocorticoid receptor (5'-[A/G]GAACAT -3') are aligned here according to their best match of a single core sequence. Underlined bases indicate exact matches with the optimal sequence at core positions. The sequences of RBSwt and of a predicted optimal sequence are included for reference. The relative footprint strength reflects the quantity of isolated receptor necessary to observe protection on a DNA fragment containing all five footprints; ++++, 200ng; +++, 400ng; ++, 600ng; +, 800ng.

Table 5

Glucocorticoid Receptor Binding Sequences

					Relative Footprint Strength
Predicted Optimal Sequence		A/G G A A C A	T A A	T G T T C T/C	
RBSwt		A G A A C A	T T T	G G A T C C	
Footprints from MTV LTR	1.1	G G G A C A	G T G	G C T G G A	++
	1.2	A G A A C A	T T A	T T C T C C	++
	1.3	A G A A C A	G T T	T G T A A C	++++
	1.4	G A G A C A	A G T	G G T T T C	+
	1.5a	A G A A C C	T T T	G A T A C C	+++
	1.5b	A G A T C A	G A A	C C T T T G	
	1.5c	A G A A C A	C T C	A G A G C T	
	1.5d	A G A A C A	T A G	G A A A A T	

consult the data in Table 3, we find that four of the bases in site b should allow favored interaction with the protein (GGA-C-), and while the T in the fourth position should reduce binding moderately, it too is acceptable. The C in the sixth postition, however, is highly unfavored (>500 fold), which could explain the significantly lower affinity at site b that I observed. Nevertheless, site b has enough homology to the favored sequence that it can act as a "pseudosite" for binding - that is, it lacks the information necessary for specific binding to the protein without aid, (consider the absence of footprinting at site b on RBS13C-256), but with the additional binding energy supplied by cooperativity it can become a binding site.

I found that cooperativity improved binding at the RBS by ~150 fold, which is equivalent to ~3 kcal/mole and apparently enough to compensate for the "missing" contacts in site b. While I measured the effect of cooperativity using the shorter T7X556 protein, there may be additional receptor-receptor contacts in the full length molecule with an accompanying increase in the effect of cooperativity. Indeed, recent reports (Kumar and Chambon, 1988; Wrange et al., 1988) have suggested that steroid receptors dimerize even in the absence of DNA, (though this is clearly not true for T7X556; L Freedman, personal communication). Cooperativity has two related effects on receptor function. First, by increasing the affinity of the ternary complex,

cooperativity allows the receptor-DNA complex a longer time to effect its changes on the transcriptional apparatus. Second, cooperativity provides the receptor with greater selectivity for its functional sites. The receptor must appropriately occupy a small number of actual GREs without being soaked up by a sea of pseudosites in the genome (Yamamoto and Alberts, 1975). Cooperativity offers the protein an energetic incentive to selectively occupy properly-spaced symmetric binding sites, effectively increasing the size and specificity of the recognition sequence.

It does not follow from this, however, that function depends on perfect symmetry. Indeed, many functional binding sites do not display sequence symmetry. None of the identified binding sites in the upstream LTR of MTV, for example, has a symmetric favored binding sequence, and only one has a coupled site with better than 2/4 favored critical bases (see Table 5; Payvar et al., 1983; see also Miksicek et al., 1986). Yet, clustered base substitution experiments have shown that all five of the binding sites contribute to the GRE function of the LTR (D. DeFranco, personal communication). Perfectly symmetric sequences may simply provide the protein with very high affinity binding sites, resulting in enough occupancy to allow function in the absence of other receptor binding sites. Since maximal inducibility is not always advantageous to the organism, we

would not expect to find perfectly symmetric binding sites at all responsive genes. In evolutionary terms, we expect instead that sequences will drift away from maximal binding until optimal function is achieved.

Recognition Sequences and Context Both Contribute to Protein Binding

One DNA-binding model, which is based on models proposed for other sequence-specific transcriptional activators (Weber and Steitz, 1984), is illustrated in Figure 11. In this model, T7X556 initially associates with the DNA in a non-specific manner, perhaps on the basis of electrostatic interactions. The protein then "slides" along the DNA testing the major groove for binding sites, by looking for its complementary hydrogen bonding pattern. From my data, the sequence that best represents this pattern is 5'-[A/G]GAACAT-3'; however, affinity and footprinting data indicate that considerable variation at the non-underlined positions and limited variation at the underlined bases is also tolerated. Thus, the microscopic affinity constant, k_a, represents the protein-DNA affinity with optimal hydrogen bonding (~10^8 M^{-1}), while the non-specific microscopic affinity constant, k_{ns}, represents the average affinity of all non-cognate sites (~5 x 10^4 M^{-1}), which presumably range from partial to poor matches of the optimal hydrogen bonding pattern.

A second T7X556 molecule will bind to the DNA and scan

Figure 11. Model for Sequence-Specific Binding by T7X556

a) T7X556 initially associates with DNA non-specifically, perhaps by electrostatic interaction with the sugar-phosphate backbone, with an average affinity per site of $5 \times 10^4\ M^{-1}$. It then "scans" the major groove looking for a hydrogen bonding pattern that complements that of its sequence-specific binding determinants.

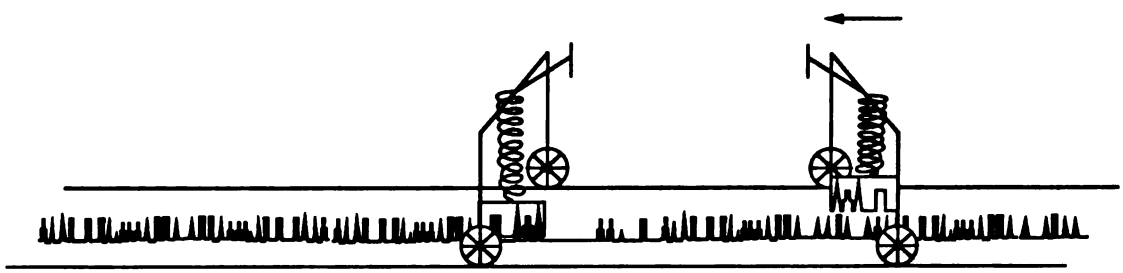
b) When the binding determinants find a matching pattern, the protein binds with high affinity, $k_a \approx 10^8\ M^{-1}$, thereby remaining at the site long enough for a second protein to interact with it. The second protein associates with and scans the DNA like the first.

c) Even when the second protein finds a suboptimal second binding site, as in RBSwt where $k_b \approx 10^6\ M^{-1}$, if it is properly spaced and oriented relative to the first, the ternary complex will be stabilized by cooperativity, $k_{ab} \approx 10^2$. Thus the overall affinity for the ternary complex can be expressed as $k_T = k_a k_b k_{ab} \approx 10^{16}\ M^{-2}$.

a

$\overline{k_{ns}} \cong 5 \times 10^4 M^{-1}$

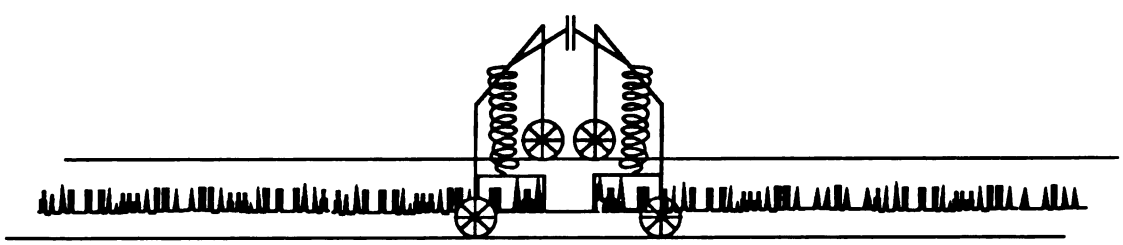
b



$k_a \cong 10^8 M^{-1}$ $\overline{k_{ns}} \cong 5 \times 10^4 M^{-1}$

c

$k_{ab} \cong 10^2$



$k_a \cong 10^8 M^{-1}$ $k_b \cong 10^6 M^{-1}$

it like the first; however, in addition to the bases in the major groove, it can also bind to the already bound T7X556 molecule. To acquire the benefit of binding to both, a second binding site must be correctly spaced and oriented to allow both sets of interactions. Thus, I would predict the optimal ternary complex site to be

5'-[A/G]GAACATAATGTTC[T/C]-3'

for which the predicted overall affinity would be $K_T = (k_a)^2 k_{ab} \approx 10^{18}\ M^{-2}$. This sequence agrees with that of functionally active oligonucleotides (Strähle et al., 1987). While I have not experimentally altered the spacing between the core binding sites, I and others (Scheiderit et al., 1984; Jantzen et al, 1987) have observed that it is consistently 3 base pairs. Since this is an odd number of bases, the binding site can never be truly symmetric. As discussed, such optimal binding sites are not usually favored biologically, and only rarely do both of the core sites match the predicted sequence. Nevertheless, as illustrated by these experiments, even the weaker affinity of a pseudosite substantially improves complex stability. Indeed, without the pseudosite the binding site is inactive *in vivo* (RBS490, Table 4).

Since the cooperative interaction between the T7X556 molecules appears to be DNA-induced, (consider the low abundance of protein dimers in solution), the core binding site may do more than simply act as a recognition sequence.

It might also induce a conformation of the protein that favors protein-protein interaction. In support of this, I did observe that certain substitutions in site a caused particularly strong reductions in the affinity of the <u>second</u> protein molecule (compare Q_2 values of RBS13 and RBS14 with other mutants). These mutations might indicate bases that are crucial for the cooperatively-active conformation.

The simple binding model described here provides an incomplete picture of receptor-DNA interactions. The sequence elements that it describes are evidently necessary for site-specific binding, but not sufficient for it. Both footprinting and affinity experiments demonstrated that T7X556 did not bind specifically to the first consensus sequence - a sequence predicted to be at least an adequate, if not an efficient, binding site. Even when I mutated the sequence to match precisely the optimal core sequence, the affinity did not increase and I did not observe footprinting at the site. Furthermore, I also observed that several substitutions outside the predicted recognition sequence caused reduced hormone induction in my transient assays, despite only modest effects on protein affinity. Thus, the protein appears to "look" for more than just a complementary hydrogen bonding pattern in the major groove; the sequences that flank the binding site must also be able to provide contexts that improve both site recognition and subsequent transcriptional activation. Indeed, protein contacts on DNA

flanking the binding site are confirmed by the hydroxyl radical footprints. Such contacts might alter not only the affinity but also the specificity of binding. It is thus possible that in different contexts even the relative importance of different consensus positions might change.

For T7X556, the flanking interactions are likely to be subtle, since mutations in this region had only modest effects on affinity. One possibility is that the protein recognizes some structural aspect of the double helix. Significant variations in the groove geometry and local helix bending have been observed and shown to be highly sequence dependent (Dickerson, 1983; Callidine and Drew, 1986; Yoon et al., 1988). Thus, even base pairs that do not directly contact protein can affect binding by altering the DNA conformation (Koudelka et al., 1987; Gartenberg and Crothers, 1988). Regardless of how the effect is mediated, by demonstrating how dramatically both binding affinity and *in vivo* activity are affected, my results highlight the importance of context in DNA recognition and the need for systematic comparisons within a constant context.

Methods and Materials

End Labeling and Purification of DNA Fragments

Plasmid DNA, which had been isolated by standard alkaline lysis and purified through two CsCl gradients, was restricted with the appropriate enzymes (Eco RI for footprinting assays and Eco RI/Bgl I (RBSx-256) or Eco RI/Hind III (RBSx-102) for affinity assays) and then treated with 0.1 units of calf intestinal phosphatase per μg of DNA for 30 min at 37°C in restriction buffer supplemented with 0.1 μM $ZnCl_2$ and 50 mM Tris pH 9.0 (all Tris buffers were pH adjusted at room temperature). After one extraction with phenol/$CHCl_3$/isoamyl alcohol (PCI, 25:24:1) and ethanol precipitation, the DNA was dissolved in 10 mM Tris pH 8.0, 1 mM EDTA (TE) and stored at 4°C for future use.

For 5' end labeling, 1-5 μg of the restricted DNA were incubated with 10 units T4 polynucleotide kinase (USB) and 160 μCi of γ-^{32}P-ATP (>7000 Ci/mmol, ICN) in 20 μl of 50 mM Tris pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine and 1 mM EDTA at 37°C for 1 hr. The reaction was stopped with 80 μl of 10 mM Tris pH 8.0, 10 mM EDTA, extracted 1x with PCI, and the DNA was precipitated by the addition of 2.5 volumes of 0.45M ammonium acetate in ethanol (Amm/EtOH) to remove the unincorporated radioactivity. The DNA was redissolved, (restricted with Bgl I for footprinting assays), and electrophoresed at 6V/cm on a 1.2% low melt agarose gel in

1x TAE buffer (40 mM Tris acetate, 2 mM EDTA). For most assays, the labeled DNA was isolated by excising the portion of the gel containing the relevant fragment, adding 3 volumes of 200 mM NaCl, 20 mM EDTA, melting the slice at 68°C, extracting it twice with phenol buffered at pH 8.0 and once with PCI, and precipitating the aqueous phase with 2.5 volumes of EtOH. However, for the DNAase I footpinting assays, the activated glass powder method was used (Vogelstein and Gillespie, 1979). In both cases the purified labeled fragment was dissolved in TE and stored at 4°C for up to two weeks.

For 3' end labeling, 1-5 μg of the restricted DNA were incubated with 5 units of DNA polymerase (Klenow fragment), and 100 μCi of α-^{32}P-dATP (6000 Ci/mmol, Amersham) in 20 μl of 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 500 μg/ml BSA. After 1 min at 25°C, 2 μl of 2.5 mM dNTP mix [A,C,G,T] were added and the reaction continued for 10 min at 25°C. The reaction was stopped with 80 μl of 10 mM Tris 8.0, 10 mM EDTA and the labeled DNA was isolated as described above.

DNAase I Footprinting

The quantities of T7X556 indicated in the text (purified as described in Freedman et al., 1988) were incubated with 20,000 cpm (Cerenkov) of end-labeled fragment in 100 μl of binding buffer (BB; 20 mM Tris pH 7.8, 2 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 200 μg/ml porcine insulin, 4

mM DTT, 0.1% NP-40) that also contained 2% (w/v) polyvinyl alcohol (PVA). After 15 min at room temperature, 100 μl of 10 mM $MgCl_2$, 5 mM $CaCl_2$ were added, followed, 1 min later, by the addition of 0.8 ng of DNAase I (DPFF, Worthington) to initiate the digestions. The reactions were stopped after 60 sec by the addition of 200 μl of stop mix (200 mM NaCl, 40 mM EDTA, 1% SDS, 250 μg/ml tRNA, 100 μg/ml proteinase K) and placed at 37°C for 15 min. Each sample was extracted once with PCI and precipitated from 2.5 volumes of EtOH. The samples were dissolved in formamide dye mix (95% formamide, 4% EDTA, 1% concentrated bromphenol blue and xylene cyanol mix (10% w/v in H_2O)), counted in a scintillation counter, and equal counts were electrophoresed on an 8% polyacrylamide sequencing gel (7M urea, 20:1 acrylamide:bisacrylamide) in 1x TBE buffer (89 mM Tris borate, 89 mM boric acid, 20 mM EDTA). Sequencing reactions for guanine-specific and cytosine- and thymidine-specific lanes were carried out according methods described in Maxam and Gilbert, 1980.

Hydroxyl Radical Footprinting

The indicated quantities of T7X556 were incubated with 20,000 cpm of end-labeled DNA fragment in 100 μl of 1x BB that contained 2% PVA but no glycerol for 15 min at room temperature. As mentioned in the text, H_2O_2, one of the reagents necessary for generating the free radicals, reduced

DNA binding by T7X556, thus requiring that the cleavage reaction be maximized for low concentrations of H_2O_2. Since the free radicals are short lived, I found it useful to initiate the reaction by mixing all of the separate reagents with the protein-DNA complexes simultaneously, thereby avoiding the decay of the free radicals that occurs when the active reagents are pre-mixed. To accomplish this, 5 μl of 78 mM $Fe(NH_2)_2(SO_4)_2 \cdot 6H_2O$ (made fresh), 5 μl of 156 mM EDTA, 10 μl of 65 mM NaAscorbate (made fresh), and 10 μl of 0.06% H_2O_2 were aliquoted to a petri dish, taking care to keep the reagents separated from one another. The reagents were then taken up, one by one, into the tip of an adjustable pipeting device allowing in ~25 μl of air after each to act as a separator. The last two reagents, the Fe and EDTA, were allowed to mix. The cleavage reactions were then initiated by injecting the reagents into the binding reaction and allowed to proceed at room temperature for 1 min. The reactions were terminated by the addition of 170 μl of stop mix, placed at 37°C for 15 min, and then purified and electrophoresed as described for DNAase I footprinting.

Electrophoresis in Nondenaturing Polyacrylamide Gels

The indicated amount of T7X556, T7X525, or a mixture of the two was incubated for 5 min at room temperature with one or two end-labeled DNA fragments in 10 μl of BB that also contained 1 μg of poly (d[I-C]) and a trace amount of marker

dye (bromphenol blue and xylene cyanol). The entire sample was loaded onto a running 7.8% nondenaturing polyacrylamide gel (75:1 acrylamide/bis), which also included 0.1% NP-40, in 0.25x TBE buffer. Typically, the gel was run at 350V in a room maintained at 4°C, after pre-running the gel for 30-60 min under the same conditions. After electrophoresis, the gels were dried and autoradiographed.

1,10 Phenanthroline Copper Footprinting

T7X556 was incubated with the indicated end-labeled DNA fragments and the resulting complexes were resolved on a nondenaturing polyacrylamide gel according to the procedure described, except that no tracking dye was used in lanes with complexes. The reactions were then initiated using slight modifications of the method described by Kuwabara and Sigman (1987). After electrophoresis, the top glass plate was removed, and the gel was transferred to a pyrex dish containing 200 ml of 10 mM Tris pH 8.0. While allowing the gel to equilibrate for 10 min at room temperature, 1 ml of 9 mM $CuSO_4$ (in H_2O) was mixed with 1 ml of 40 mM 1,10 phenanthroline (in 100% EtOH) and allowed to sit for 1 min until 18 ml of H_2O was added, and the entire solution was transferred to the gel. The reaction was then initiated by adding 100 μl of mercaptoproprionic acid (diluted fresh in 20 ml of H_2O) to the gel, and allowed to proceed for 1 min at room temperature without shaking. The reaction was

stopped with 10 ml of 1.16% 2,9 dimethyl 1,10 phenanthroline in 100% EtOH. After 2 min, the wet gel was autoradiographed and gel slices containing the indicated complexes were excised. These were soaked in 500 μl gel elution buffer (200 mM NaCl, 20 mM EDTA, 1% SDS, 10 μg/ml tRNA) overnight at 40°C. After discarding the gel fragments, the buffer was extracted with equal volumes of PCI and precipitated with 2.5 volumes of EtOH. The DNA was redissolved in formamide dye mix, and equal counts were loaded onto an 8% sequencing gel.

Methylation Interference

To partially methylate the DNA, 1 μl of dimethyl sulfate was added to 200 μl of TE containing the end-labeled DNA fragment and incubated for 4 min at room temperature. The reactions were stopped by the addition of 50 μl of G-STOP mix (1.5M NaAc, pH 7, 200 μg/ml tRNA, 7% 2-mercaptoethanol) and precipitated with 750 μl of EtOH (chilled). The DNAs were redissolved in TE, incubated with protein, and the resulting complexes were resolved on a nondenaturing polyacrylamide gel, as described. The wet gel was autoradiographed and gel fragments containing the different complexes were excised, and soaked in 500 μl gel elution buffer for 8-12 hr at 40°C. The gel fragments were then discarded and the buffer was extracted with equal volumes of PCI. The aqueous layers were precipitated with

2.5 volumes EtOH, and the DNAs were redissolved in 50 μl water. To this, 50 μl of 20% piperidine was added and the samples were heated to 90°C for 30 min. The DNA was precipitated by the addition of 100 μl of 600 mM NaAc and 500 μl of EtOH. The samples were redissolved in formamide dye mix, and equal counts were loaded onto an 8% sequencing gel.

Absolute Affinity Constant Determination

Measurement of the equilibrium affinity constant for T7X556 binding to RBSwt-256 was done as follows. A mixture of highly purified T7X556 (>99% pure) and purified RBSwt-256 in BB was serially diluted. To each dilution, a trace amount of end-labeled RBSwt-256 was added as an indicator of the fraction of DNA that was bound by protein, and the resulting complexes were resolved on a nondenaturing polyacrylamide gel. (In an independent experiment it was shown that the binding of protein to DNA was unaffected by the order of addition of labeled and unlabeled DNA.) After drying and autoradiography, the gel bands were quantified by counting excised gel slices in a Beckman LS 1801 scintillation counter (Cerenkov).

To analyze the data I assumed the following equilibrium:

$$D + P \overset{K_1}{=} D\text{-}P + P \overset{K_2}{=} D\text{-}P_2 + P \overset{K_3}{=} D\text{-}P_3$$

where D represents RBSwt-256 and P reprsents T7X556. This equilibrium is valid as long as the concentration of protein is small enough to avoid any significant concentration of complex 4. Under these conditions, we can write

$$[P_{active\text{-}free}] = P_{total\text{-}active} - [D\text{-}P] - 2[D\text{-}P_2] - 3[D\text{-}P_3]$$

which can be rearranged to

$$[P_{active\text{-}free}] = D_T\ (\alpha - (r_1 + 2r_2 + 3r_3)) \qquad (13)$$

where D_T is the total DNA, r_1, r_2, and r_3 represent the fraction of DNA shifted into complex 1, 2, and 3, respectively and α is $P_{total\text{-}active}/D_T$.

We can recast the equilibrium binding equations to be

$$K_1 = (D\text{-}P)/(D)(P) = [r_1/(1 - r_T)](P_{active\text{-}free})^{-1} \qquad (14)$$

$$K_2 = (D\text{-}P_2)/(D\text{-}P)(P) = [r_2/r_1](P_{active\text{-}free})^{-1} \qquad (15)$$

where $r_T = r_1 + r_2 + r_3$. We can then substitute equation (13) into equations (14) and (15) and with rearrangement we obtain

$$(\alpha - (r_1 + 2r_2 + 3r_3))(1 - r_T)/r_1 = (K_1)^{-1}(D_T)^{-1} \qquad (2)$$

$$(\alpha - (r_1 + 2r_2 + 3r_3))r_1/r_2 = (K_2)^{-1}(D_T)^{-1} \qquad (3)$$

The values of r_1, r_2, and r_3 are all experimentally determined for each value of D_T. When the left halves of equations (2) and (3) are plotted versus the $(D_T)^{-1}$, the slope of the resulting line is the reciprocal of the affinity constant, provided the line passes through the origin. Thus, the value of α was adjusted until the intercepts were 0 and the slopes were taken as the values of the dissociation constants. These experiments were repeated

three times and I report the mean ± SEM for all three.

Relative Affinity Constant Determination

Since there are many potential protein binding sites on RBSwt-256, complex 1 represents a heterogeneous population of protein-DNA complexes, all containing one molecule of T7X556 but varying in the site occupied by that molecule. The abundance of each species, $(D-P)_i$, depends on its microscopic affinity constant $k_i = (D-P)_i/[(D)(P)]$. We can express the concentration of all complex 1 species as the sum of the concentrations of each individual species as follows

$$(\text{complex } 1) = (D-P)_1 + (D-P)_2 + (D-P)_3 + \ldots + (D-P)_n \quad (16)$$

where n represents the number of potential binding sites on the fragment. We estimate that n = 2(length - 30) ≈ 450 assuming that the protein requires at least 15 bp to bind at the ends of the fragment (see figure 3). We can substitute this into the affinity expression to obtain

$$K_1 = (\text{complex } 1)/[(D)(P)]$$

$$= \{(D-P)_1 + (D-P)_2 + (D-P)_3 + \ldots + (D-P)_n\}/[(D)(P)] \quad (17)$$

which can be recast as

$$K_1 = k_1 + k_2 + k_3 + \ldots + k_n \quad (18)$$

The footprinting experiments suggested that two of the binding sites had an affinity significantly larger than the others. We can then simplify equation (18) by assuming an average nonspecific affinity at the rest of the sites to

give

$$K_1 = nk_{ns} + k_a + k_b \quad (4)$$

where k_{ns}, k_a, and k_b represent the microscopic affinity constant at nonspecific sites, site a, and site b, respectively, and n is the number of nonspecific binding sites.

Using similar reasoning, we observe that a second protein molecule can bind to the same three classes of binding site. As before, it can bind to a nonspecific site; however, there will be 30 fewer such sites since the first protein molecule will occlude that many bp, (assuming it blocks binding in both directions). It can also bind to site a, although for most of the molecules this site will already be occupied, since this is a high affinity binding site. To correct for this, we can adjust the contribution of site a using a term that expresses the fraction of molecules that have protein bound at sites other than site a:

$$Y_a = (nk_{ns} + k_b k_{ab})/K_1$$

which also includes the term k_{ab} representing the contribution of the cooperative interaction observed when T7X556 occupies sites a and b. Similarly, for site b

$$Y_b = (nk_{ns} + k_a k_{ab})/K_1$$

Thus we can express the equilibrium constant K_2 as

$$K_2 = (n - 30)k_{ns} + Y_a k_a + Y_b k_b \quad (5)$$

To determine the relationship between $Q_1 = K_{1wt}/K_{1mut}$ and

$q_a = k_{awt}/k_{amut}$, we can express Q_1 using equation (4) to obtain

$$Q_1 = \{nk_{ns} + k_a + k_b\}_{wt}/\{nk_{ns} + k_a + k_b\}_{mut}$$

For those mutants that differ from wild type only in site a and that were compared to RBSwt-102, this can be simplified to

$$Q_1 = \{150k_{ns} + k_{awt} + k_b\}/\{450k_{ns} + k_{amut} + k_b\}$$

If we substitute the values $k_a + k_b = 1900k_{ns}$ and $k_a = 58k_b$ and solve for k_{awt}/k_{amut}, we find that

$$q_a = k_{awt}/k_{amut} = Q_1/(1.1 - 0.26Q_1) \qquad (12)$$

Saturation Mutagenesis

The RBS derivatives were produced by chemically synthesizing the following oligonucleotide:

5'-GCCGTCTAGACTCGAG<u>GCTCAGATCAGAACATTT</u>GGATCCTCTAGAG-3'

For the substitution mutants, nucleoside phosphoramidite precursors that had been "doped" to a final concentration 1.7% of each of the three nondesignated bases were used at the underlined positions. Since the eight bases at the 3' end form a palindrome, two molecules of the oligomer can hybridize and act as mutual primers for second strand synthesis (Oliphant and Struhl, 1988). To mutually hybridize the molecules, 5 μg of the oligomer in 10 μl of 30 mM Tris pH 7.5, 150 mM NaCl, 30 mM $MgCl_2$, 15 mM DTT, 0.1 mg/ml BSA were incubated for 15 min at 37°C, and then allowed to cool slowly to room temperature. The second strand was synthesized by adding 16 μl H_2O, 3 μl of a

mixture of all four dNTPs (2.5 mM each), 1 μl of α-^{32}P-dGTP (10 mCi/ml, 3000 Ci/mmol), and 5 units of DNA polymerase I (Klenow fragement) and incubating the sample for 1 hr at 37°C. (By generating the second strand using the first as a template, any mutations introduced into the first strand by the "doping" procedure, were also introduced into the second strand.) The sample volume was increased to 100 μl with TE before extracting once with PCI and precipitation with 5 μl 3M NaAc and 2.5 volumes of EtOH. The sample was redissolved and restricted with 80 units of Xba I for 2 hr at 37°C. The sample was then electrophoresed on a 12% polyacrylamide gel in 1x TBE buffer; the appropriate band (36 bp) was excised and soaked overnight at 40°C in gel elution buffer. After discarding the gel fragments, the buffer was extracted with equal volumes of PCI and precipitated with 2.5 volumes of EtOH. The sample was redissolved in 100 μl TE and serial dilutions were then used in standard ligation reactions with 75 ng of Xba I restricted and phosphatased vector (pOTCO, Sakai et al., 1988). To minimize the number of double insertions, bacterial colonies were picked from plates representing the ligation reaction with the lowest concentration of insert that gave significantly more colonies than background.

Plasmid Sequencing

Potential mutants were screened by sequencing crude

miniprep DNA using a rapid sequencing protocol (described below). To reduce background signal and to simplify the overall procedure, we used end-labeled sequencing primer instead of incorporating radioactive nucleotides in the sequencing reactions. Plasmids containing mutations were then purified by alkaline lysis followed by two successive bandings in CsCl-ethidium bromide gradients. The concentrations and purity of all plasmids were verified by UV absorption and agarose gel electrophoresis. In addition, all plasmids were resequenced after preparation and purification.

In a final reaction volume of 20 μl, 40 ng of "universal" sequencing primer (M13, USB) was incubated with 2 μl of γ^{32}P-ATP (7000 Ci/mmol, 160 mCi/ml, ICN) and 1 μl of T4 polynucleotide kinase (USB, 30 U/μl) in kinase buffer (50 mM Tris pH 7.6, 10 mM $MgCl_2$, 0.1 mM spermidine, 0.1 mM EDTA, 5 mM DTT; which can be made as a 10x stock and frozen in aliquots) for one hr at 37°C. The reaction was stopped by the addition of 80 μl of 10 mM Tris pH 7.6, 10 mM EDTA, 80 μg/ml tRNA and then extracted once with PCI. To remove the unincorporated label, the aqueous fraction was precipitated three times by adding 2.5 volumes of 0.45 M NH_4Ac in EtOH and redissolving the pellet each time in 100 μl of TE. After the final precipitation the end-labeled primer was dissolved in 100 μl TE.

To 1.5 ml microfuge tubes, 2-3 μg of miniprep DNA were

added in a volume preferably less than 10 μl. To each plasmid, 50 μl of 0.2 M NaOH, 1 mM EDTA, 8 ng/ml end-labeled primer (~200 cps/sample by mini-monitor) were added and the samples were allowed to sit at room temperature for 5 min. The samples were neutralized by the addition of 20 μl of 5M NH_4Ac, 250 μg/ml tRNA and immediately precipitated by adding 200 μl of EtOH. The samples were vortexed and immediately centrifuged in a standard microfuge for 10 min. During this spin a 37°C and a 75°C temperature block were set up and filled to the brim with water. Each pellet was redissolved in 8 μl H_2O and 3 μl of a 5X Sequenase Buffer (200 mM Tris pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl)/0.1M DTT mix (2:1) were added to each.

Samples were placed at 37°C for 10 min to allow annealing of primer to template. During this incubation, 2.0 μl of the ddA, ddC, ddG, and ddT termination mixes (10 μM ddXTP [X = A, C, G, or T], 100 μM dATP, 100 μM dCTP, 100 μM dGTP, 100 μM dTTP, 62.5 mM NaCl) were aliquoted to the wells of a microtiter assay plate (such as the Falcon 3911 MicroTestIII). The samples were then moved to room temperature and allowed to sit for 5 min while the Sequenase was diluted 1:20 in cold TE to a final volume = (3 μl × the number of samples) + 5 μl.

The microtiter dish containing the aliquoted termination mixes was placed on top of the 37°C temperature block so that the nearly overflowing water contacted all of

the well bottoms. To start the reactions, 3 μl of the diluted Sequenase were added to the first plasmid, mixed well, and then 3 μl of this mix were then added to each of the wells containing the termination mixes for that plasmid. Reactions were allowed to proceed for 5 min, after which 4μl of formamide dye mix were added to each well. The microtiter dish was covered and placed on top of the 75°C temperature block ensuring good water contact as above, and the dish was heated for 2 min. The dish was placed on ice and the samples were loaded on a standard sequencing gel at 2.6 μl per lane, which gave an easily readable exposure overnight without an intensifying screen.

Cell Culture and Transfections

CV-1 monkey kidney cells were propagated in Dulbecco's modified Eagle medium containing 5% fetal bovine serum. Cells were maintained at 37°C in a humidified incubator (Rh 91%) maintained at 8% CO_2. All transfections were repeated 3-4 times, in separate experiments, and the results were averaged as described in the text.

In sterile 1.5 ml microfuge tubes, 0.8 μg of plasmid RSVGR (Sakai et al., 1988) and 2.4 μg of the indicated reporter plasmids were mixed in a final volume of 0.3 ml of 0.1x TE. To these samples, 100 μl of 1 M $CaCl_2$ was added, mixed, and allowed to sit for 5 min. Then 400 μl of 2x HBS buffer (50 mM HEPES pH 6.90, 280 mM NaCl, 1.5 mM

$Na_2HPO_4 \cdot 7H_2O$) were added and the tubes were vortexed vigorously and allowed to sit for 15 min at room temperature. Half of each sample was then added to each of two paired cell culture plates (2×10^5 cells per 6 cm dish), which had been freshly fed with 4 ml of growth medium. The calcium phosphate-DNA precipitate was allowed to remain on the cells for 12-16 hr, after which the medium was removed, the cells were rinsed once with TBS buffer (10 mM Tris pH 7.4, 150 mM NaCl), and fresh medium with or without 0.1 μM dexamethasone was added to the paired plates. Cells were then incubated for 24 hr before harvesting.

Extract Preparation and CAT Assays

Cells were rinsed once with TBS·E buffer (10 mM Tris pH7.4, 150 mM NaCl, 2 mM EDTA), and using a rubber policeman, they were lifted from the culture dish in 1 ml of the same buffer and transferred to 1.5 ml microfuge tubes. The cells were pelleted by cetrifugation at 5,000 rpm for 1 min in a microfuge and resuspended in 125 μl of 0.25 M Tris pH 7.4. In this buffer, the cells were then lysed by 3 freeze-thaw cycles in a dry ice/EtOH bath and a 68°C water bath, respectively, ensuring that the samples were maintained at the high temperature for a total of 10 min. The membranes and insoluble protein were pelleted by centrifugation at 10,000 rpm for 10 min in a microfuge, and the supernatant was transferred to a clean 1.5 ml microfuge

tube. The protein concentration of each sample was assayed according to Bradford (1976).

The enzymatic assay was run using equal quantities of extract protein, typically 2 μg, which was brought up to a final volume of 100 μl with 0.25 M Tris pH7.4. To this was added 4 μl of ^{14}C-chloramphenicol (~0.2 μCi/μl, Amersham), 10 μl of 20 mM acetyl CoA, and 66 μl of H_2O. The samples were incubated at 37°C for 5 hr and the reactions were terminated by the addition of 600 μl of ethyl acetate. The samples were shaken vigorously, centrifuged at 10,000 rpm for 1 min in a microfuge, and the organic phases were recovered. They were then rotoevaporated until dry, and redissolved in 30 μl of $CHCl_3$. The samples were then spotted onto a 20 cm × 20 cm silica gel thin layer chromatography (TLC) plate, and the substrates and products were resolved in a freshly made 95:5 $CHCl_3$:MeOH buffer system. The results were visualized by autoradiography of the TLC plates, and quantitated by scintillation counting of substrate and product spots in 5 ml of scintillant (Safety Solve, RPI).

Bibliography

Berg, J.M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. USA 85, 99-102.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of potein utilizing the principle of protein-dye bidning. Anal. Biochem. 72, 248-254.

Calladine, C.R. and Drew, H.R. (1986). Principles of sequence-dependent flexure of DNA. J. Mol. Biol. 192, 907-918.

Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983). DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. Cell 33, 489-499.

Derbyshire, K.M., Salvo, J.J., and Grindley, N.D.F. (1986). A simple and efficient procedure for saturation mutagenesis using mixed oligodeoxynucleotides. Gene. 46, 145-152.

DeFranco, D., and Yamamoto, K.R. (1986). Two different factors act separately or together to specifiy functionally

distinct activities at a single transcriptional enhancer. Mol. Cell. Biol. 6, 993-1001.

Dickerson, R.E. (1983). Base sequence and helix structure variation in B and A DNA. J. Mol. Biol. 166, 419-441.

Ebright, R.H., Cossart, P., Gicquel-Sanzey, and Beckwith, J. (1984). Molecular basis of DNA sequence recognition by the catabolite gene activator protein: Detailed inferences from three mutations that alter DNA sequence specificity. Proc. Natl. Acad. Sci. USA 81, 7274-7278.

Evans, R.M. (1988). The steroid and thyroid hormone receptor superfamily. Science 240, 889-895.

Fairall, L., Rhodes, D., and Klug, A. (1986). Mapping of the sites of protection on a 5S RNA gene by the Xenopus transcription factor IIIA - A model for the interaction. J. Mol. Biol. 192, 577-591.

Freedman, L.P., Luisi, B.F., Korzun, Z.R., Basavappa, R., Sigler, P.B., and Yamamoto, K.R. (1988). The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. Nature 334, 543-546.

Gartenberg, M.R., and Crothers, D.M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. Nature 333, 824-829.

Glantz, S.A. (1981) Primer of Biostatistics. McGraw-Hill, Inc., New York.

Godowski, P.J., Picard, D., and Yamamoto, K.R. (1988). Signal transduction and transcriptional regulation by glucocorticoid receptor - LexA fusion proteins. Science 241, 4812-4816.

Green, S. and Chambon, P. (1987). Estradiol induction of a glucocorticoid-responsive gene by a chimeric receptor. Nature. 325, 75-78.

Green, S., Kumar, V., Theulaz, I., Wahli, W. and Chambon, P. (1988). The N-terminal DNA-binding "zinc finger" of the estrogen and glucocortcoid receptors determines target gene specificity. EMBO J. 7, 3037-3044.

Jantzen, H.-M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R., and Schütz, G. (1987). Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. Cell 49, 29-38.

Klein-Hitpass, L., Ryffel, G.U., Heitlinger, E., and Cato, A.C.B. (1988). A 13 bp palindrome is a functional estrogen responsive element and interacts specifically with estrogen receptor. Nucl. Acids. Res. 16, 647-663.

Klock, G., Strähle, U., and Schütz, G. (1987). Estrogen and glucocorticoid responsive elements are closely related but distinct. Nature 329, 734-736.

Koudelka, G.B., Harrison, S.C., and Ptashne, M. (1987). Effect of non-contacted bases on the affinity of 434 repressor and Cro. Nature 326, 886-888.

Kumar, V. and Chambon, P. (1988). The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell 55, 145-156.

Kuwabara, M.D., and Sigman, D.S. (1987). Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: Escherichia coli RNA polymerase-lac promoter complexes. Biochemistry 26, 7234-7238.

Liu-Johnson, H.-N., Gartenberg, M.R., and Crothers, D.M. (1986). The DNA binding domain and bending angle of E. coli

CAP protein. Cell 47, 995-1005.

Martinez, E., Givel, F., and Wahli, W. (1987). The estrogen-responsive element as an inducible enhancer: DNA sequence requirements and conversion to a glucocorticoid-responsive element. EMBO J. 6, 3719-3727.

Maxam, A.M., and Gilbert, W.C. (1980). Sequencing end-labeled DNA wit base-specific chemical cleavages. Meth. Enzymol. 65, 497-559.

McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell 25, 385-398.

Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikström, A.-C., Gustafsson, J.-Å., and Yamamoto, K.R. (1986). Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. Cell 46, 389-399.

Miesfeld, R., Godowski, P.J., Maler, B.A., and Yamamoto, K.R. (1987). Glucocorticoid receptor mutants that define a small region sufficient for enhancer activation. Science 236, 423-427.

Miesfeld, R., Sakai, D., Inoue, A., Schena, M., Godowski, P.J., and Yamamoto, K.R. (1988). Glucocorticoid receptor sequences that confer positive and negative transcriptionl regulation. In Steroid Hormone Action, UCLA Symposium on Molecular and Cellular Biology (ed. G. Ringold). Alan R. Liss, Inc., New York. 193-200.

Miksicek, R., Heber, A., Schmid, W., Danesch, U., Posseckert, G., Beato, M., and Schütz, G. (1986). Glucocorticoid responsiveness of the transriptional enhancer of moloney murine sarcoma virus. Cell 46, 283-290.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4, 1609-1614.

Oliphant, A., and Struhl, K. (1988). Defining the consensus sequences of E. coli promoter elements by random selection. Nucl. Acid. Res. 16, 7673-7683.

Pabo, C.O., and Sauer, R.T. (1984). Protein-DNA recognition. Ann. Rev. Biochem. 53, 293-321.

Payvar, F., Wrange, Ö., Carlstedt-Duke, J., Okret, S., Gustafsson, J.-Å., and Yamamoto, K.R. (1981). Purified

glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo. Proc. Natl. Acad. Sci. USA 78, 6628-6632.

Payvar, F., DeFranco, D., Firesone, G.L., Edgar, B., Wrange, Ö., Okret, S., Gustafsson, J.-Å., and Yamamoto, K.R. (1983). Sequence-specific binding to MTV DNA at sites within and upstream of the transcribed region. Cell 35, 381-392.

Párraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., Klevit, R.E. (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. Science 241, 1489-1492.

Picard, D., and Yamamoto, D.R. (1987). Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. EMBO J. 6, 3333-3340.

Picard, D., Salser, S.J., and Yamamoto, K.R. (1988). A movable and regulable inactivation function within the steroid binding domain of the glucocorticoid receptor. Cell 54, 1073-1080.

Rhodes, D. and Klug, A. (1986). An underlying repeat in some transcriptional control sequences corresponding to half a

double helical turn of DNA. Cell 46, 123-132.

Rusconi, S., and Yamamoto, K.R. (1987). Functional dissection of the homone and DNA binding activities of the glucocortioid receptor. EMBO J. 6, 1309-1315.

Sakai, D.S., Helms, S., Carlstedt-Duke, J., Gustafsson, J.-Å., Rottman, F.M., and Yamamoto, K.R. (1988). Hormone-mediated repression: a negative glucocorticoid response element from the bovine prolactin gene. Genes and Dev. 2, 1144-1154.

Scheidereit, C., Geisse, S. Westphal, H.M., and Beato, M. (1983). The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumor virus. Nature 304, 749-752.

Scheidereit, C., Westphal, H.M., Carlson, C., Bosshard, H., and Beato, M. (1986). Molecular model of the interaction between the glucocorticoid receptor and the regulatory elements of inducible genes. DNA 5, 383-391.

Schleif, R. (1988). DNA binding by proteins. Science 241, 1182-1187.

Severne, Y., Wieland, S., Schaffner, W., and Rusconi, S.

(1988). Metal binding "finger" structures in the glucocorticoid receptor defined by site-directed mutagenesis. EMBO J. 7, 2503-2508.

Strähle, U., Klock, G., and Schütz, G. (1987). A DNA sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression. Proc. Natl. Acad. Sci. USA 84, 7871-7875.

Suck, D. and Oefner, C. (1986). Structure of DNAase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA. Nature 321, 620-625.

Suck, D., Lahm, A., and Oefner, C. (1988). Structure refined to 2 Å of a nicked DNA octanucleotide complex with DNAase I. Nature 332, 464-468.

Tsai, S.Y., Carlstedt-Duke, J., Weigel, N.L., Dahlman, K., Gustafsson, J.-Å., Tsai, M., and O'Malley, B.W. (1988). Molecular interactions of steroid hormone receptor with its enhancer element: evidence for receptor dimer formation. Cell 55, 361-369.

Tullius, T.D., and Dombroski, B.A. (1986). Hydroxyl radical "footprinting": high resolution information about DNA-protein contacts and application to lambda repressor and Cro

protein. Proc. Natl. Acad. Sci. USA 83, 5469-5473.

Tullius, T.D., Dombroski, B.A., Churchill, M.E.A., and Kam, L. (1988). Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. Meth. Enzymol. Recombinant DNA part E, in press.

Veal, J.M., and Rill, R.L. (1988). Sequence specificity of DNA cleavage by bis(1,10-phenanthroline)copper(I). Biochemistry 27, 1822-1827.

Vogelstein, B. and Gillespie, D. (1979). Preparative and analytical purification of DNA fragments from agarose. Proc. Natl. Acad. Sci. USA. 76, 615-619.

Vrana, K.E., Churchill, M.E.A., Tullius, T.D., and Brown, D.D. (1988). Mapping functional regions of transcription factor TFIIIA. Mol. Cell. Biol. 8, 1684-1696.

Weber, I.T., and Steitz, T.A. (1984). A model for the non-specific binding of catabolite gene activator protein to DNA. Nucl. Acid. Res. 12, 8475-8487.

Wrange, Ö., Eriksson, P., and Perlmann, T. (1989). The purified activated glucocorticoid receptor is a homodimer. J. Biol. Chem. 264, 5253-5259.

Yamamoto, K.R., and Alberts, B. (1975). The interaction of estradiol-receptor protein with the genome: an argument for the existence of undetected specific sites. Cell 4, 301-310.

Yamamoto, K.R. (1985). Steroid receptor regulated transrciption of specific genes and gene networks. Ann. Rev. Genet. 19, 209-252.

Youderian, P., Bouvier, S. and Susskind, M.M. (1982). Sequence determinants of promoter activity. Cell. 30, 843-853.

Yoon, C., Privé, G.G., Goodsell, D.S., and Dickerson, R.E. (1988). Structure of an alternating-B DNA helix and its relationship to A-tract DNA. Proc. Natl. Acad. Sci. USA 85, 6332-6336.

Zaret, K.S., and Yamamoto, K.R. (1984). Reversible and persistent changes in chromatin structure accompany activation of glucocorticoid-dependent enhancer element. Cell 38, 29-38.

UNIVERSITY OF CALIFORNIA · UC San Francisco LIBRARY

UC San Francisco LIBRARY
UNIVERSITY OF CALIFORNIA

FOR REFERENCE

NOT TO BE TAKEN FROM THE ROOM

BRODART CAT. NO. 23 012 PRINTED IN U.S.A.