Hayes, J.J. & Tullius, T.D. The missing nucleoside experiment: a new technique to study recognition of DNA by protein. Biochemistry 28, 9521-9527

ARTICLE *in* BIOCHEMISTRY · NOVEMBER 1989

Impact Factor: 3.02 · DOI: 10.1021/bi00450a041 · Source: PubMed

2 AUTHORS:

Jeffrey J Hayes
University of Rochester
111 PUBLICATIONS **6,155** CITATIONS
SEE PROFILE

Thomas D Tullius
Boston University
92 PUBLICATIONS **9,030** CITATIONS
SEE PROFILE

The Missing Nucleoside Experiment: A New Technique To Study Recognition of DNA by Protein†

Jeffrey J. Hayes and Thomas D. Tullius*

Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218

Received June 12, 1989; Revised Manuscript Received July 28, 1989

ABSTRACT: We report a new technique for quickly determining which nucleosides in a DNA molecule are contacted by a sequence-specific DNA-binding protein. Our method is related to the recently reported "missing contact" experiment [Brunelle, A., & Schleif, R. F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6673–6679]. We treat the DNA molecule with the hydroxyl radical to randomly remove nucleosides. The ability of protein to bind to gapped DNA is assayed by gel mobility shift. Nucleosides important to protein binding are identified by sequencing gel electrophoresis. The missing nucleoside experiment can be used to scan a DNA molecule at single-nucleotide resolution in one experiment. The bacteriophage λ repressor–O_R1 and cro–O_R1 complexes were analyzed to evaluate the method. For both proteins, the most important contacts are located in the protein monomer that binds to the consensus half of the operator. These contacts correspond well to those found by mutational studies, and in the cocrystal structure of the λ repressor–operator. The missing nucleoside data show that the amino-terminal arms of λ repressor make energetically important contacts with positions 7 and 8 and the central dyad base pair of the operator. The amino-terminal arm that makes the most extensive contacts to DNA appears to be the one that emanates from the repressor monomer that binds to the consensus half of the operator, in agreement with the cocrystal structure. The λ cro protein does not have an amino-terminal arm, and the missing nucleoside experiment clearly shows a lack of contacts to DNA in the central region of the operator in this complex.

The molecular interactions responsible for recognition of DNA by proteins are beginning to be unravelled. Recent crystallographic analyses of cocrystals of DNA and protein have provided the most detailed look at these interactions (Jordan & Pabo, 1988; Anderson et al., 1987; McLarin et al., 1986). However, since crystal structures are not available for most protein–DNA complexes, other methods are needed for determination of the loci of specific protein–DNA recognition.

Mutational analysis is a powerful adjunct to direct structural studies. Site-directed mutagenesis can be used to investigate contacts with protein at each base pair in a recognition sequence, but this approach requires many time-consuming manipulations to scan even a small binding site. Very often, only one member of a base pair is involved in specific recognition. Since mutagenesis changes the identity of both bases in a base pair, it is not easy to determine which base is the important contact (McClarin et al., 1986; Jordan & Pabo, 1988).

A related approach, chemical interference, involves assessing the effect of chemical modification of the DNA on protein binding. In these experiments, the DNA molecule is alkylated either at the backbone phosphates (ethylation interference) or at the bases (methylation interference), before protein binding is attempted. Modified positions that interfere with protein binding are revealed upon induction of backbone cleavage at the alkylated sites of the DNA molecules that were unable to bind to protein. Ethylation interference experiments (Siebenlist & Gilbert, 1980) can reveal positions of close association of the protein with the phosphates of the DNA backbone. Methylation interference experiments show which guanine and adenine bases are contacted in the major and minor groove, respectively (Siebenlist & Gilbert, 1980). Although information on base-specific contacts is available from methylation interference studies, only contacts with purines can be detected.

Brunelle and Schleif (1987) recently introduced the "missing contact" method to determine the bases that contact protein. This technique can be used to assess the contribution to protein binding of each member of a base pair independently, at virtually all of the nucleotides in a DNA molecule. In this experiment, bases were chemically removed from DNA containing the binding site for the bacteriophage λ repressor. Reaction conditions were adjusted so that any individual DNA molecule suffered at most the loss of a single base. Repressor was then added to the sample of modified DNA. It was assumed that if a base important to binding was missing in a particular DNA molecule, protein would not bind. The bases found by this experiment to make contact with repressor were in general accord with those implicated by crystallographic and mutagenesis experiments. In the missing contact method, Maxam–Gilbert sequencing chemistry is used to remove the bases; three separate experimental samples are needed to determine the bases important for protein binding.

We report that randomly generated single-stranded *gaps* in a DNA molecule, introduced by treatment of the DNA with the hydroxyl radical, in a similar way reveal the *nucleosides* that make contact with a DNA-binding protein. A single chemical reaction, and thus a single lane on a sequencing gel, is all that is needed for the determination. A large DNA molecule can be quickly scanned for the bases that are responsible for specific recognition by a protein. We use this method, which we call the "missing nucleoside" experiment, to analyze two well-understood protein–DNA complexes, the complexes of the bacteriophage λ repressor and cro proteins

†This work was supported by Grants GM40894 and GM41930 (T.D.T.) from the National Institutes of Health. T.D.T. is a fellow of the Alfred P. Sloan Foundation, a Camille and Henry Dreyfus Teacher–Scholar, and the recipient of a Research Career Development Award from the National Institutes of Health (CA01208).

0006-2960/89/0428-9521$01.50/0

with the O_R1 binding site. We compare our results to protein–DNA contacts detected crystallographically and genetically. Our results demonstrate the validity of the missing nucleoside technique, and in addition add to our understanding of the protein–DNA contacts in these systems. While this paper was in preparation, a preliminary report of a similar experiment on progesterone receptor binding to DNA (Chalepakis & Beato, 1989) appeared.

Experimental Procedures

Materials. Plasmid pOR1 was purified as described previously (Tullius & Dombroski, 1986). λ repressor and λ cro proteins were generous gifts from Dorothy Beckett and Gary Ackers, and Carl Pabo, respectively. The sequences of the top and bottom strands of the O_R1 operator site are

TACCTCTGGCGGTGATA (top strand)

ATGGAGACCGCCACTAT (bottom strand)

Radioactive Labeling. Plasmid pOR1 was cleaved at its unique *Bgl*II site, radioactively labeled by standard methods, and subsequently cut with *Hpa*II to generate a 200 base pair long singly end labeled restriction fragment that contains the O_R1 operator site. We found that the ends of the fragment must be further than 15–20 base pairs from the binding site to be assayed, to prevent dissociation of short oligomers from the gapped DNA molecule (Rimphanitchayakit et al., 1989).

Hydroxyl Radical Cleavage. The labeled DNA molecule was randomly gapped by reaction with the hydroxyl radical. We use the $[Fe(EDTA)]^{2-}/H_2O_2$/ascorbate reagent system to generate the hydroxyl radical in solution. The conditions of cleavage were as described (Tullius & Dombroski, 1985), except that the final concentration of Fe(II) was 100 μM. After cleavage, the DNA was precipitated in ethanol twice and then dissolved in a buffer consisting of 10 mM Tris·HCl/0.1 mM EDTA (pH 8.0).

Formation of Protein–DNA Complexes. Buffers for protein binding were prepared as described (Tullius & Dombroski, 1986). Protein was added to labeled, gapped DNA that had been dissolved in binding buffer. DNA bound to protein was separated from free DNA on a native polyacrylamide (mobility shift) gel as described (Wolffe, 1988). The amount of protein or of unlabeled competitor DNA in the binding mixture was adjusted to give approximately 95% formation of complex as judged by the intensities of the DNA bands on the mobility shift gel. Mobility shift gels in the λ repressor experiments were run at 4 °C, while those in the experiments with cro protein were carried out at 25 °C. Radioactive bands containing bound and free DNA were excised from the gel, and the DNA were eluted. DNA was then applied to a denaturing polyacrylamide electrophoresis (sequencing) gel.

Results

The Missing Nucleoside Experiment. The hydroxyl radical cleavage reaction was used to generate DNA fragments which contained on average fewer than one randomly placed one-nucleoside gap per fragment. The hydroxyl radical effects oxidative cleavage of the DNA phosphodiester backbone by reaction with the deoxyribose residues in the DNA strand. The frequency of cleavage does not depend on the base or sequence, and is nearly equivalent for all sites in linear double-stranded DNA free in solution (Tullius & Dombroski, 1985). Although the resultant strand break produced by the hydroxyl radical is often referred to as a "nick" (Chalepakis & Beato, 1989), it is, in fact, a one-base gap with predominantly phosphomonoester termini (G. E. Shafer and T. D. Tullius, unpublished results). We therefore call this the "missing nucleoside" experiment, because hydroxyl radical treatment results in the loss of a nucleoside (a base and its attached deoxyribose) from the DNA molecule.

The DNA molecule we used in these experiments was derived from plasmid pOR1 (Johnson, 1980) and contained the O_R1 binding site for λ repressor and cro. The gapped DNA molecule was mixed with λ repressor or cro, and DNA bound to protein was then separated from free DNA by electrophoresis on a native polyacrylamide gel. Bands containing bound and free DNA were excised from the native gel, and the DNA was eluted and run on a sequencing electrophoresis gel for determination of the cleavage patterns in the bound and free samples. Since the DNA strand is already broken in the gapping reaction, no other treatment is necessary to "develop" the pattern.

Gaps in the DNA molecule that interfere with binding enough that the protein and DNA no longer migrate as a complex on the gel are detected by the experiment. A low-intensity or missing band on the sequencing gel in the lane containing DNA that was bound to protein, or a high-intensity band in the lane containing free DNA, identifies a nucleoside that is important to formation of the protein–DNA complex. Conversely, uncontacted nucleosides within the binding site also yield a positive experimental signal, but with the opposite pattern of band intensity. This is in contrast to protection ("footprinting") methods, which give a negative result for bases not involved in protein contact.

One must be aware that introducing gaps in a DNA molecule might influence protein binding by leading to structural changes in the DNA backbone (Koudelka et al., 1987), as well as by removing specific nucleoside contacts. While nearly all of the missing nucleoside signals we observe for λ repressor and cro can be interpreted as due to the loss of a contact with a nucleoside, effects of changes in DNA structure should also be considered for other protein–DNA systems.

Results for λ Repressor and cro. The autoradiograph from a typical experiment with λ repressor is shown in Figure 1. The lanes which contain bound and free DNA, fractionated from the same sample [compare lanes 3 and 4 (lanes numbered from left to right), Figure 1], clearly give complementary information. Densitometer scans of lanes on the autoradiograph (Figure 2) show that positions indicated to be contacts by the missing nucleoside experiment are offset one or two bases to the 3′ side of regions protected from hydroxyl radical cleavage in a footprinting experiment (Tullius & Dombroski, 1986). Figures 3 and 4 show analogous experimental results for cro protein.

The data from the densitometer scans of the λ repressor and cro experiments (Figures 2 and 4) are mapped on the O_R1 sequence in Figure 5. Strongly contacted nucleosides are marked by vertical bars. With repressor, for example, these nucleosides are found at positions T1′, A2′, G-dyad, C8, G7, G6, and T5 on the top strand, and at T1, A2, A7′, G6′, and A5′ on the bottom strand. Positions C4′, T7′, and G8′ on the top strand, and G4′ and C7 on the bottom strand, show weaker but still significant contact signals. At positions where a very weak signal is obtained, the clevage frequencies of neighboring sites in the free DNA control (Figure 1, lanes C) must be taken into account. With such an analysis, the loss of nucleosides C3′, C6′, T7′, A3, and A1 appears to have a weak but detectable effect on binding of repressor. At a few sites, no unequivocal decision can be reached based on the data.

The missing nucleoside experiment also allows identification of nucleosides not contacted by the protein. These positions

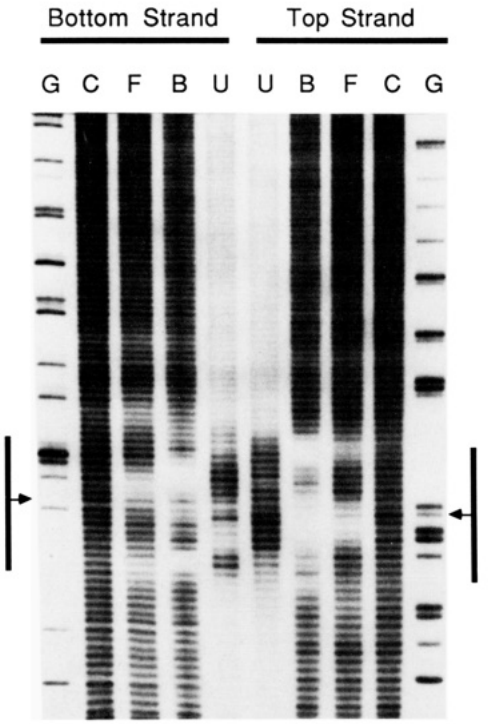

FIGURE 1: DNA sequencing gel revealing bases contacted and uncontacted by λ repressor protein when complexed with the O_R1 binding site. Lanes G are Maxam–Gilbert G-specific sequencing reactions performed on the intact fragment radioactively labeled on the top or bottom strand (see Experimental Procedures) as indicated. Lanes marked C (control) and F (footprint) represent the hydroxyl radical cleavage patterns of the fragment in the absence and presence of bound protein, respectively. Lanes B and U are the bound and unbound DNA fractions, respectively, which are derived from the missing nucleoside experiment (see text). The arrows point to the central base in the operator, and the vertical bars span the 17 bp operator sequence.

are indicated in Figure 5 by horizontal lines. Remarkably, we find that some sites *within* the operator could be gapped and still bind repressor or cro. Any position outside the operator sequence, including those immediately adjacent, could be removed without effect on binding of repressor or cro. This is in contrast to results from the missing contact experiments of Brunelle and Schlief (1987), in which repressor binding was reduced by loss of certain of these bases.

λ repressor and cro bind as dimers to the operator sequence. Thus, one would expect the pattern from an analytical method which is sensitive to base contacts also to be symmetrical about the (pseudo)dyad in the operator. The cro pattern clearly shows this symmetry (Figure 5). As well, many of the nucleosides that are contacted (and not contacted) by repressor have symmetry-related counterparts on the opposite strand (Figure 5). However, we also find asymmetry at a few sites. The large contact signals for repressor on the top strand at C8, dyad-G, and G8′ have no counterparts on the bottom strand. G4 on the top strand yields a small uncontacted signal, while G4′ on the bottom is strongly contacted. T5′ (top strand) gives a strong uncontacted signal, but position A5 (bottom) gives only a weak uncontacted signal. Position A5 is flanked by two uncontacted bases (C6 and C4), while T5′ is flanked by two bases that give moderate contacted signals (C4′ and C6′).

Discussion

We developed the missing nucleoside method with the λ repressor–cro–O_R1 system because of the wealth of data in the literature on protein–DNA contacts for these proteins with the bacteriophage λ operators. We focus our comparisons here

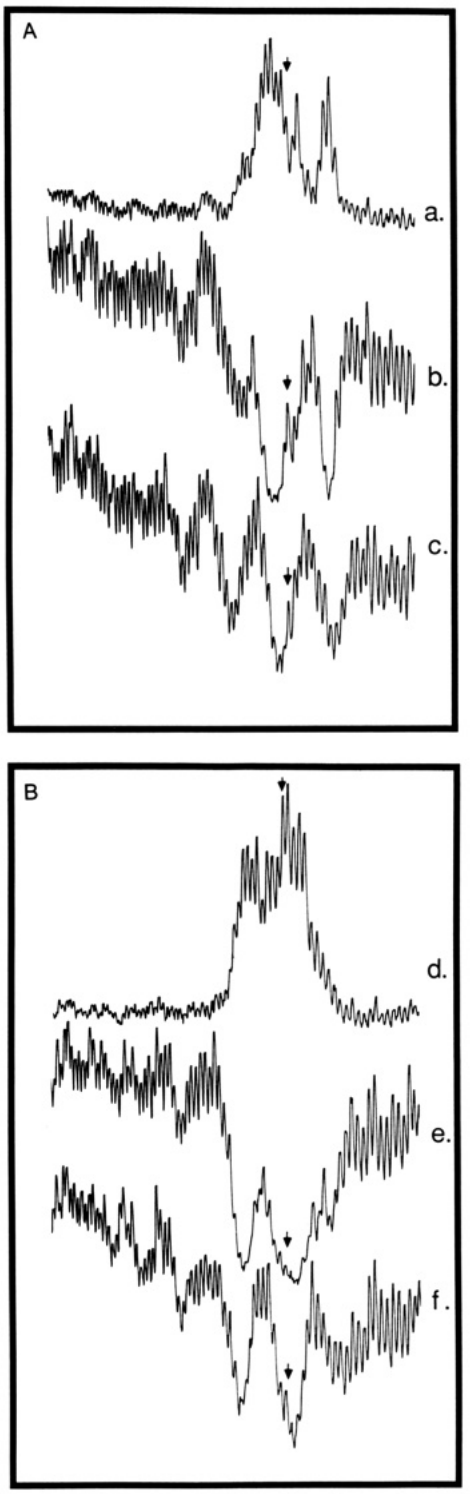

FIGURE 2: Desitometer scans of the λ repressor–O_R1 missing nucleoside experiment shown in Figure 1, in which the bottom strand (A) or the top strand (B) of the operator was radiolabeled. (A) Scans a, b, and c correspond to lanes U, B, and F ("bottom strand"), respectively. (B) Scans d, e, and f correspond to lanes U, B, and F, respectively, from the lanes marked "top strand" in Figure 1. Arrows mark the central base in the 17 bp O_R1 operator sequence.

on four previous reports: saturation point mutagenesis studies of the binding of λ repressor (Sarai & Takeda, 1987) and cro (Takeda et al., 1989) to O_R1, the missing contact study of Brunelle and Schleif (1987) on repressor, and the λ repressor–O_L1 cocrystal structure recently published by Jordan and Pabo (1988).

The six operators of phage λ each have one half-site that is an exact (or nearly exact) consensus sequence and another

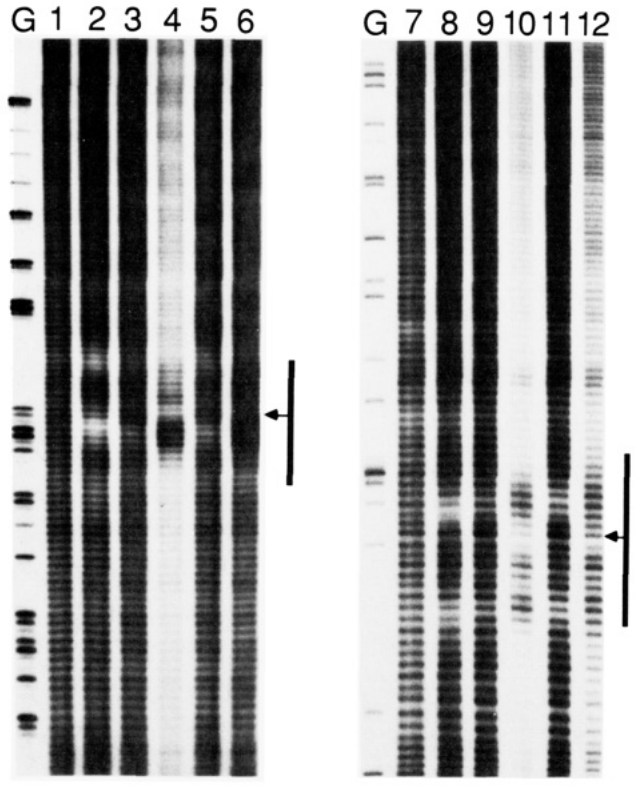

FIGURE 3: Autoradiograph of a sequencing gel on which was separated the products from a missing nucleoside analysis of cro protein binding to O_R1. Left (lanes 1–6) and right (lanes 7–12) autoradiographs are from experiments in which the top and bottom strands, respectively, were radiolabeled. Lanes marked G, Maxam–Gilbert G-specific sequencing reactions; lanes 1 and 7, hydroxyl radical control digestion of naked DNA; lanes 2 and 8, hydroxyl radical footprint of cro protein; lanes 3 and 9, bound fractions from the missing nucleoside experiment; lanes 4 and 10, unbound fraction complementary to samples in lanes 3 and 9; lanes 5 and 11, same as lanes 3 and 9 except that 2 μg of nonspecific unlabeled competitor DNA was added before separation of bound from unbound complexes on the native gel; lanes 6 and 12, unbound fraction complementary to samples in lanes 5 and 11.

half-site that diverges more from the consensus. The O_R1 operator we used conforms to this sequence arrangement. Since the results of the missing nucleoside experiments on repressor and cro show some degree of symmetry about the dyad, our discussion will focus on the consensus half of O_R1. Unless otherwise noted, the discussion of a particular position will apply to the two symmetry-related base pairs.

Contacts of λ Repressor with O_R1. Base pair A·T1 shows a relatively strong contacted signal for the thymine and a weak but detectable signal for the adenine. Brunelle and Schleif (1987) observed a similar pattern in their missing contact study. The cocrystal structure (Jordan & Pabo, 1988) suggests that the 5-methyl group of the T is involved in creating a hydrophobic pocket which could associate with the γ carbon of Gln-44 and the methylene carbons of Glu-34 (Jordan & Pabo, 1988). Indeed, substitution of uracil for T1 reduces the binding energy of repressor as much or more than any other mutation at this position (Sarai & Takeda, 1987). However, these interactions are predicted to be weak based on distances found in the crystal. We also observe an identical, large signal at the symmetry-related nucleoside T1′, although Brunelle and Schleif (1987) report no contact at T1′. Thus, it would seem that the proposed hydrophobic interactions mediated by T1 might be more important than they appear in the crystal structure. An alternative explanation for the very large signal at T1 could involve the additive effect of a moderate missing nucleoside signal coupled with a moderate signal due to the

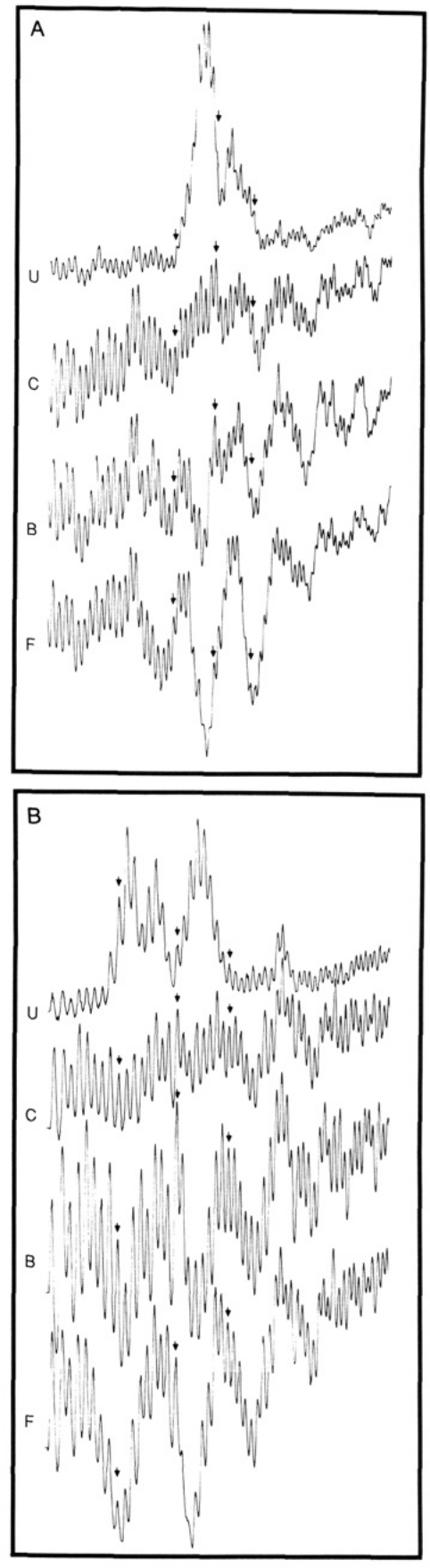

FIGURE 4: Densitometer scans of the cro protein–O_R1 missing nucleoside experiment in which the top strand (A) or the bottom strand (B) of the O_R1 operator was radiolabeled. (A) Scans marked U (unbound), C (control), B (bound), and F (footprint) correspond to lanes 4, 1, 3, and 2, respectively, in Figure 3. (B) Scans U, C, B, and F correspond to lanes 10, 7, 9, and 8 in Figure 3. Arrows mark the first, last, and central base in the 17 bp O_R1 operator sequence.

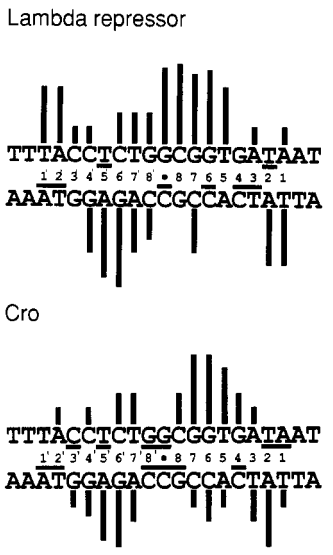

FIGURE 5: Compilation of the missing nucleoside data from λ repressor and cro protein plotted on the O_R1 operator sequence. The height of the bar approximates the relative strength of the contact signal for a particular nucleoside. The horizontal bars on the interior of the sequence indicate bases scored as not contacted by the missing nucleoside assay. Bases with no indication were indeterminate cases. The nonconsensus half of the operator is numbered 1′–8′, and the consensus half 1–8. The base pair at the center of the operator is marked by a filled circle.

disruption of DNA secondary structure as a result of the gapped base (see below).

We do not understand why a small contacted signal is observed at A1 since the cocrystal structure indicates no base contacts at this site. A similar observation by Brunelle and Schleif (1987) leads us to believe the effect is perhaps due to a subtle change in the secondary structure of the helix caused by the apurinic site (their study) or a one-base gap (this study). This structural change might adversely affect precisely aligned outer phosphate contacts between A2 and T1 (also T1′ and A2′) which the crystal structure shows are much more extensive than in the interior of the complex (Jordan & Pabo, 1988). This explanation is also consistent with the assumption that most of the binding energy contributed by this base pair is a result of hydrophobic interactions with T1. The symmetry-related counterpart to this small signal (at A1′) is not observed, consistent with other results (Brunelle & Schleif, 1987).

Our results predict that position 2 is strongly contacted entirely on the adenine half of the base pair, while the thymine is not involved in any contacts. This is consistent with the crystal structure, which shows a complicated arrangement of interactions with A2 but none with T2 (Jordan & Pabo, 1988). Mutagenesis experiments show that any substitution at this position results in a large drop in binding energy, while substitution of uracil for thymine has no effect on binding. Brunelle and Schleif (1987) reported a similar signal for A2, but also a small contacted signal at T2 in contrast to our results.

Our results indicate that base pair 3 is not contacted, since only a weak signal at A3, and none at G3′ or at the complementary bases of these base pairs, is observed. Jordan and Pabo (1988) suggest that the 5-methyl group of T3 is involved in the formation of a hydrophobic pocket, but the mutational data imply that this is not a large effect (Sarai & Takeda, 1987). The substitution of U for T results in an increase in binding affinity, and no other mutation at this position causes an appreciable loss of binding energy for repressor (Sarai & Takeda, 1987). Methylation of G3′ at N7 (O_R1, lower strand) causes a slight increase in binding of repressor (Hochschild & Ptashne, 1986). This result suggests that some small hydrophobic interaction can be made and thus account for the small but detectable signal we observe at A3. In sum, however, the evidence in the literature suggests that base pair 3 is probably not involved in operator discrimination by repressor (Sarai & Takeda, 1987; Hochschild et al., 1986). We find that no strong contacts are made at this position by the protein.

Base position 4 is completely conserved in all six operator sequences. Brunelle and Schleif found that both G4 and G4′ gave a moderate contacted signal, while C4′ and C4 gave a small and uncontacted signal, respectively. However, despite these indications of structural symmetry, our missing nucleoside data suggest that base pair 4 is contacted differently in the consensus and nonconsensus halves of the operator. Specifically, no signal is detectable at G4; C4 gives a strong uncontacted signal; G4′ gives a moderate contacted signal, while the situation at C4′ is indeterminate from our results. The cocrystal structure of the repressor/O_L1 complex indeed shows a small asymmetry in base contacts at this position, but opposite from what our data would predict. The cocrystal structure shows that G4 makes a specific contact with Ser-45, while because of conformational differences in the nonconsensus half, G3′ is also shifted toward Ser-45 (Jordan & Pabo, 1988). This additional interaction might help compensate in the event of the loss of the G4′ interaction in the consensus half of the repressor–O_R1 complex in solution. In addition, the substitution of any other base pair at this position on the consensus half of O_R1 results in only a small decrease in binding affinity for repressor (Sarai & Takeda, 1987). Thus, the effect of a gap at his position might be expected to be difficult to detect in our missing nucleoside assay. However, regardless of these arguments, at worst two of the four possible base sites at position 4 (C4 and G4′) are predicted correctly by our data.

The signal at position 5 indicates symmetrical *contacts* about the dyad even though the *sequence* itself does not exhibit this symmetry in the O_R1 operator. In particular, a strong contact is indicated at T5 in the consensus half and at A5′ in the nonconsensus half of the operator. A strong noncontacted signal is indicated for the opposite base in these pairs at A5 and T5′. The nonconsensus half of the operator sequence used in this study (O_R1) is different from that in the cocrystal (O_L1, which is symmetrical around the dyad at base pair 5) at this position. The crystal structure shows that the methyl group of T5 participates in hydrophobic contacts with Gly-46 and the β carbon of Ser-45 (Jordan & Pabo, 1988). Thus, for the consensus half of O_R1, the missing nucleoside data are consistent with an important hydrophobic contact at T5 (and no contact with A5). This result also agrees with the results of Brunelle and Schleif (1987) at T5.

The analysis of the results for position 5 in the nonconsensus half of the operator is more complicated because of the difference in sequence between O_R1 and O_L1. In this case, comparison with the mutational data is indispensable. When the consensus sequence T·A is mutated to A·T, the loss in binding energy is large and equal to that which results when U is substituted for the consensus T5. Thus, an important hydrophobic interaction with the 5-methyl group of T5 is indicated. However, also, when G·C or C·G is substituted for T·A at position 5, the binding energy decreases even further than in the case of the A·T mutation. This suggests that G·C or C·G base pairs lack the ability to make either of *two* types of interactions with the protein. One type (hydrophobic with

T5) is observed in the cocrystal in which both halves of O_L1 contain the consensus T·A sequence. Another type occurs in the case of the A·T base pair in the nonconsensus half of O_R1 where a different or additional interaction is indicated by mutagenesis data (Sarai & Takeda, 1987) and detected in our missing nucleoside experiment. The missing nucleoside experiment rules out the trivial explanation that a T on either strand can make hydrophobic interactions at base pair 5′ since our results show that T5′ in the nonconsensus half gives a large uncontacted signal. The interaction must therefore be with A5′ and might be formed only in lieu of the T5 interaction observed in the consensus half of the operator.

On the basis of the mutagenesis data, the most energetic contact between O_R1 and λ repressor is at position 6 of the operator (Sarai & Takeda, 1987). This position is conserved in all operator half-sites except the nonconsensus half of O_L3. The cocrystal structure shows a complex network of interactions at this position with Lys-4 and Asn-55 of the repressor. However, these interactions occur entirely with the G half of the base pair. The results we obtain from the missing nucleoside experiment are completely consistent with this arrangement of contacts. Specifically, we see some of the strongest contacted signals on either strand at position 6, while strong uncontacted and small contacted signals are observed at C6 and C6′, respectively.

The missing nucleoside results for the central five base pairs of the operator will be discussed together with the structural unit of the λ repressor dimer that contacts them. The repressor has an N-terminal arm that wraps around the center of the operator and contacts the backside of the DNA in the major groove (Pabo et al., 1982). The three amino-terminal-most residues of repressor have been shown to be important for protection of G7, G8, and the dyad G from methylation, and for making position G8′ sensitive to methylation. Base position 8 has been proposed to be involved in operator discrimination by repressor (Sarai & Takeda, 1987; Hochschild et al., 1983), and the N-terminal arm has been suggested to be involved in this function (Hecht et al., 1983; Eliason et al., 1985). The supposition that the arms contact the central bases of the operator is borne out by the cocrystal structure (Jordan & Pabo, 1988). However, the electron density of the arms is not as clear as the rest of the protein, especially in the nonconsensus half of the complex, and the assignment of all the specific contacts is difficult (Jordan & Pabo, 1988).

We propose that the structure of the N-terminal arm is asymmetric with respect to each half of the dimer. Specifically, the missing nucleoside data show that the top strand of the operator in the consensus half-site is strongly contacted at positions G7, C8, and the dyad G. Meanwhile, the symmetry-related positions on the bottom strand show only moderate contacted signals at A7′ and C8′, while the dyad C is uncontacted. Also, position C7 (lower strand, consensus half) is strongly contacted, but position T7′ (upper strand, nonconsensus half) is only moderately contacted.

These results suggest that the N-terminal arm that extends from the monomer that contacts the consensus half of the operator binds much more tightly to the DNA than the nonconsensus arm. This proposal corresponds well with the cocrystal structure which shows that the arm that contacts the consensus half-site is the more ordered. The asymmetry in methylation protection at position 8 also implies that the consensus arm is more closely associated with the DNA (Pabo et al., 1982). In addition, these results, when taken with the fact that the central base of the dyad (G-dyad, upper strand) is strongly contacted while its pair (C-dyad) is uncontacted, might suggest that only one of the two arms of the dimer can occupy the floor of the major groove in the center of the operator at one time. The inherent lack of true 2-fold symmetry at the dyad base pair could mean that only one specific interaction at this position is available for the two arms of the dimer such that only one arm at a time can interact at this position. It is interesting that the removal of just the last three residues of the protein (Ser^1-Thr^2-Lys^3) exposes base G8 and the central dyad-G base to methylation (Pabo et al., 1982) and that the cocrystal shows a possible interaction between base pair 8 in the consensus half of the operator and Thr-2 (Jordan & Pabo, 1988). Perhaps another serine–guanine bond is formed (Ser^1–dyad-G) at the central base pair similar to that found at position 4.

Summary of the Missing Nucleoside Analysis of the λ Repressor–O_R1 Complex. The λ repressor–operator system used for this study has been extensively analyzed in the past, and a cocrystal structure has been recently published. Even so, the results obtained with the missing nucleoside method can still contribute to the understanding of the repressor–operator complex. Specifically: (1) the results suggest that hydrophobic interactions found at bases T1 and T1′ are important, while those at position 3 are not as significant; (2) while the contacts made by the N-terminal arms of the repressor are not easily discernible in the crystal structure, the missing nucleoside data show that these arms probably make extensive and energetically important contacts with positions 7 and 8 and the central dyad base; (3) only the guanine of the central dyad base pair is contacted, and this contact is probably made with the repressor monomer which contacts the consensus half of the operator; (4) the majority of the contacts made by the N-terminal arm are on the same strand as the contacted G at position 6 for that repressor monomer; and (5) the most energetically important contact in the nonconsensus half of the complex involves position 6′, while in the consensus half the contacts with positions 7 and 8 and the dyad are at least as important as the contact at position 6.

cro Protein–O_R1 Complex. As a second application of the missing nucleoside technique, we analyzed the cro protein–O_R1 complex. cro and λ repressor give almost identical DNase I and hydroxyl radical footprints when bound to the same DNA site (Tullius & Dombroski, 1986). However, these proteins are expected to make very different sets of specific DNA–protein contacts. Thus, despite the similarities in footprints, the cro and repressor complexes should yield different signals in the missing nucleoside assay.

A total mutational analysis of the cro protein–O_R1 complex was recently published by Takeda et al. (1989). They showed that most of the important base-specific contacts occur in the consensus half of the operator. In addition, Metzler and Lu (1989) have shown evidence for an asymmetric interaction of the cro dimer with the O_R3 operator binding site. These results correlate well with our missing nucleoside analysis, in which the strongest protein–DNA contact signals are in the consensus half of the operator (Figure 5). Thus, this complex, like the λ repressor complex (see above), shows many signs of asymmetry about the pseudodyad axis.

Takeda and co-workers model the cro complex as making contacts with the consensus half of O_R1 at positions T1, A2, and T3 on the bottom strand, and at T2, A3, G4, T5, G6, and G7 on the top strand (our numbering system). In general, we observe contact signals which correspond to each of these positions. Positions G7, G6, and T5 show the largest signals, in agreement with the mutational data. In fact, contacts at these positions on the top strand (T5, G6, G7) appear to be

much stronger than their symmetry-related counterparts (A5′, G6′, A7′) with respect to the other contacts in each strand (Figure 5). Thus, contacts at these three bases might be the basis for the different strengths of interaction of each half of the operator. An important contact is also predicted to occur at position 2 (Takeda et al., 1989). Nucleoside A2 gives a large signal in our assay while its partner, T2, appears to be completely uncontacted, fully consistent with the other studies (Takeda et al., 1989; Metzler & Lu, 1989).

Another interesting observation concerns the center of the operator site. cro protein lacks the amino-terminal arms of repressor which appear to interact with the center of the 17 base pair operator (see above). Consequently, Takeda et al. (1989) show that the central three base pairs of O_R1 contribute little to the stability of binding. Our analysis agrees with these results in that this central region is largely devoid of any contact signals on either strand (Figure 5). This is in contrast to the large contact signals that we observe in this region in the missing nucleoside analysis of λ repressor (Figure 5).

Conclusions

The analysis of the missing nucleoside experiment presented in this paper has shown that the method gives valid results when tested on an extensively studied system such as the λ repressor–operator complex. The method has advantages over similar missing contact methods (Brunelle & Schleif, 1987) in that the premodification of the DNA-binding site is easily accomplished in one step and is effective at all sites in a DNA fragment. Thus, only one sequencing gel lane is required for analysis of the binding site. The advantages over site-directed mutagenesis techniques are even greater, since only one missing nucleoside experiment is required to analyze the importance of all the bases in even a very large binding site. We are presently extending this analysis to such a system, the 45 base pair internal control region of the *Xenopus* 5S RNA gene.

Acknowledgments

We thank Gary Ackers, Dorothy Beckett, and Carl Pabo for providing samples of λ repressor and cro proteins and Carl Pabo for a critical reading of the manuscript.

Registry No. Hydroxyl radical, 3352-57-6.

References

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature (London) 326*, 888–891.

Brunelle, A., & Schleif, R. F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6673–6679.

Chalepakis, G., & Beato, M. (1989) *Nucleic Acids Res. 17*, 1783.

Eliason, J. L., Weiss, M. A., & Ptashne, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2339–2344.

Hecht, M. H., Nelson, H. C. M., & Sauer, R. T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2676–2680.

Hochschild, A., & Ptashne, M. (1986) *Cell 44*, 925–933.

Hochschild, A., Douhan, J., III, & Ptashne, M. (1986) *Cell 47*, 807–816.

Johnson, A. D. (1980) Dissertation, Harvard University, Boston, MA.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893–899.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1987) *Nature 326*, 886–888.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Green, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526–1541.

Metzler, W. J., & Lu, P. (1989) *J. Mol. Biol. 205*, 149–164.

Pabo, C. O., Krovatin, W., Jeffrey, A., & Sauer, R. T. (1982) *Nature 298*, 441–443.

Rimphanitchayakit, V., Hatfull, G. F., & Grindley, N. D. F. (1989) *Nucleic Acids Res. 17*, 1035–1050.

Sarai, A., & Takeda, Y. (1987) *UCLA Symp. Mol. Cell. Biol., New Ser. 69*, 57–64.

Siebenlist, U., & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 122–126.

Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 439–443.

Tullius, T. D., & Dombroski, B. A. (1985) *Science 230*, 679–681.

Tullius, T. D., & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5469–5473.

Wolffe, A. P. (1988) *EMBO J. 7*, 1071–1079.