152–157 Nucleic Acids Research, 1994, Vol. 22, No. 2

Missing-base and ethylation interference footprinting of P1 plasmid replication initiator

Peter P.Papp+ and Dhruba K.Chattoraj*
Laboratory of Biochemistry, National Cancer Institute, NIH, Bethesda, MD 20892, USA

Received November 4, 1993; Revised and Accepted December 2, 1993

ABSTRACT

RepA, the replication initiator protein of plasmid P1, binds to specific 19 bp sequences on the plasmid DNA. Earlier footprinting studies with dimethylsulfate identified the guanines that contact RepA through the major groove of DNA. In this study, base elimination was used to identify the contribution of all four bases to the binding reaction. Depurination and depyrimidation of any base in the neighborhood of the contacting guanines was found to decrease RepA binding. These results are consistent with the notion that RepA contacts bases of two consecutive major grooves on the same face of DNA. We also observed that depurination but not methylation of three guanines (G3, G8 and G9) affected binding. We identified the DNA phosphate groups (3 in the top strand, one of which mapped between G8 and G9, and 4 in the bottom strand, one of which was adjacent to C3) that strongly interfered with RepA binding upon ethylation. These results indicate that certain bases (e.g. G3, G8 and G9) may not contact RepA directly but contribute to base and backbone contacts by maintaining proper structure of the binding site.

INTRODUCTION

The *repA* gene of P1 plasmid encodes the plasmid specific initiator protein, RepA (1,2). The gene is flanked by two sets of repeats of nearly identical 19 bp sequences, called iterons, which are the binding sites of the RepA protein (3). The upstream set of five iterons (# 10 to # 14), called *incC*, is a part of the replication origin of the plasmid and the downstream set of nine iterons (# 1 to # 9) constitutes the copy number control locus, *incA*. RepA binding to *incC* is essential for the initiation of plasmid replication (4). Since *prepA*, the promoter of the *repA* gene resides within *incC*, binding of RepA to *incC* also represses *prepA* and autoregulates its synthesis (4). RepA binding to *incA* is responsible for the low copy number of the plasmid (5). We have been studying the binding reaction *in vitro* with a goal to understand how RepA accomplishes its multiple critical functions.

Previous footprinting experiments, by methylation interference and methylation protection, indicated that the guanines in positions 1, 11, and 12 contact RepA through two consecutive major grooves on the same face of DNA. Hydroxyl radical protection experiments also confirmed that RepA primarily binds to one face of DNA (6).

In this study, we wanted to identify bases other than the guanines that could contact RepA. For this purpose 'missing base' footprinting experiments were attempted (7). In these experiments, any of the four bases can be singly eliminated from DNA by chemical treatments and the effect of the missing base on the binding reaction can be tested. In addition to base-contacts, we also tested for phosphate contacts by ethylation interference experiments since phosphate contacts usually contribute significantly to the binding energy (8–10). Both base elimination and phosphate ethylation from several positions significantly interfered with RepA binding. The results of these additional footprinting experiments were consistent with the notion that RepA contacts bases in the major grooves and the regions of backbone that flank the major grooves on the same face of DNA. Moreover, several guanine residues were identified where depurination but not methylation decreased RepA binding. These residues appear not to contact RepA directly but their presence may be required for proper structure of the site so that other base and backbone contacts can be made. Directly or indirectly, the degree to which individual base pairs contribute to RepA binding appeared to be correlated with the degree of base conservation among natural RepA binding sites (1,6). These results imply that RepA is the only protein that interacts with the iterons.

MATERIALS AND METHODS

Preparation of labeled DNA fragment

A 58 bp *Eco*RI–*Hin*dIII fragment of plasmid pSD14 was used throughout (11). This fragment carried iteron # 14 whose sequence is the consensus for all 14 iterons of miniP1 (1). pSD14 DNA was prepared from the DH5α strain of *E.coli* using a Qiagen plasmid kit and purified by Cs-banding. After digesting the plasmid DNA either with *Eco*RI or *Hin*dIII (New England BioLabs), end filling reactions were performed by Klenow polymerase in the presence of [α-^{32}P]dATP. The DNA was purified by ethanol precipitation and a second digestion was done

* To whom correspondence should be addressed

+ Present address: Department of Biology, Carlton University, Ottawa, Ontario K1S 5B6, Canada

with either enzymes *Hin*dIII or *Eco*RI. The fragments were separated from the rest of the plasmid DNAs by electrophoresis on a 5% polyacrylamide gel and recovered from the gel by the crush and soak method (12).

RepA–DNA binding reactions

The binding reactions with end-labeled DNA and RepA were performed at room temperature for 20 minutes in a buffer consisting of 20 mM Tris.HCl, pH 8, 40 mM KCl, 60 mM NaCl and 0.1 mM EDTA. The purity of the RepA protein used in the RepA–DNA binding reactions was ~50%. The specific binding of this partially purified RepA did not differ from the binding of highly purified (purity >99%) preparation of RepA protein as determined by DNase I and dimethylsulfate (DMS) protection footprinting techniques (data not shown). Unlike the highly purified protein, where addition of ATP and heat shock proteins DnaJ and DnaK (11) was obligatory for specific binding in our binding buffer, no such addition was necessary with the partially pure RepA preparation.

Missing base interference footprinting

RepA-binding interference experiments using depurinated and depyrimidated DNAs were done following Brunelle and Schleif (7). DNA was modified by the G, G+A and C+T reactions as described (Dupont-NEN sequencing kit). All reactions contained 10 ng end-labeled fragment and 4 μg calf thymus DNA.

For the G reaction, the DNAs were in 200 μl water to which 5 μl 20% DMS was added and the reactions were stopped after 7 minutes. For the G+A reaction, the DNAs were in 24 μl water to which 4 μl piperidine formate was added. The reactions were stopped by adding 240 μl of hydrazine stop buffer. For the C+T reaction, the DNAs were in 24 μl water to which 30 μl hydrazine was added and the reactions were stopped after 15 minutes by the addition of 200 μl hydrazine stop buffer.

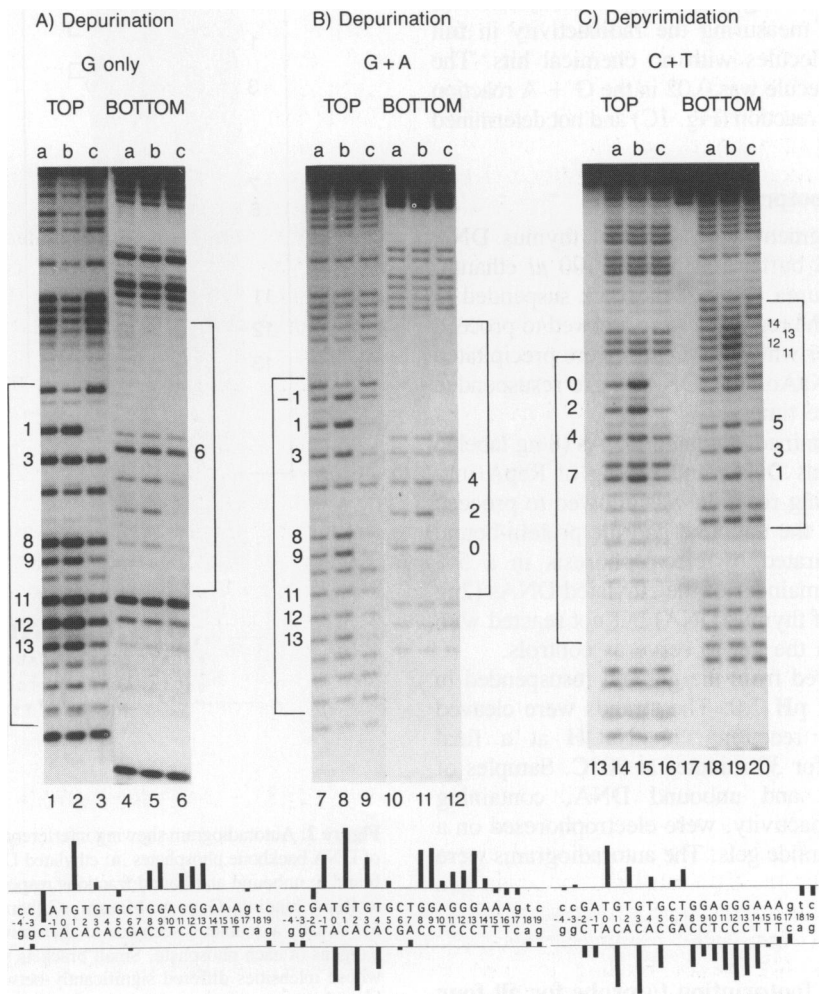

Figure 1. Autoradiograms showing interference of RepA binding due to depurination and depyrimidation for both top and bottom strands of a fragment carrying a single iteron. Iteron sequences are shown by brackets. **a**: chemically modified DNA not subjected to RepA binding; **b** and **c**: unbound and bound fractions respectively of modified DNA subjected to RepA binding. In lanes 13 and 17, the chemically modified DNA has not been subjected to piperidine cleavage. Coordinates of the bands representing significant interference are shown in the figure except in lane 6 where the single G of the bottom strand of the iteron is marked for reference purposes. Relative strengths of interference are shown by histograms below the autoradiograms. The height and sign of the bars were deduced from band intensities in lanes (b−c)/a. Positive values are shown above the line for the top strand and below the line for the bottom strand. The range of heights for the top and bottom strands were −0.78 to 1.18 and −0.10 to 0.04 respectively in (A), −0.23 to 1.33 and −0.15 to 0.99 in (B), and −0.035 to 0.93 and −0.21 to 1.12 in (C). The scale can be seen in Fig. 4A.

The binding reactions contained the modified DNAs (8 ng labeled + 3.2 μg calf thymus) and 2.5 μg of the RepA protein in a volume of 60 μl. The total RepA concentration was about 600 nM, and $K_{D,\ apparent}$ for specific DNA binding was ≅20 nM. The initiator concentration was such that it would have saturated binding if the DNA were not chemically modified. RepA and DNA were allowed to bind for 20 minutes after which the unbound and protein-bound DNA fractions were separated by electrophoresis in a 5% polyacrylamide gel. As controls, the modified DNAs (2 ng labeled + 0.8 μg calf thymus) but without reactions with RepA were also loaded on the gel. The DNAs were recovered from the gel and were subjected to piperidine cleavage. Equivalent amounts of radioactivity from the three samples (control DNA not reacted with RepA, and the bound and unbound DNAs reacted with RepA), were electrophoresed in a 11% denaturing polyacrylamide gels. The video images of autoradiograms were quantified using the software package NIH Image 1.41 developed by W.Rasband (NIMH). The radioactivity in the dried gels were also quantified directly using Fujix BAS2000 image analyzer. The high sensitivity of the system was particularly convenient in measuring the radioactivity in full length fragments, i.e. molecules with no chemical hits. The average number of hits/molecule was 0.02 in the G + A reaction (Fig. 1B), 0.2 in the C + T reaction (Fig. 1C) and not determined for the G reaction.

Ethylation interference footprinting

To 10 ng end-labeled fragment and 4 μg calf thymus DNA suspended in 100 μl DMS buffer (NEN kit), 100 μl ethanol, saturated with ethylnitrosourea (1 g Sigma stock suspended in 5 ml ethanol) was added. The reactions were allowed to proceed for 50 minutes at 50°C after which the DNAs were precipitated in the presence of 0.3 M NaAc. The DNAs were resuspended and precipitated three more times.

The binding reaction contained ethylated DNAs (8 ng labeled DNA + 3.2 μg calf thymus DNA) and 2.5 μg of RepA in a volume of 60 μl. The binding reaction was allowed to proceed for 20 minutes after which the unbound and the protein-bound DNA fractions were separated by electrophoresis in a 5% polyacrylamide gel. The remainder of the ethylated DNAs (2ng labeled DNA + 0.8 μg calf thymus DNA) but not reacted with RepA were also loaded on the gel to serve as controls.

The DNAs were recovered from the gel and resuspended in 10 mM sodium phosphate, pH 7.0. The strands were cleaved at modified positions by reaction with NaOH at a final concentration of 150 mM for 30 minutes at 90°C. Samples of the control, RepA-bound and unbound DNA, containing equivalent amounts of radioactivity, were electrophoresed on a 11 % denaturing polyacrylamide gels. The autoradiograms were quantified as before.

RESULTS

Missing-base interference footprinting to probe for all four base contacts with RepA

An end-labeled 58 bp DNA fragment carrying a single RepA binding site was used to bind RepA. The bases in the site have been given the coordinates −2 to +16 (6). The fragments were chemically modified by G specific depurination, G+A depurination and C+T depyrimidation reactions as described in the Materials and Methods. The G specific depurination experiments (Fig. 1A) showed that from the top strand the elimination of guanines from positions G1, G3, G8, G9, G11 to 13 interfered with RepA binding. In previous experiments, methylation of G1 and G11 to G13 interfered with RepA binding (6). In addition to these four positions, RepA also protected G9 from methylation upon RepA binding (6). Most likely, these five guanines are in close proximity to RepA in RepA−iteron complexes. In contrast, G3 and G8 might not contact RepA directly, but their elimination could have interfered with RepA

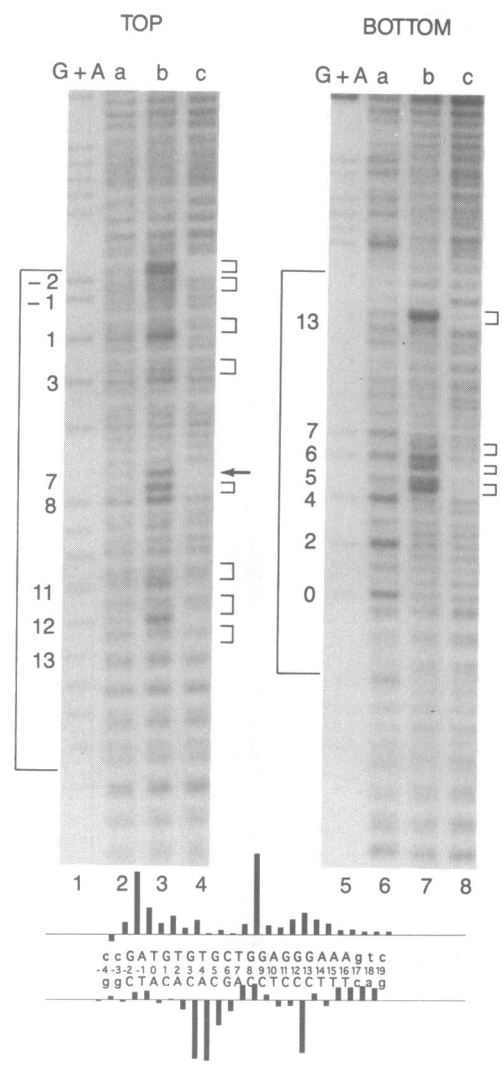

Figure 2. Autoradiogram showing interference of RepA binding due to ethylation of DNA backbone phosphates. **a**: ethylated DNA not subjected to RepA binding, **b** and **c**: unbound and bound fractions respectively of ethylated DNA subjected to RepA binding. G + A: Maxam−Gilbert sequencing products (lanes 1 and 5). Usually a doublet is seen due to alternative ethylation possibilities at the two oxygens of each phosphate. Small brackets show the doublets (resolved or not) whose intensities differed significantly between lanes b and c, and they were identified by coordinates of the corresponding base-specific cleavage bands running about 0.5 position faster in lanes 1 and 5. Other symbols are as in Fig. 1. Seven phosphates that interfered most strongly, namely the ones 3′ to bases at −2, −1 and 8 in the top strand and, 4, 5, 6 and 13 in the bottom strand, are shown in Fig. 3. In addition to the doublet, often a third band is seen at the position of Maxam−Gilbert sequencing markers e.g. positions 12, 11, 8, 7 (arrow), 3, 1 etc., in the top strand and positions 0, 2, 4 to 7 etc in the bottom strand. These bands resulted from base rather than backbone ethylation and were not included in quantification of interference (see text for details). The histograms were drawn as in Fig. 1 and the range of heights for the top and bottom strands were −0.061 to 1.59 and −0.32 to 1.19 respectively.

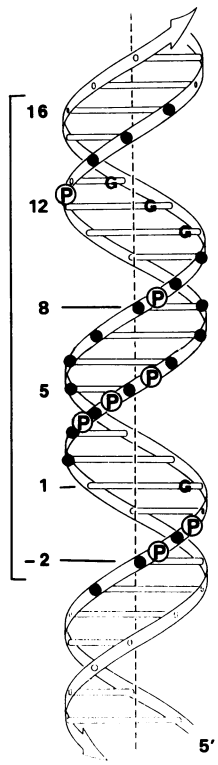

Figure 3. Map of base and backbone contacts of RepA on B-form DNA deduced from methylation (G) and ethylation interference (P), and hydroxyl radical protection studies (filled circles).

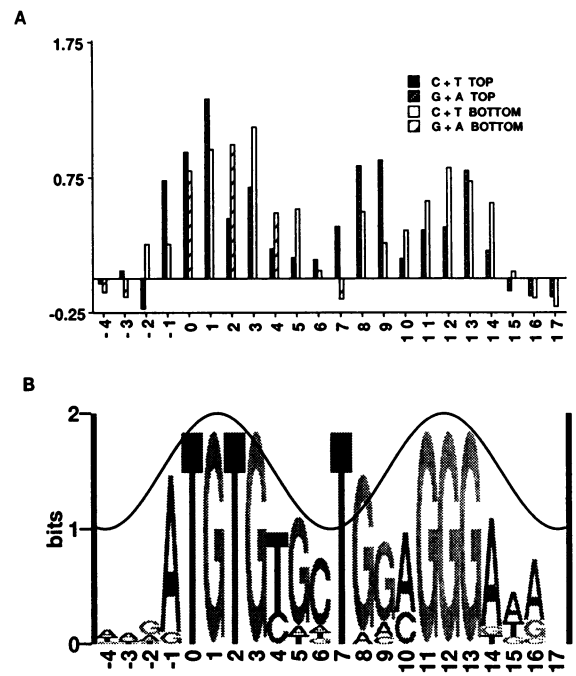

Figure 4. Comparison of the effect of base elimination on DNA binding by RepA (A) and the degree of base conservation among 14 natural RepA binding sites of P1 (B). The histogram in (A) shows the data combined from Figs. 1B and 1C except that positive values for both the top and bottom strands are plotted above the base line. The bar heights represent relative degree of interference to RepA binding upon base elimination. (B) is modified from Fig. 5a of reference 6 where bases are shown in a stack with heights in proportion to their relative frequencies. The height of each stack is the sequence conservation at that position in bits. A completely conserved base appears as a letter 2 bits high. The cosine wave represents one face of B-form DNA, the crests representing the major grooves. The correspondence between the bar heights (A) and conservation in bits (B) is weak only at positions 7, 9, 15 and 16. See text for possible reasons.

binding by altering the structure of the iteron. From the bottom strand of the iteron, elimination of the sole guanine, G6, did not cause interference with the binding reaction.

In the G+A depurination experiments, in addition to the seven Gs identified in Fig. 1A, elimination of adenines from position −1 in the top strand and from positions 0, 2 or 4 in the bottom strand interfered with binding (Fig. 1B). These experiments also confirmed that elimination of G3 can be significantly detrimental to binding (lanes 8,9).

C or T depyrimidation from any position between −2 to 14 significantly interfered with RepA binding although the magnitude of the interference varied across the site (Fig. 1C).

In summary, it appears that except for positions 15 and 16 at the end of the iteron, bases at all other positions contribute significantly to the binding energy. Either they contact RepA directly or their presence is required to orient the DNA or the protein so that proper contacts can be made elsewhere in the site.

Ethylation interference footprinting to identify phosphate contacts

In these experiments, end-labeled DNA fragments carrying on an average of one ethylated phosphate per fragment were used to bind RepA. The positions of the interfering ethylated phosphates were localized as described by others (8, 10; Fig. 2). Since the DNA fragments used were labeled at the 3′ ends, the bands whose intensity decreased in the RepA-bound fraction correspond to the bases (in the Maxam−Gilbert sequencing lanes) whose 3′ phosphate ethylation interfered with RepA binding. As pointed out by Rimphanitchayakit *et al.* (10), ethyl nitrosourea treatment, in addition to phosphate ethylation, could also modify DNA bases. As a result, up to three bands could be seen in ethylated DNA for each band of the lane representing the products of Maxam and Gilbert sequencing reactions (e.g. position 12, lanes 1 and 2 of top strand). The fastest moving species of the triplet migrated similarly to the products of sequencing reactions and, therefore, assumed to end with a 3′ phosphate. In other words, these bands were generated due to base modifications and, therefore, were not included in our analysis. If our interpretation of these bands is correct, then base ethylation also interfered with binding at positions 12, 7 and 1 in the top strand and position 5 in the bottom strand. The other two slower migrating species were assumed to represent the products of ethylated phosphates and ended with either 3′-OH or 3′ ethylated phosphates. These two species often did not resolve and, therefore, the combined intensity of the doublets were used to determine the strengths of interference. The doublets representing positions of phosphate ethylation interference have been marked with brackets (Fig. 2). These phosphates were 3′ to the bases at positions −2, −1, 1, 3, 8, 11, 12 and 13 in the top strand and 4, 5, 6 and 13 in the bottom strand. The interference from positions 1, 3, 11, 12, and 13 of the top strand was considerably weaker than that from other positions and may not represent phosphate contact as discussed below.

DISCUSSION

The interactions of RepA with one of its natural binding sites have been studied by various footprinting techniques in this paper. The same site was also studied in the past (6) and the results of the two studies are summarized in Fig. 3. From these results the following points emerge:

RepA binds to one face of B-DNA

Earlier results of methylation interference, methylation protection and hydroxyl radical footprinting experiments and the results of ethylation interference experiments in the present study are consistent with RepA binding to one face of B-form DNA. The phosphates whose ethylation interfered with RepA binding were principally those that were positioned next to the sugar residues that were protected from hydroxyl radical attack. But whereas all residues protected from hydroxyl radicals could be accommodated on one face of the DNA, ethylated phosphates 3′ to positions 1, 3, 11, 12 and 13 were on the face opposite to that of the residues protected from hydroxyl radicals. This observation does not necessarily indicate that the opposite face is also contacted by RepA. There are two oxygens in a phosphate group that can be ethylated and the oxygen at the 'inside edge' of the sugar−phosphate backbone when ethylated can protrude into the major groove significantly to interfere with base-specific contacts (9). With the exception of position 3, the other four positions were next to the G-residues that most likely contact RepA through the major groove (Fig. 1A). Also ethylation at all five of these positions interfered weakly with RepA binding cmpared to the ones that mapped on the RepA contacting face. Due to these considerations, we did not include these positions of weak interference as candidates for phosphate contact (Fig. 3), and believe that all the footprinting data can be interpreted to mean that RepA binds to one face of the DNA.

RepA contacts bases in the major grooves

This conclusion was inferred from earlier methylation interference and protection experiments (6). The results of both of these experiments indicated that G1 and G11 to G13 were likely to contact RepA through the major grooves. Depurination at these positions also severely affected binding (Fig. 1A). Depurination and depyrimidation experiments further showed that loss of bases from the contacting major grooves (positions −1 to 5 and 10 to 14) affect binding more profoundly than the bases from positions whose minor groove face RepA (Fig. 3). Notable exceptions were G8 and G9. The contribution of these two bases to binding appeared to depend on the probing technique. G8 appeared to be important for binding only upon depurination (Fig. 1A). Methylation of G9 did not interfere with binding but the residue was partially protected from methylation by RepA (6). These results indicate that if the base is contacted it must be at positions other than N_7 (e.g. a single hydrogen bond can be made at O_6 even if N_7 is methylated). Clearly, depurination of both G8 and G9 affected binding more severely than did methylation. Since G8 and G9 are located in a region of intimate backbone contacts, we favor the idea that the bases are required for proper orientation of the backbone rather than for direct contacts with RepA. In fact, G9 could be changed to any of the other three bases without affecting binding as judged by the band shift assay (6). These considerations make us believe that base-specific contacts are unlikely at G9.

Relationship of base conservation and binding energy

From the footprinting experiments of this study, a reasonable correspondence between the information content (degree of conservation) of 14 natural RepA binding sites and the magnitude of interference upon base elimination was apparent (Fig. 4). For example, the positions of lowest conservation (−2, 6, 15 and 16) were also the ones where base elimination had the least effect. The reason for low but significant conservation at these positions may be unrelated to the binding of RepA to a single iteron as studied here. RepA most likely forms a higher order structure by binding to its multiple sites (13). By DMS footprinting *in vivo*, hypersensitive sites between the iterons that included the terminal A residues (positions 14 to 16) have been seen (3). The converse correlation was also true: the highly conserved positions also affected binding most profoundly. These results suggest that the base conservation is there for iteron specific binding of RepA and for the formation of higher order structure of RepA−iteron complexes. Whether some conservation is there for binding of a second factor, as we have proposed earlier on theoretical grounds (6), remains to be determined.

Comparison of RepA−iteron and λ repressor−operator complexes

The λ repressor-operator contacts have been studied by various chemical probes as well as by high resolution X-ray crystallography (9). These studies predicted nine base-specific contacts, eight of which were revealed by the missing base contact technique (7). However, base elimination from one of these eight positions, G4′, only weakly affected binding although methylation of the base strongly interfered with binding and the base made a H-bond with Ser-45 in the crystal. One possible explanation for this discrepancy could be that the band shift assay used to select the bound fraction protected the weak complexes in which G4′ had been eliminated. The discrepancy at the ninth position (G8) also can be explained similarly. G8 was contacted by the N-terminal arm of the repressor in the crystal but its elimination only marginally affected binding. In the P1 iteron, a T to A change at position 7 showed a less than two-fold increase in K_D when binding was monitored by the band shift assay but showed about a nine-fold increase in k_{off}, the dissociation rate constant (6; data not shown). This is most likely due to the fact even weak complexes become stable in the low ionic strength of the gel buffer (because of reduction in k_{off} values) and, therefore, small affinity differences between wild type and mutant sites are not scored effectively in band shift assays.

For λ repressor, bases not previously predicted to contact the repressor were seen to reduce binding upon their elimination. It remains to be determined whether the integrity of the bases at these positions is important for the structure of the site or whether these bases are indeed involved in base specific contacts. In the case of P1, similar ambiguities remain for positions G3, G8 and G9. We favor the idea that these bases contribute to binding indirectly, as it is known that non-contacting bases are critically important for binding (6, 14, 15). Missing base contact probing can be a powerful means of identifying both base specific contacts and the bases that contribute to binding indirectly.

ACKNOWLEDGEMENTS

We thank Deepak Bastia for recommending missing-base footprinting, Gauranga Mukhopadhyay for RepA protein, Alokes Mazumdar for helpful suggestions on ethylation, and Tom

Schneider, Justin Dibbens and Michael Yarmolinsky for critical reading of the manuscript.

REFERENCES

1. Abeles, A.L., Snyder, K.M. and Chattoraj, D.K. (1984) J. Mol. Biol. 173, 307–324.
2. Wickner, S. and Chattoraj, D.K. (1987) Proc. Nat. Acad. Sci. USA 84, 3668–3672.
3. Sozhamannan, S. and Chattoraj, D.K. (1993) J. Bacteriol. 175, 3546–3555.
4. Chattoraj, D.K., Snyder, K.M. and Abeles, A.L. (1985) Proc. Nat. Acad. Sci. USA 82, 2588–2592.
5. Pal, S.K., Mason, R.J. and Chattoraj, D.K. (1986) J. Mol. Biol. 192, 275–285.
6. Papp, P.P., Chattoraj, D.K. and Schneider, T.D. (1993) J. Mol. Biol. 233, 219–230.
7. Brunelle, A. and Schlief, R.F. (1987) Proc. Nat. Acad. Sci. USA 84, 6673–6676.
8. Mazumdar, A. and Adhya, S. (1989) J. Mol. Biol. 208, 217–223.
9. Jordan, S.R. and Pabo, C.O. (1988) Science 242, 893–899. 10. Rimphanitchayakit, V., Hatfull, G.F. and Grindley, N.D.F. (1989) Nucleic Acids Res. 17, 1035–1050.
11. DasGupta, S., Mukhopadhyay, G., Papp, P.P., Lewis, M.S. and Chattoraj, D.K. (1993) J. Mol. Biol. 232, 23–34.
12. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. USA 78, 560–564.
13. Mukhopadhyay, G. and Chattoraj, D.K. (1993) J. Mol. Biol. 231, 19–28.
14. Koudelka, G.B. and Carlson, P. (1991) Nature 355, 89–91.
15. Somers, W.S. and Phillips, S.E.V. (1992) Nature 359, 387- 393.