Volume 5 Number 9 September 1978 Nucleic Acids Research

# DNAase footprinting: a simple method for the detection of protein-DNA binding specificity

David J.Galas and Albert Schmitz

Département de Biologie Moléculaire, Université de Genève, Genève, Switzerland

Received 6 July 1978

ABSTRACT

A method for studying the sequence-specific binding of proteins to DNA is described. The technique is a simple conjoining of the Maxam-Gilbert DNA-sequencing method and the technique of DNAase-protected fragment isolation. Fragments of a 5' end-labelled, double-stranded DNA segment, partially degraded by DNAase in the presence and absence of the binding protein, are visualized by eletrophoresis and autoradiography alongside the base-specific reaction products of the Maxam-Gilbert sequencing method. It is then possible to see the protective "footprint" of the binding protein on the DNA sequence. The binding of lac repressor to lac operator is visualized by "footprinting" as an example. Equilibrium estimates indicate that 10-fold sequence-specificity (differential binding constant) could be studied easily using this technique.

## INTRODUCTION

The specificity of DNA-protein interaction, as a fundamental recognition process in molecular biology, has been the object of several experimental methods designed to characterize this specificity (1-6). Proof of the specificity of a DNA-protein binding reaction has ultimately rested with the isolation and characterization of a fragment of DNA which is protected from DNAase degradation by the DNA-binding protein (2,7,8,9). In the course of our study of the binding specificity of lac repressor protein, we have devised a modification of this technique which is much simpler and which has potential applications to a wider spectrum of protein-DNA binding studies, some of which would be very difficult to carry out by isolating a protected DNA fragment. In this paper we describe this method and, with an eye to its wider applications, discuss its advantages and limitations.

 

The protected fragment isolation procedure (8) involves the digestion of a homogeneous sample of DNA by a DNAase in the presence of the binding protein being studied. The remaining intact piece of DNA is isolated chromatographically and then radioactively labelled and sequenced. In addition to the several steps involved, this method has the potential disadvantage that the yield of protected fragment may be very small, if the binding constant for a specific sequence is not much greater than the binding constant for the rest of the DNA. The method proposed here avoids these additional steps and, most notably, permits the detection of specificity ultimately limited only by the ability to detect differences in intensity on autoradiographs. By comparing intensities in the same reaction and employing microdensitometry, the detectability limits of sequence specificity could be substantially reduced. The "footprinting" technique can also be used to detect multiple binding sites on the same DNA fragment, and potentially allow quantitation of differential binding among them. In this sense it is similar to the methylation modification technique in which protein mediated changes in a base methylation reaction with dimethyl sulfate are studied (1,5). It is different in that protection from DNAase action indicates physical blocking of the DNAase molecule. While the methylation modification also depends on the binding of the protein to a DNA sequence, changes in the methylation pattern of the purines occur by an unknown mechanism. DNAase protection is therefore inclusive of methylation protection; i.e. it is possible to have DNAase protection without methylation modification (see section D).

## MATERIALS AND METHODS

Lac Repressor: The lac repressor of a strain carrying the I12-X86 double mutation in the I gene of E.coli was purified as described in reference 13. This repressor possesses a highly increased affinity for the lac operator and non-operator DNA as compared to wildtype repressor (13).

DNA fragments: a) The 48 base-pair DNA fragment used in this study was isolated from a pMB9 plasmid derivative carrying E.coli lac operon DNA (pMC1, kindly provided by M. Calos). The fragment was isolated after restriction enzyme fractionation of

the plasmid DNA and elution from polyacrylamide gels (14). The sequence of this fragment is known (16). Radioactive labelling and sequence reactions followed the procedure of Maxam and Gilbert (14).

b) The isolation of the lac operator DNA containing fragment has been reported (13). The plasmid used was kindly provided by L. Johnsrud.

DNAase I digestion: DNAase I (Worthington, bovine pancreatic) was added to a concentration of 0.05-5.0 µg/ml to the DNA fragments in 100 µl of DNAase I buffer (10 mM cacodylate buffer (pH8), 10 mM $MgCl_2$, 5 mM $CaCl_2$, 0.1 mM dithiothreitol) at room temperature. The reaction was stopped after 30 sec by adding 25 µl of a 3 M ammonium acetate, 0.25 M EDTA solution, containing 0.15 mg/ml sonicated calf thymus DNA (Serva). The DNA in this sample was then concentrated and desalted by two ethanol precipitations and prepared for electrophoresis on a 20% polyacrylamide gel by resuspension in 0.1 M NaOH, 1 mM EDTA as described in reference 14.

Protection against DNAase I digestion: the radioactively labelled lac operator containing DNA fragment in DNAase I buffer with 0.08 M KCl was incubated for 15 min at room temperature with purified I12-X86 repressor at 5 µg/ml before being subjected to the DNAase I digestion (0.14 mg/ml) described above.

## RESULTS AND DISCUSSION

### A. DNAase Partial Degradation Products

It is known from previous studies that DNAase I cleavage involves some sequence specificity (10,11). For the present method to be applicable this specificity must be weak enough to permit cleavage between virtually all the DNA bases in the region of interest, so that inhibition at each of these sites may be detected. In spite of the known specificity of the DNAase reaction the partial degradation fragments, produced under the conditions described in the previous section, show that detectable cutting occurs between every base pair in the regions we examined. Two such reactions are shown in figures 1 and 2. The stopping of the DNAase reaction at different degrees of completion results in different distributions of fragment lengths. In

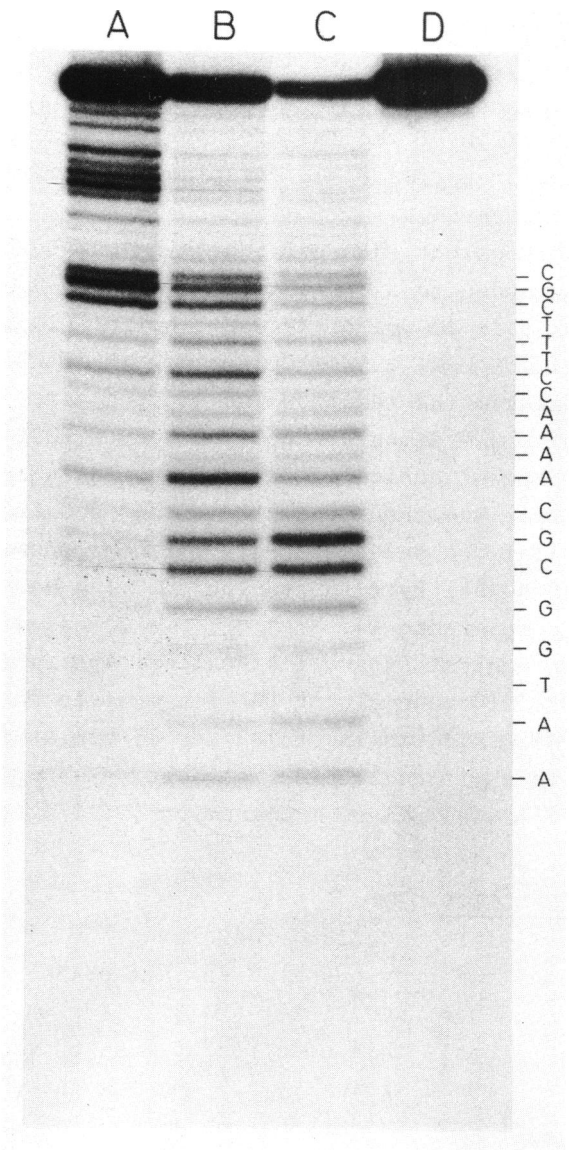


Figure 1 Autoradiograph of a partial DNAase I digestion of a 48 base pair, radioactively labelled DNA fragment electrophoresed on a 20% polyacrylamide gel (14). The DNAase I concentrations were: A, 0.05 µg/ml; B, 0.5 µg/ml; C, 5.0 µg/ml; D, stopping solution was added before the DNAase I (at 5.0 µg/ml). The sequence assignments were made (and confirmed) on the basis of parallel runs of sequencing reactions and DNAase I digestions. The sequence agrees with Calos (16).

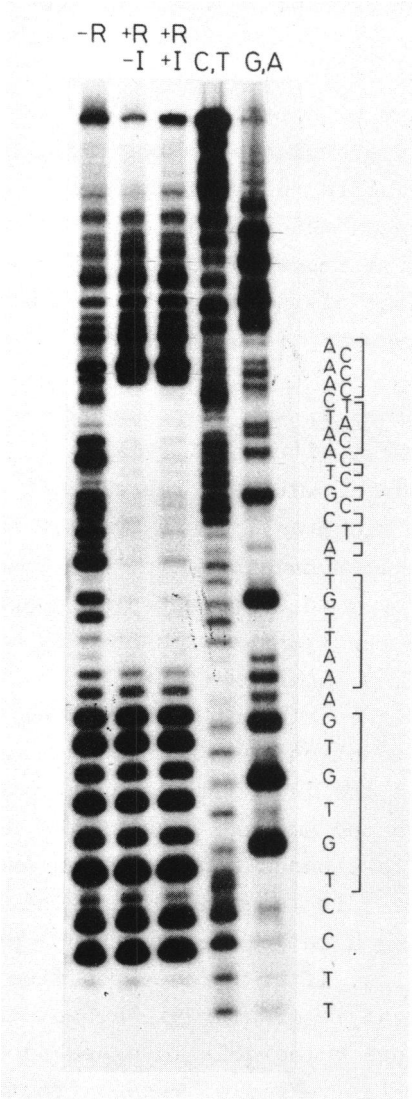


Figure 2 DNAase I footprint of the binding of I12-X86 repressor to lac operator DNA. The first three lanes (from the left) show partial DNAase I digests of a radioactively labelled lac-operator-containing DNA fragment for 30 seconds using 0.14 μg/ml DNAase. -R indicates the absence of repressor in the reaction mixture; +R, the presence of repressor (5 μg/ml); +I indicates that the inducer IPTG (isopropyl-B-D-thiogalacto-side) was present at 0.03 M. The lanes marked C,T and G,A represent sequencing reactions with hydrazine and dimethyl-sulfate respectively, as described in ref. 14.

figure 1 an example of a short (30 sec) reaction of DNAase I with a 48 base pair DNA fragment using different concentrations of DNAase is shown. It is immediately obvious that the fragment yield is not uniform with respect to length. The cutting apparently occurs preferrentially away from the labelled end of the DNA fragment (cutting near the unlabelled end is not clearly observable). This phenomenon leads to a change in intensity from the long to the short fragment region; otherwise, a mild reaction, which left most of the DNA intact, would yield an uniform distribution of fragment sizes, as in the Maxam-Gilbert sequencing method.

In figure 1, lane A, the reaction is shown to be incomplete by the fact, that most of the DNA remains in the uncleaved position. In this case, which is the closest of the three reactions to "one-hit" kinetics*, it is evident that the smaller fragments are much less numerous than the longer pieces. The smaller fragments seem to be produced more abundantly only in the more extensive reactions represented by lanes B and C.

The reaction set shown in figure 1 and in a densitometric form in figure 3, is consistent with the following picture of the size specific reaction. The nearer the base to the 5' end of the DNA fragment the less the reaction rate for cleavage at that base. Since the average size of the products of a complete reaction is 4 bases (10), such an end effect is not unexpected. In figure 3 it is evident that the fragment yield for all pieces requiring a cut closer to the 5' end than about 10 bases is quite low. After further reaction the shorter fragments are present in greater yield, but the longer fragments, which are more vulnerable to a second cut eventually suffer a loss in relative yield. These effects combine to produce a broad peak in yield which moves progressively toward the shorter fragments (see figure 3).

This interpretation of the data is certainly not proven by these experiments, but it does provide a reasonable working

* By "one-hit" kinetics we mean the condition in which those fragments produced have a high probability of coming from a single cut of the original piece of DNA.

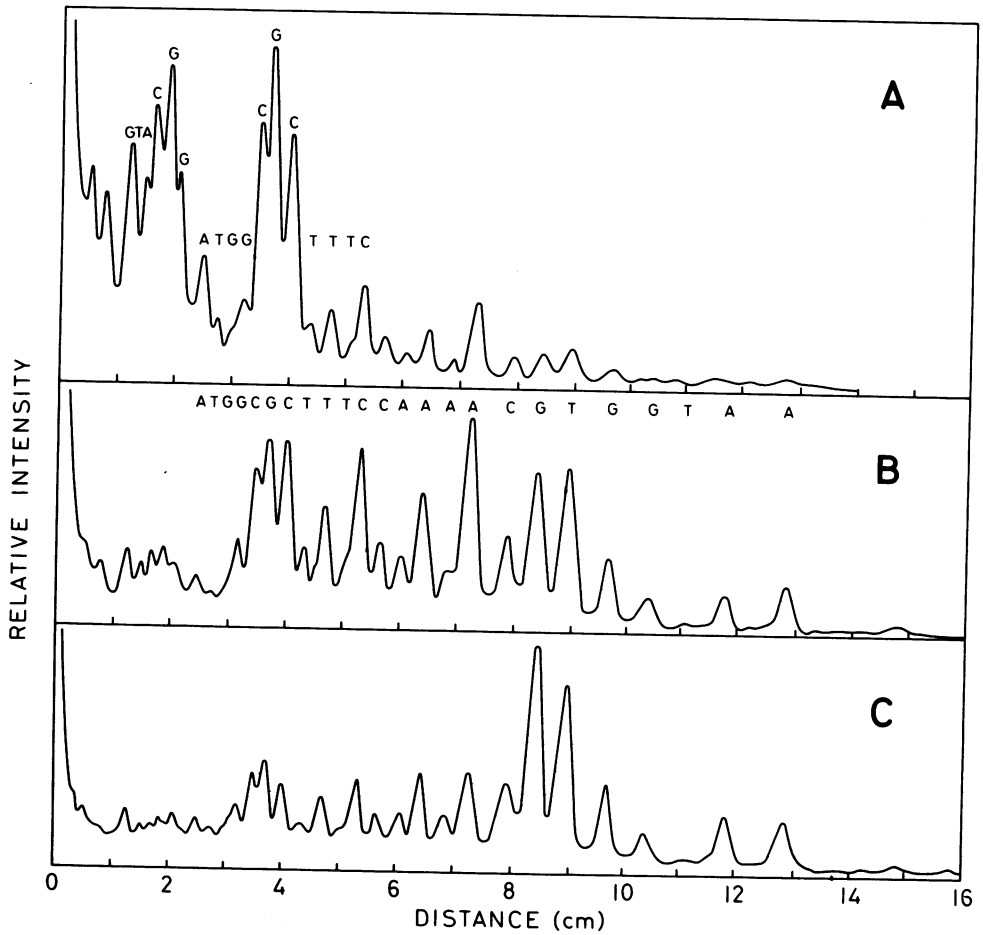


Figure 3. Microdensitometer scans of the gel in figure 1. The smoothed scans of lanes A, B and C are shown in panels A, B and C respectively.

model for the progression of the DNAase hydrolysis reaction. Our principal concern for the applicability of the method is that there are conditions under which the relative fragment yields are broadly distributed, and most of the bases are represented by detectable bands. This does appear to be the case. In practice, of course, such conditions would be determined empirically. Preliminary results indicate that DNAase II works equally well in partial digestion reactions.

B. Protection of Lac Operator by Lac Repressor

As an example of the use of this technique we have examined

the "footprint" of the repressor of the lac operon on the operator, an interaction originally determined by Gilbert and co-workers using protected fragment isolation and methylation protection (1,8). In figure 2 the footprint of a tight-binding mutant repressor (13) on the operator region is shown. This is a particularly simple case: the binding is strong and highly specific for only one site.

In examing the base-specific cleavage reactions alongside the DNAase degradations it is clear that the bands do not line up. To properly interpret the sequence then, the details of the cleavage reactions must be considered. The base-specific reactions (14) leading to the removal of the base and sugar moieties leave a phosphate terminated 3' end; while DNAase I cleaves the phosphodiester bond so that the 3' end is OH terminated. The mobilities of the fragments with equal numbers of bases will therefore differ according to how they are produced, and this must be considered in interpreting the autoradiographs. Two fragments of equal size, one produced by chemical removal of the n th base and one by enzymatic cleavage (DNAase I) to the 5' side of the n th base, will migrate in the gel such that the DNAase-produced fragment will be below the other fragment. This information, in concert with the sequencing reactions, permits the identification of the bases in the DNAase I reaction lanes.

In figure 2 the region protected from DNAase cleavage by the repressor is 25 bases long, and is essentially the same as the protected fragment isolated and sequenced by Gilbert and co-workers (15). The supression of the DNAase cleavage in the protected region is at least 70-fold (as determined by microdensitometry, data not shown). In this case, the use of a tight-binding mutant to enhance an already strong binding has permitted the use of low protein concentration. Since this is a rather special case, it is of great importance to the general applicability of the method to see how the binding affinities of the protein influence the yields of protected and unprotected fragments in the reaction.

C. Differential Fragment Yields Due to DNA Binding Proteins

In attempting to detect specific binding by this method there are two central issues to be considered: 1) the effect of

the binding protein concentration on the reaction rate for DNA hydrolysis at different sites, and 2) the detectability limits of different fragment yields. We would like to estimate what differential binding constants could be reliably distinguished, at reasonable concentrations of the reactants. To examine these relationships let us adopt a very simplified picture of the relevant reactions.

The binding of the protein to the i th site on the DNA* is assumed to go like:

$$(1) \qquad P + S_i \underset{k_{d,i}}{\overset{k_{a,i}}{\rightleftharpoons}} C_i$$

where P is the unbound protein, $S_i$ is the i th site on the DNA, and $C_i$ is a complex of protein and DNA in which the i th site is protected from hydrolysis by DNAase. The association constant, $k_{a,i}$, is most conveniently defined as the rate at which all DNA-protein complexes form, in which the i th site is protected. The dissociation of the same set of complexes is likewise described by $k_{d,i}$. These are composite quantities, in general. For a site within a specific binding sequence (like the lac operator), however, the contribution of this sequence to the composite may completely dominate. At the same time the DNAase reaction may proceed. This is described by the simplified scheme:

$$(2) \qquad D + S_i \xrightarrow{k_i} S_i' + D$$

(valid as long as the DNAase concentration is not near saturation) where D is the DNAase, and $S_i'$ is the i th site on the DNA which has been cleaved. The reaction constant $k_i$ depends on the site (DNAase specificity), but this variation can be accounted for as discussed previously. The important variables are the changes in the yields of cleaved sites, $S_i'$, on the binding of the protein.

* We will name the sites on the DNA by the convention that they are sites for DNAase cleavage. Thus they are single bases; whether they are covered or not covered by the protein is the crucial issue.

Further simplifications are possible when the constraints on these reactions are considered. The requirement that we get a broad spectrum of fragments, and that the reaction is reasonably close to one-hit kinetics, means that the DNAase reaction never proceeds very far. The number of sites, $S_i$, available for protein binding will therefore not change appreciably during the DNAase reaction. We can safely assume then that the binding is approximately in equilibrium, so that

$$(3) \qquad C_i = \frac{k_{a,i}}{k_{d,i}} P S_i ,$$

and the conservation conditions: total sites, $S_{it} = C_i + S_i$ and total protein $P_t = P + \sum_i C_i$. The rate of cleavage at each site in the DNA is assumed to be proportional to the concentration of the DNA unprotected at that site. Then the rate equation for the production of fragments is:

$$(4) \qquad \frac{dS'_i}{dt} = k_i D S_i$$

Since our primary interest is in detecting changes in the yield of fragments at a particular site, on the binding of the protein, the important variables are: the ration of the yield at one site in the presence of the protein to the yield in its absence, and the ratio of yield at a specific binding site to the yield at other sites, both in the presence of the protein. It is convenient for us to distinguish here between specific binding sites and the remainder of the DNA by the simplification of specifying only two binding constants: $K$ denotes the constant for a specific site and $\overline{K}$ the constant elsewhere. For our purposes we don't require an exact solution to the kinetic problem. If all reaction conditions are constant then the ratios of interest are approximated, using equation (4), as:

$$(5) \qquad \begin{aligned} &( S'_i )_{+\text{protein}} / ( S'_i )_{-\text{protein}} = ( S_i )_{+\text{protein}} / S_t \\ &( S'_i / S'_j )_{+\text{protein}} = ( S_i / S_j )_{+\text{protein}} \end{aligned}$$

Define this last ratio as R. In this approximation then, the DNAase reaction simply samples the equilibrium distribution of sites unobstructed by the binding protein.

To obtain an expression for the above ratios in terms of

the binding constants and protein concentration is straightforward. However, for most cases of interest further approximations apply which render the expressions much more transparent. First, the specific binding constant is expected to be much larger than the non-specific constant; $K \gg \overline{K}$. Next, the DNA segments in the reactions must necessarily be small enough (the fragment will usually be 50-100 base pairs long) so that the number of non-specific sites does not exceed the number of specific sites by a large factor, not more than 5-fold, for example. Under these conditions the absorption of protein by the non-specific sites will not change the available protein concentration by much if the total protein concentration is near saturation for the specific sites. One must check each case individually, of course, to acertain the validity of these considerations. The equilibrium concentration of free DNA sites, for specific and non-specific binding respectively, is then

$$(6) \qquad S_i = \frac{S_t}{1 + K P} , \quad S_j = \frac{S_t}{1 + \overline{K} P}$$

The value of the ratio R is therefore

$$(7) \qquad R = \frac{k_i ( 1 + \overline{K} P )}{k_j ( 1 + K P )}$$

While this expression is strictly valid only for equilibrium conditions and high protein concentrations (relative to DNA concentration), it provides a reasonable estimate for the yields expected from the DNAase degradation and therefore of the relative intensities of the bands in an autoradiograph.

In figure 4 equation 7 is plotted for several values of $\overline{K}/K$, assuming $k_i/k_j=1$. It is clear from eqn. (7) that the limit to the discrimination ratio R, achieved at high protein concentration, is $\overline{K}/K$. Suppose we can easily see a 5-fold difference in intensity. Taking the ratio of the binding constants to be 0.1, to consider a case of very weak specificity*, the curve in figure 4 gives us the required value of R

* The specificity of the lac repressor for the operator is of the order of $10^6$; i.e. $\overline{K}/K = 10^{-6}$.

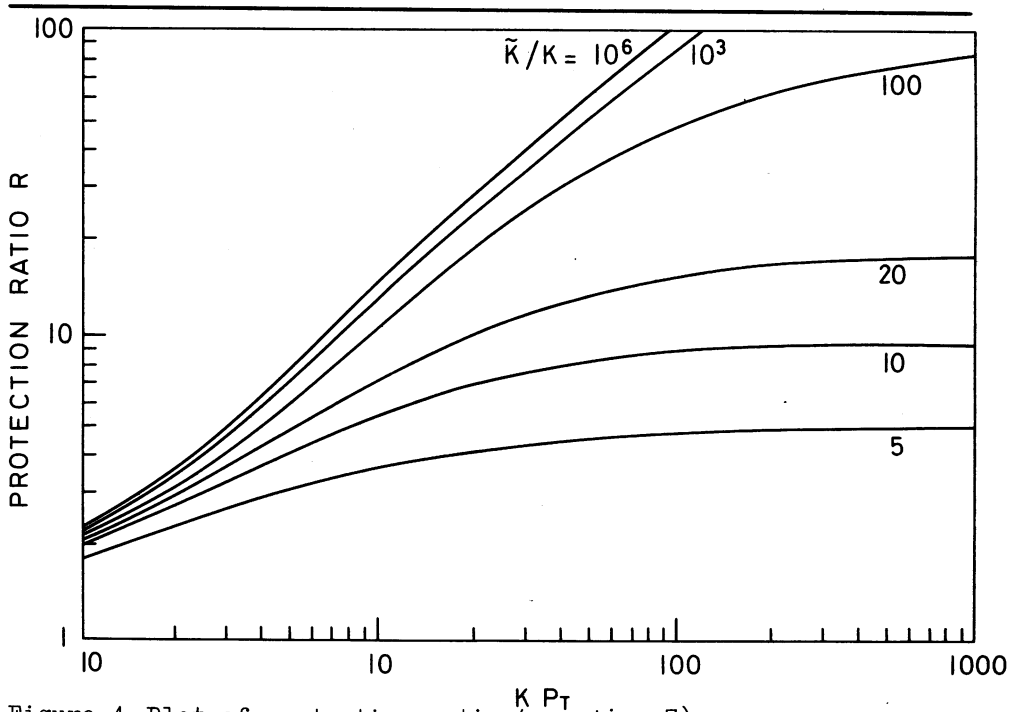


Figure 4 Plot of protection ratio (equation 7) versus the product $K \cdot P_t$ for several ratios of specific-site binding constant to non-specific constant, $\bar{K}/K$ (defined in text).

at KP=10. If $K = 10^6\ M^{-1}$, for example, the protein concentration required is $10^{-5}$M. At this concentration equation (6) predicts a reduction in the availability of sites to DNAase of a factor of 2, for non-specific sites, and a factor of 11 for specific sites. This case is considered to be a rather extreme example. Much lower values of $\bar{K}/K$ are expected for proteins involved in gene control, for instance.

D. Protection of a Non-Operator Sequence

We have detected sequence-specific binding of the I12-X86 repressor protein to non-operator DNA sequences using this technique. Figure 5 shows the DNAase footprint of the repressor on a 76 base pair fragment of non-operator DNA. This fragment, of known sequence, was not known to bind lac repressor. At the same repressor concentration as in figure 2 the DNAase cleavage protection is barely discernable (lane B). At ten times higher concentration than B the protection is quite obvious (lane C), but still much less than for the operator. The fragments

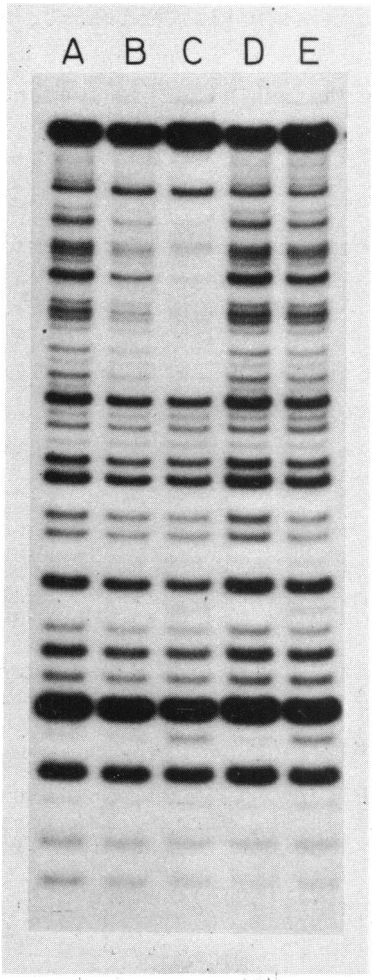


Figure 5 DNAase I footprint of the binding of I12-X86 repressor to a 76 base pair DNA fragment not containing lac operator DNA. Lane A: no repressor in the reaction mixture present; B and C: repressor 5 and 50 µg/ml in the reaction mixture; D and E: repressor 5 and 50 µg/ml in the presence of 0.035 M IPTG.

outside the protected region remain unchanged in intensity for both repressor concentrations. Note that the protection is completely reversed in the presence of IPTG (lanes D and E). Further details and a discussion of these results will be presented elsewhere (Schmitz and Galas, in preparation).

Although the protection from DNAase is quite marked under these conditions, no changes in the methylation pattern of this

fragment were seen, even at higher repressor concentrations (data not shown). This then is an example of protein DNA interaction in which the protein binding is sequence specific, as seen by DNAase footprinting, but no methylation modification results.

It is apparent from the theoretical estimates and the examples given here that the DNAase footprinting method should be applicable to a wide variety of DNA-binding proteins.

ACKNOWLEDGEMENTS

We would like to thank Dr. Jeffrey Miller for his encouragement and for useful discussions throughout the course of this work. This work was supported by a grant from the Swiss National Fund (F.N. 3.179.77).

REFERENCES

1. Gilbert,W., Maxam,A.M. and Mirzabekov,A.D. (1971) in Control of Ribosome Synthesis, Alfred Bension Symposium IX, Kjeldgaard,N. and Maaloe,O., Eds., pp. 139-148, Munksgaard, Copenhagen
2. Brown,K.D., Bennett,G., Lee,F., Schweingruber,M.E. and Yanofsky,C. (1978) J.Mol.Biol. 121, 153-177
3. Pribnow,D. (1975), Proc. Nat. Acad. Sci. USA 72, 784-788; and (1975) J.Mol.Biol. 99, 419-443
4. Gilbert, W. (1976) in RNA Polymerase, Losick,R. and Chamberlin,M., Eds., pp. 193-205, Cold Spring Harbor Laboratories, Cold Spring Harbor, New York
5. Mirzabekov,A.D. and Melnikova,A.F. (1974) Molec. Biol. Rep. 1, 379-384
6. Walz,A. and Pirrota,V. (1975) Nature 254, 118-121
7. Maniatis,T. and Ptashne,M. (1973) Nature 246, 133-137
8. Gilbert,W. (1972) in Polymerization in Biological Systems, pp. 245-259, Elsevier, North-Holland, Amsterdam
9. Johnson,A., Meyer,B.J. and Ptashne,M. (1978) Proc. Nat. Acad. Sci. USA 75, 1783-1787
10. Bernardi,A., Gaillard,C. and Bernardi,G. (1975) Eur. J. Biochem. 52, 451-457
11. Bernardi,A., Ehrlich, S.D. and Thiery,J. (1973) Nature 246, 36-40
12. Matsuda,M. and Ogoshi,H. (1966) J. Biochem. (Tokyo) 59, 230-237
13. Schmitz,A., Coulondre,C. and Miller, J.H. (1978) J.Mol.Biol. 123, in press
14. Maxam,A.M. and Gilbert,W. (1977) Proc. Nat. Acad. Sci. USA 74, 560-564
15. Maxam,A.M., Gilbert,W. (1973) Proc. Nat. Acad. Sci. USA 70, 3581-3584
16. Calos,M. (1978) Nature, in press.