Accessibility of DNA–epoxycellulose to sequence-specific DNA-binding proteins

Pedro A. LAZO

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, U.S.A.

(Received 26 September 1983/Accepted 9 November 1983)

Phage λ DNA was covalently coupled to epoxy-activated cellulose to form a stable DNA–cellulose matrix for affinity chromatography of sequence-specific DNA-binding proteins. The accessibility of three specific six-base sequences, GGATCC (*Bam*HI), GAATTC (*Eco*RI) and AAGCTT (*Hind*III) was studied quantitatively and qualitatively by restriction analysis followed by labelling of their recessed ends. All sites are randomly accessible. The site accessibility is variable, *Bam*HI > *Hind*III > *Eco*RI, and within the range 20–100% depending on base composition and internal structure of the sequence. DNA–epoxycellulose, because of its high efficiency of coupling, capacity, stability and accessibility, can be of great help in the isolation and characterization of sequence-specific DNA-binding proteins.

The coupling of DNA to solid supports is a widely used technique in molecular biology for the isolation of RNA. The immobilization of DNA has been done either through their terminal phosphate groups using carbodi-imides (Gilhan, 1971) or through their nucleotides by the use of diazotized aromatic amines (Noyes & Stark, 1975; Alwine *et al.*, 1979) or CNBr (Arndt-Jovin *et al.*, 1975; Siddell, 1978). These techniques have shown great variability in three key parameters, efficiency of the coupling reaction, capacity and accessibility or availability of DNA for RNA hybridization. They vary in the range of 10 to 80% for each of the three parameters, but not in the same way for an individual method (Bunemann, 1982; Bunemann *et al.*, 1982). Although some have been used for protein purification, no study has been performed regarding the accessibility of specific sequences.

Recently a new method for coupling DNA to a solid support has been reported (Moss *et al.*, 1981). It involves the activation of cellulose by the bifunctional oxirane-1,4-butanediol diglycidyl ether and subsequent linkage of DNA through its amino groups. This epoxy-activated cellulose has a higher efficiency of coupling and a higher capacity than the other procedures (Moss *et al.*, 1981). The method has already been successfully used for the purification of avian lipoprotein RNA (Moss *et al.*, 1981).

The state of DNA in the cellulose matrix might

Abbreviation used: bp, base pairs.

not be suitable for RNA or proteins. RNA accessibility requires long stretches of single strand DNA while protein accessibility needs short segments of double strand DNA (in some cases it might be single strand). So far nothing is known about the real possibilities and state of these techniques for the isolation of sequence-specific DNA-binding proteins. At first sight we might be doubtful of their usefulness since the coupling procedures might destroy the sequence by either base modification, strand separation or creation of steric hindrance.

In a model situation bacteriophage λ DNA was coupled to epoxycellulose and the protein accessibility to three short specific sequences was tested: GGATCC, GAATCC and AAGCTT that are recognized by *Bam*HI, *Eco*RI and *Hind*III restriction endonucleases, respectively. The results show that all of them are accessible in a high proportion and remained linked to the cellulose matrix. This method can be of great help in the identification, isolation and characterization of sequence-specific DNA-binding proteins. This is a field of ever-increasing importance, though still in its infancy, if we want to understand the regulation of many DNA functions in which specific proteins are certainly involved (Takeda *et al.*, 1983).

Materials and methods

Reagents and materials

Epoxy-activated cellulose, bacteriophage λ DNA, agarose and restriction endonucleases

*Bam*HI, *Eco*RI and *Hind*III were from Bethesda Research Laboratories (Gaithersburg, MD, U.S.A.). DNA polymerase I and its Klenow fragment were from Boehringer (Indianapolis, IN, U.S.A.). The radioactive nucleotides [α - 32 P]dATP and [α - 32 P]dGTP, both with a specific radioactivity of 795 mCi/ μ mol, were from New England Nuclear (Boston, MA, U.S.A.). Nitrocellulose filters (HAWP, 0.45 μ m pore size, 25 mm diameter) were from Millipore (Bedford, MA, U.S.A.). All other reagents were of analytical grade.

Coupling DNA to epoxycellulose

The coupling of 50 μ g of DNA to 50 mg of epoxycellulose was performed as recommended by the manufacturer (Bethesda Research Laboratories) for 1 or 8 h. Before use, DNA-epoxycellulose was washed eight times in the appropriate buffer for restriction analysis. DNA was determined by absorbance at 260 nm and distribution of tracer amounts of radiolabelled λ DNA.

Restriction analysis

DNA (2–5 μ g) coupled to epoxycellulose was suspended in 20 μ l of the restriction buffer and 1 unit of restriction enzyme/ μ g of DNA was added. Restriction buffers have the following composition: 100 mM-Tris/HCl (pH 7.2)/5 mM-MgCl₂/2 mM- β -mercaptoethanol/50 mM-NaCl for *Eco*RI; 20 mM-Tris/HCl (pH 7.4)/7 mM-MgCl₂/60 mM-NaCl for *Hind*III; and 20 mM-Tris/HCl (pH 7.0)/7 mM-MgCl₂/2 mM- β -mercaptoethanol/100 mM-NaCl for *Bam*HI. Digestion was performed overnight at 37°C. The digestion was stopped by washing three times with 150 μ l of 50 mM-NaCl/5 mM-EDTA/30 mM-Tris/HCl, pH 8.0. The supernatants were collected and ethanol-precipitated for further quantitative and qualitative analysis.

Labelling of DNA

Either 0.2 or 1 μ g of DNA coupled to epoxycellulose was resuspended in 20 μ l of 50 mM-NaCl/10 mM-MgCl₂/1 mM-dithiothreitol/10 mM-Tris/HCl (pH 7.4); 1 μ l of undiluted labelled nucleotide (dGTP for the *Bam*HI digest and dATP for the *Eco*RI and *Hind*III digests) and 0.5 units of Klenow fragment of DNA polymerase I were added for labelling of their recessed 3' ends. The reaction mixture was incubated for 1 h at room temperature after which it was extensively washed and passed through a nitrocellulose filter. The filters were dried and radioactivity was counted.

The end labelling of fragments released by digestion was performed in a similar way. The incubation was stopped by ethanol precipitation using 16 μ g of carrier tRNA. The precipitate was analysed by agarose-gel electrophoresis or sus-

pended in 0.5 ml of 50 mM-NaCl/10 mM-Tris/HCl, pH 7.4, and passed through Sephadex G-50 (fine grade) to separate the nucleotides; the peak with DNA was used for counting radioactivity.

Agarose-gel electrophoresis

Agarose gels (0.8%) in 40 mM-sodium acetate/2 mM-EDTA/20 mM-Tris/acetate (pH 8.0) were used to analyse the released fragments. DNA was transferred to a nitrocellulose filter (Southern, 1975) and autoradiographed.

Results

Coupling of λ DNA to epoxycellulose

The yield of the coupling reaction was in the range 82–93% of the input of DNA after 1 or 8 h. Since the coupling reaction conditions were developed for DNA with an average length of 700 bp (Moss *et al.*, 1981) and λ DNA has a size of 48502 bp (Sanger *et al.*, 1982), the coupling time was shortened to 1 h with no change in the yield. A similar yield was obtained using a plasmid 10 kbp long. This λ DNA-epoxycellulose was stable for at least 2 months without loss of coupled DNA.

Release of DNA fragments by restriction endonuclease digestion

To observe accessibility of specific sequences within phage λ coupled to epoxycellulose the release of fragments of DNA after digestion with restriction enzymes was performed. Three restriction enzymes were used, *Bam*HI, *Eco*RI and *Hind*III. This analysis will give us two types of information: (1) what individual sites were accessible, and (2) what fraction of DNA is free of attachment points to epoxycellulose outside the restriction sites. As can be seen in Fig. 1, all fragments but the largest one are released from the cellulose matrix. They have an intensity inversely related to size; this is particularly noticeable in the *Hind*III digest where a wider and smaller DNA size range is covered. This increasing intensity represents an increase in concentration, since the end labelling is size-independent, and equimolar fragments will give an equal intensity, as in the control digestion. A partially digested fragment with a size of 0.7 kbp, presumably derived from the 0.6 and 0.1 kbp fragments, was released following digestion with *Hind*III. The cohesive end termini are also free as the fragments carrying them, the 4.3 kbp band from the *Hind*III digest, and the smaller one in the *Bam*HI and *Eco*RI digests, are also released. The observations on released fragments tell us that a fraction of all restriction sites are accessible and form part of segments of DNA without any attachment point to epoxycellulose. The quantification of the released fragments (Table 1) tells us that they are a small fraction of

total DNA and variable depending on the sequence: 10% for *Bam*HI, 3% for *Hind*III and 1.6% for *Eco*RI. We must note that all restriction sites under study are concentrated in half the length of

phage λ , so the relative access per length unit is higher.

End accessibility of digested DNA coupled to epoxycellulose

To quantify the actual number of sites that have been digested, but not observed as released DNA because of the linkage through bases outside the restriction site, the remaining DNA bound to epoxycellulose was end-labelled. This end labelling was done on the assumption that if a site is accessible to a restriction enzyme, there should be no steric hindrance to another enzyme, the Klenow fragment of *E. coli* DNA polymerase I. The results are shown in Table 1 and are expressed as pmol of incorporation of labelled dNTP. For *Bam*HI digestion each site will incorporate 2mol of dGTP per mol of restriction site, and 10mol per mol of phage λ . In *Eco*RI digestion each site incorporates 4mol of dATP, that will make 20mol/mol of phage λ . In *Hind*III digestion each site incorporates 2mol of dATP that will result in 14mol of dATP incorporated per mol of phage λ . The results are also shown (Table 1) in relative terms, assuming that there is no base modification or that the base modification is 3% due to the coupling reaction (Bunemann *et al.*, 1982). A 3% base modification will knock out one site out of five for *Bam*HI and *Eco*RI and 2.5 sites out of 7 for *Hind*III. Clearly most of the sites are accessible to the restriction enzymes, particularly those of *Bam*HI.

Fig. 1. Fragments released from λ DNA (λ) and λ DNA–epoxycellulose (λ EC) by digestion with restriction endonucleases

The λ EC lanes contained 3 μ g of DNA for *Bam*HI and *Eco*RI digestion and 5 μ g for *Hind*III digestion. The λ lanes contained 180ng of DNA for *Bam*HI and *Eco*RI and 200ng for *Hind*III.

Discussion

The high yield of the coupling reaction, over 80%, confirms the original report by Moss *et al.* (1981) and is well above that of other methods (Bunemann *et al.*, 1982).

The observation that there is no difference in

Table 1. Determination of accessible sites by end labelling after digestion with restriction enzymes
n is the number of determinations. Values are means $\pm$ s.d.

Sequence	Total possible incorporation of dNTP (pmol/ μ g of DNA)		Observed incorporation of dNTP (pmol/ μ g of DNA)			Relative values observed/possible (%)		
	No base modification (A)	3% base modification (B)	Linked to epoxycellulose	Free	Total (C)	C/A	C/B	<i>n</i>
<i>Bam</i> HI GGATCC CCTAGG	0.3125	0.2344	0.2050 $\pm$ 0.0360	0.0314 $\pm$ 0.0015	0.2364	75	100	10
<i>Eco</i> RI GAATTC CTTAAG	0.6250	0.4687	0.0982 $\pm$ 0.0578	0.0100 $\pm$ 0.0011	0.1082	17	23	5
<i>Hind</i> III AAGCTT TTCGAA	0.4375	0.2800	0.0882 $\pm$ 0.0203	0.0132 $\pm$ 0.0003	0.1014	23	36	6

coupling for a wide range of DNA size, from 600bp (Moss *et al.*, 1981) to 48kbp (the present work), as well as no effect of shortening the reaction time by 80%, suggests that the coupling is essentially complete in a short time. This is important, because the potential denaturing effect of the coupling procedure might separate the strands. This seems to be the case where short fragments (600bp) are used, since 90% of the DNA is accessible to RNA (Moss *et al.*, 1981). In the case of a long molecule of DNA, due to its higher stability, the coupling seems to take place as a double strand as suggested by the number of sites accessible to restriction enzymes and by the identification of specific fragments released by restriction endonuclease digestions.

The analysis of restriction sites (Fig. 1 and Table 1) shows that all of them are accessible in a significant proportion. The loss of some sites is consistent with a 3% base inactivation or modification of nucleotides; a modified base will eliminate a restriction site. The base reactivity is not the same for all nucleotides and depends on the reaction volume. The relative reactivity of nucleotide homopolymers in 250 μ l is: 72 for poly(dT), 45 for poly(dA) and poly(dC) and 20 for poly(dG) on a scale of 100 (Moss *et al.*, 1981). The results in Table 1 using a completely different DNA seem to agree with this reactivity order. The sequence with higher GC content (GGATCC, *Bam*HI) is the most accessible or intact. Those with higher AT content (GAATTC for *Eco*RI and AAGCTT for *Hind*III) are less accessible; this decreased accessibility is more manifest when the more reactive bases are in a row (*Eco*RI) rather than split by GC (*Hind*III). The behaviour of the released fragment (Fig. 1, Table 1) follows quantitatively the same order.

Among the factors that can prevent the recognition of a specific sequence are: (1) direct base modification within the specific sequence, (2) modification of a nearby base, since the linkage might either distort the sequence configuration or create a steric hindrance with the linking arm, or (3) if two different bases, not necessarily within the sequence, but in different strands, react, they can modify the sequence as above or by local strand separation. The results reported here cannot discriminate between the different possibilities,

but no matter what the mechanism of inactivation is, the remaining intact sites are enough to detect sequence-specific-DNA proteins. These values can be an underestimate if there is partial digestion. Though the digestion was for 10h, incompleteness cannot be ruled out, and one partial digest was in fact observed for *Hind*III. The loss of specific sites by base modification can be bypassed by increasing the copy number of the sequence of interest by molecular cloning techniques.

The determination of accessibility by restriction analysis and end labelling can be used as a random check, by proper restriction enzyme selection, of the state of a DNA-epoxycellulose matrix before using it with valuable protein samples. The present method as it is now, due to its high yield, capacity, accessibility to specific DNA sequences, stability of linkage and lack of risk during preparation might be of great help in the identification and purification of sequence-specific DNA-binding proteins.

I am grateful to my colleague Marjorie Shapiro for bringing epoxycellulose to my attention, to Dr. Martin Weigert for support and to Terry Baumgardner and Ella Cloeren for typing the manuscript. I have been supported by an N.I.H. Fogarty International Fellowship from the U.S.P.H.S.

References

- Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R. & Wahl, G. M. (1979) *Methods Enzymol.* **68**, 220-242
- Arndt-Jovin, D. J., Jovin, T. M., Bahr, W., Frischauf, A. M. & Marquardt, M. (1975) *Eur. J. Biochem.* **54**, 411-413
- Bunemann, H. (1982) *Nucleic Acids Res.* **10**, 7180-7196
- Bunemann, H., Westhoff, P. & Herrman, R. G. (1982) *Nucleic Acids Res.* **10**, 7163-7179
- Gilham, P. T. (1971) *Methods Enzymol.* **21**, 191-197
- Moss, L. G., Moore, J. P. & Chan, L. (1981) *J. Biol. Chem.* **256**, 12655-12658
- Noyes, B. E. & Stark, G. R. (1975) *Cell* **5**, 301-320
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Peterson, G. B. (1982) *J. Mol. Biol.* **162**, 729-773
- Siddell, S. G. (1978) *Eur. J. Biochem.* **92**, 621-629
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517
- Takeda, Y., Ohleudorf, D. H., Anderson, N. F. & Matthews, B. W. (1983) *Science* **221**, 1020-1026