The effect of sequence specific DNA methylation on restriction endonuclease cleavage

Michael McClelland

Department of Molecular and Population Genetics, The University of Georgia, Athens, GA 30602, USA

Received 9 September 1981

ABSTRACT

Sequence specific DNA methylation sometimes results in the protection of some or all of a restriction endonucleases' cleavage sites. This is usually, but not always, the result of methylation of one or both strands of DNA at the site characteristic of the corresponding "cognate" modification methylase. The known effects of sequence specific methylation on restriction endonucleases are compiled.

THE EFFECT OF COGNATE METHYLATION BY MODIFICATION METHYLASES ON THEIR CORRESPONDING ENDONUCLEASE*

As predicted by Arber in 1965, restriction modification (R/M) systems consist of a sequence specific endodeoxyribonuclease, which acts on both strands of double stranded DNA, and/or corresponding modification methylase that recognizes the same sequence and protects the organisms' DNA from cleavage (1). Methylation takes place at the ⁶N position of adenine or ⁵C of cytosine using S-adenosyl methionine as the methyl donor (38). Methylation of only one strand (hemimethylation) is generally sufficient to protect against the endogenous endonuclease (46). This is necessary since newly replicated hemimethylated DNA would otherwise be degraded before the daughter strand could be methylated.

There are three known R/M systems classified as Type I, II, and III (2), (27), (38): Type I and III have their restriction endonuclease and modification methylase activities in the same protein. Both types require ATP and a divalent cation for endonuclease activity but differ in that Type I cuts at random after identifying an unmethylated recognition sequence (29) whereas Type III cuts site-specifically (2); Type II R/M

*Note: The nomenclature for identifying modification methylases is adapted from that of Smith and Nathans for restriction endonuclease (51) by prefixing the restriction endonuclease designation by a capital M to denote the complimentary methylase.

systems are the most studied and consist of a separate restriction endonuclease and modification methylase. The restriction endonuclease does not require ATP for its activity. A typical Type II restriction endonuclease cuts within its "palindromic" recognition sequence and its cognate modification methylase methylates symmetrically within the same sequence at what I shall call the "cognate sites". For example, the Hind II restriction endonuclease and MHind II modification methylase recognize:

$$\begin{array}{c} 5' \text{ G T P y } \text{Pu}^{\text{m}} \text{ A C } 3' \\ 3' \text{ C A Pu } \text{Py T G } 5' \end{array}$$

(45) $\text{m}^{\text{m}}\text{A}$ and $\text{m}^{\text{m}}\text{A}$ represent the ^6N methyl adenines produced by MHind II which protect against cleavage by Hind II at the arrows $\downarrow$.

There are a number of exceptions to the general rule of symmetrical double stranded methylation. These include MMbo II which has not been isolated but by virtue of its recognition sequence:

$$\begin{array}{c} 5' \text{ GAAGA } 3' \\ 3' \text{ CTTCT } 5' \end{array} \quad (8)$$

must have some unusual properties. One strand contains no cytosines and the other strand contains no adenines. Methylation must take place on only one strand, or asymmetrically at an adenine and a cytosine. MEcoPI is from a Type III system and methylates at $\begin{array}{c} 5' \text{ AG }^{\text{m}} \text{ ACC } 3' \\ 3' \text{ TC TGG } 5' \end{array}$ (2). Apparently only one strand is methylated; the other strand contains no adenines. It is a moot point why restriction does not occur during replication. It is possible that the concentration of ATP is sufficiently low around newly replicated DNA that the ATP dependent cleavage does not occur (2). Finally, Dpn I represents a unique restriction system. The endonuclease only cleaves the methylated sequence $\begin{array}{c} 5' \text{ G }^{\text{m}} \text{ A TC } 3' \\ 3' \text{ C T }^{\text{m}} \text{ AG } 5' \end{array}$ (28) when both adenines are methylated.

Strains of Diplococcus pneumoniae which carry this enzyme have no modification methylase. Other strains of this species have a conventional Type II R/M system. They contain Dpn II which cuts at GATC and a cognate modification methylase which methylates $\text{G}^{\text{m}}\text{ATC}$ and protects against Dpn II (28).

THE EFFECT OF COGNATE METHYLATION BY HETEROLOGOUS METHYLASES

Some endonuclease recognition sequences can be protected from cleavage by "cognate site" methylation with heterologous methylases that have recognition sequences overlapping but not identical to those of the endonu-

Table I is organized by length of the restriction endonuclease recognition sequence and then alphabetically by sequence. Isoschizomers are listed alphabetically by name. Only the effects of methylation at the N^6 position of adenine and C^5 of cytosine are considered. References to the purification of restriction enzymes and the determination of their recognition sequences can be found in R.J. Roberts review (43).

TABLE I. # denotes known modification methylase specificity

* denotes probable modification methylase specificity

J = A or C, K = G or T, N = A, C, G, or T, R = A or G, Y = C or T, X = A or T

Restriction Enzyme (55)	Recognition Sequence	Methylated Sequences Cut	Methylated Sequences Not Cut	Effect Of Methylation Unknown	References
<u>AluI</u>	AGCT	?	AG ^m CT	^m AGCT	20
<u>HapII</u>	CCGG	?	C ^m CGG [#]	^m CCGG	14, 59
<u>HpaII</u>	CCGG	^m CCGG	C ^m CGG [#]	-----	14, 33, 59
<u>MspI</u>	CCGG	C ^m CGG	^m CCGG ^(a)	-----	14, 24, 54, 59
<u>BstEIII</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C	37, 43
<u>DpnI</u>	GATC	G ^m ATC	GATC	(only cuts methylated DNA)	28
<u>DpnII</u>	GATC	?	G ^m ATC [*]	GAT ^m C	28, 43, 58
<u>FnuAII</u>	GATC	?	G ^m ATC ^(a)	GAT ^m C	37, 43
<u>FnuCI</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C	43
<u>FnuEI</u>	GATC	G ^m ATC	?	GAT ^m C [*]	31, 43
<u>MboI</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C	12, 18, 35, 43, 53
<u>MnoIII</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C ^(d)	37, 43
<u>MosI</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C	43
<u>MphI</u>	GATC	?	G ^m ATC ^(d)	GAT ^m C	43
<u>PfaI</u>	GATC	G ^m ATC	?	GAT ^m C [*]	43, 55
<u>Sau3A</u>	GATC	G ^m ATC	GAT ^m C [*]	-----	12, 14, 36, 42, 43, 53
<u>HhaI</u>	GCGC	-----	G ^m CGC [#] GCC ^m C	-----	14, 32, 34, 52
<u>BsuRI</u>	GGCC	?	GG ^m CC [#]	GGC ^m C	21
<u>HaeIII</u>	GGCC	GGC ^m C	GG ^m CC [#]	-----	32, 33
<u>TaqI</u>	TCGA	T ^m CGA	TCG ^m A	-----	20, 36, 54
<u>TthI</u>	TCGA	T ^m CGA	TCG ^m A	-----	47
<u>HinFI</u>	GANTC	GANT ^m C	?	G ^m ANTC	20, 39
<u>Sau96</u>	GGNCC	?	GGN ^m CC	GGN ^m CC	39
<u>ApyI</u>	CCXGG	C ^m CXGG	?	^m CCXGG [*]	10, 42, 43
<u>AtuBI</u>	CCXGG	?	C ^m CXGG ^(d)	^m CCXGG	43, 44
<u>AtuII</u>	CCXGG	?	C ^m CXGG ^(d)	^m CCXGG	43
<u>BstNI</u>	CCXGG	C ^m CXGG	?	^m CCXGG ^(b)	43
<u>EcaII</u>	CCXGG	?	C ^m CXGG ^(d)	^m CCXGG	43
<u>EclII</u>	CCXGG	?	C ^m CXGG ^(d)	^m CCXGG	43
<u>EcoRII</u>	CCXGG	?	C ^m CXGG [#]	^m CCXGG	6, 7, 35, 43, 38
<u>MphI</u>	CCXGG	?	C ^m CXGG ^(d)	^m CCXGG	25, 43
<u>BbvI</u>	GCGC	?	G ^m CXGC [#]	GCXG ^m C	22, 57
<u>AvaII</u>	GGXCC	?	GGXC ^m C	GGX ^m CC	3
<u>EcoPI</u>	AGACC (type III)	?	AG ^m ACC [#]	^m AGACC AGA ^m CC GGT ^m CT	2, 23

TABLE I. (cont.)

Restriction Enzyme (55)	Recognition Sequence	Methylated Sequences Cut	Methylated Sequences Not Cut	Effect Of Methylation Unknown	References
<u>Mbo</u> II	GAAGA	?	GAAG ^m A cytosine methylation	GA ^m AGA G ^m AAGA	3, 20
<u>Ava</u> I	CYCGRG	?	cytosine methylation	?	5, 14, 26
<u>Aos</u> II	GRCGYC	?	GR ^m CGYC	GRCGY ^m C	20, 54
<u>Acc</u> I	GTJKAC	?	GTJK ^m AC	GTJKA ^m C	36
<u>Hind</u> II	GTYRAC	GTYRA ^m C	GTYR ^m AC [#]	-----	20, 45
<u>Hae</u> II	RGCCGCY	?	RG ^m CGCY	RGCG ^m CY	14
<u>Hind</u> III	AAGCTT	GTYRA ^m C	GTYR ^m AC [#]	-----	20, 45
<u>Bgl</u> II	AGATCT	AG ^m ATCT	?	^m AGATCT AGAT ^m CT	4, 12, 40
<u>Cla</u> I	ATCGAT	?	ATCG ^m AT [#]	AT ^m CGAT ATCGAT ^m	36
<u>Pvu</u> II	CAGCTG	?	cytosine methylation	?	11
<u>Sma</u> I	CCCGGG	?	CC ^m CGGG	C ^m CCGGG mCCCGGG	14, 17, 41
<u>Xma</u> I	CCGCGG	CC ^m CGGG	?	C ^m CCGGG mCCCGGG	60
<u>Sac</u> II	CCGCGG	?	cytosine methylation	?	14
<u>Xor</u> II	CGATCG	?	cytosine methylation	?	14
<u>Xho</u> I	CTCGAG	?	CT ^m CGAG CTCG ^m AG	^m CTCGAG	14, 36, 54
<u>Pst</u> I	CTGCAG	?	cytosine methylation	?	11, 20
<u>Sal</u> PI	CTGCAG	?	cytosine methylation	?	9
<u>Eco</u> RI	GAATTC	?	GA ^m ATTC [#]	G ^m AATC GAATT ^m C	13, 19
<u>Bam</u> HI	GGATCC	GGATC ^m C	GGAT ^m CC [#]	GG ^m ATCC ^(c)	12, 22, 33
<u>Sal</u> I	GTCGAC	?	GT ^m CGAC	GTCGA ^m C GTCG ^m AC	14, 36, 54
<u>Hpa</u> I	GTTAAC	GTTA ^m C	?	GTT ^m AAC GTTA ^m AC	20
<u>Atu</u> CI	TGATCA	?	TG ^m ATCA ^(d)	TGAT ^m CA TGATC ^m A	23, 43, 49
<u>Bcl</u> I	TGATCA	?	TG ^m ATCA ^(d)	TGAT ^m CA TGATC ^m A	4, 23, 43
<u>Cpe</u> I	TGATCA	?	TG ^m ATCA ^(d)	TGAT ^m CA TGATC ^m A	15, 23, 43
<u>Eco</u> B	TGA(N) ₈ TGCT (Type I)	?	TG ^m A(N) ₈ TGCT [#] AGC ^m A(N) ₈ TCA [#]	?	30, 56

THE EFFECT OF NON-COGNATE METHYLATION

Non-cognate sites are those that are not methylated by an endonuclease corresponding modification methylase. Methylation within some endonuclease recognition sequences at non-cognate sites can sometimes protect DNA from cleavage. For example, MHha I (G^mCGC) methylation and GCG^mC methylation *both* protect DNA from Hha I (GCGC) endonuclease cleavage (32) (52). Conversely, methylation within a recognition sequence may have no effect on the restriction endonuclease; BamHI (GGATCC) cuts at GGATC^mC (33) but not at the MBamHI methylation site GGAT^mCC (22).

Presumably only "non-cognate methylation" which interferes with DNA/enzyme interactions will block restriction cutting. Methylation at any ⁶N of adenine or ⁵C of cytosine which is involved in the interaction of an endonuclease with its recognition sequence could render the DNA resistant to cleavage. Methylation within a recognition sequence which does

- 5863

not result in protection of the DNA from a restriction endonuclease can arise for two reasons; the endonuclease may interact with the base complementary to the one methylated or it may interact with some part of the methylated base which is not sterically shielded by the methyl group (50). Since it is not possible to predict the effect of methylation without knowing the sites of interaction of an endonuclease with its recognition sequence, this must be determined empirically.

IMPLICATIONS

The information presented here indicates that caution may be necessary when using some restriction endonucleases to study DNA that has an uncharacterized methylation pattern. Some DNAs may be protected from cleavage due to methylation at any of a number of sites within a particular endonuclease recognition sequence.

Studies on eucaryotic DNA methylation have been performed using Hpa II (CCGG) (16) and Msp I (CCGG) (18). Hpa II cuts m CCGG (12) but not C^m CGG (33) and Msp I cuts C^m CGG (59) but not m CCGG (54). Since these enzymes have been well characterized as to the effect of methylation at cytosines in their recognition sequence, semi-quantitative studies on the level of m CG and m CC (the two preponderant sites of methylation in eucaryotes) are possible (54). Unfortunately, it is not clear what effect methylation at each possible adenine and cytosine in the recognition sequence of many restriction endonucleases may have on their ability to cut DNA. For example, Xho I does not cut CT^m CGAG (54) or $CTCG^m$ AC (36), but its ability to cut m CTCGAG is not known. Until the effect of methylation at this site is known, the inability of Xho I to cut a particular methylated sequence cannot be attributed to m CG methylation.

Similarly, physical mapping of some DNAs can lead to ambiguous results if some restriction sites are protected from cleavage by methylation at either cognate or non-cognate sites. (For example, see (3) and (53).) The data in Table I gives clues to the methylation sites of many unstudied modification methylases; sites that do not result in inhibition of a restriction endonuclease when methylated cannot be used by the corresponding modification methylase; thus for instance Sau3A cuts at G^m ATC indicating that MSau3A probably methylates at GAT^m C.

Clearly, it is important that more of the effects of methylation on the ability of endonucleases to cleave DNA should be determined.

ACKNOWLEDGEMENTS

I would like to thank Dr. R.A. Lansman, Dr. D. Vapnek, Dr. M. Bittner and Dr. R. Meagher for helpful discussions. This work was supported in part by a grant from the National Institutes of Health, RF 10-21-RR093-009.

REFERENCES

1. Arber, W. (1965) *Am. Rev. Microbiol.* **19**, p. 365.
2. Bächli, B., Reiser, J., and Pirrotta, V. (1979) *J. Mol. Biol.* **128**, pp. 143-163.
3. Backman, K. (1980) *Gene* **11**, pp. 167-171.
4. Bingham, A.H.A., Atkinson, T., Sciaky, D., and Roberts, R.J. (1978) *Nucleic Acids Res.* **5**, 3457-3467.
5. Bird, A.P., Taggart, M.H., Smith, B.A. (1979) *Cell* **4**, 889-902.
6. Bogdarina, I.G., Vagabova, L.M., and Buryanov, Y.I. (1976) *FEBS Lett.* **68**, pp. 177-180.
7. Boyer, H.W., Chow, L.T., Dugaiczky, A., Hedgpeth, J., and Goodman, H.M. (1973) *Nature New Biology* **244**, pp. 40-43.
8. Brown, N.L., Hutchison, C.A. III, and Smith, M. (1980) *J. Mol. Biol.* **140**, pp. 143-148.
9. Carter, J.A., Chater, K.F., and Bruton, C.J. (1980) *Nucleic Acids Res.* **8**, pp. 4943-4954.
10. Dilauro, R. Unpublished results.
11. Dobritsa, A.P. and Dobritsa, S.V. (1980) *Gene* **10**, pp. 105-112.
12. Dreiseikelman, B., Eichenlaub, R., and Wackernagel, W. (1979) *Biochem. Biophys. Acta* **562**, pp. 418-428.
13. Dugaiczky, A., Hedgpeth, J., Boyer, H.W., and Goodman, H.M. (1974) *Biochemistry* **13**, 503-512.
14. Ehrlich, M. and Wang, R.Y.H. (1981) *Science* **212**, pp. 1350-1357.
15. Fisherman, J., Gingeras, T.R., and Roberts, R.J. Unpublished observations.
16. Garfin, D.E. and Goodman, H.M. (1974) *Biochem. Biophys. Res. Comm.* **59**, pp. 108-116.
17. Gautier, F., Bunemann, H., and Grotjahn, L. (1977) *Eur. J. Biochem.* **80**, pp. 175-184.
18. Gelinas, R.E., Myers, P.A., Roberts, R.J. (1977) *J. Mol. Biol.* **114**, pp. 169-180.
19. Greene, P.J., Betlach, M.C., Goodman, H.M., and Boyer, H.W. (1974) *Methods Mol. Biol.* **7**, pp. 87-111.
20. Gruenbaum, Y., Cedar, H., and Razin, A. (1981) *Nucleic Acids Res.* **9**, pp. 2509-2515.
21. Gunthert, V., Freund, M., and Trautner, T.A. (1978) Abstracts of 12th FEBS Symposium, Dresden.
22. Hattman, S., Keisler, T., and Gottehrer, A. (1978) *J. Mol. Biol.* **124**, pp. 701-711.
23. Hattman, S., Brooks, J.E., Masurekar, M. (1978) *J. Mol. Biol.* **126**, pp. 367-380.
24. Jentsch, S., Günthert, V. and Trautner, T.A. (1981) *Nucleic Acids Res.* **9**, 2753-2759.
25. Jiang, B.C. and Myers, P. Unpublished observations.
26. Kaput, J. and Sneider, T.W. (1979) *Nucleic Acids Res.* **7**, pp. 2303-2322.
27. Kauc, L. and Piekarowicz, A. (1978) *Eur. J. Biochem.* **92**, pp. 412-426.
28. Lacks, S. and Greenberg, B. (1977) *J. Mol. Biol.* **114**, pp. 153-168.
29. Lautenberger, J.A., Kan, N.C., Lackey, D., Linn, S., Edgell, M.H., and Hutchison, C.A. III (1978) *Proc. Natl. Acad. Sci. USA* **75**, pp. 2271-2275.

30. Lautenberger, J.A. and Linn, S. (1972) *J. Biol. Chem.* 247, pp. 6176-6182.
- 30a. Levy, N.P. and Walker, (1981) *Biochem.* 20, pp. 1121-1127.
31. Lui, A.P.C., McBride, B.C., Vovis, G.F., and Smith, M. (1979) *Nucleic Acids Res.* 6, pp. 1-15.
32. Mann, M.B. and Smith, H.O. (1977) *Nucleic Acids Res.* 4, pp. 4211-4221.
33. Mann, M.B. and Smith, H.O. (1979) *Proc. of the conference on Trans-methylation*, Eds. Usdin, E., Borchardf, R.T. and Greveling, C.R. Elsevier/North-Holland; New York, P. 483.
34. Mann, M.B. and Smith, H.O. Unpublished observations.
35. May, M.S. and Hattmann, S. (1975) *J. Bacterial* 122, pp. 129-138.
36. McClelland, M. Manuscript to be submitted.
- 36a. Modrich, P. (1977) *Quant Rev. Biophys.* 12, pp. 315-369.
37. Myers, P.A. and Roberts, R.J. Unpublished observations.
38. Nathans, D. and Smith, H.O. (1974) *Ann. Rev. Biochem.* 44, p. 273.
39. Pech, M., Streeck, R.E., Zachau, H.G. (1979) *Cell* 18, pp. 883-893.
40. Pirrotta, V. (1976) *Nucleic Acids Res.* 3, pp. 1747-1760.
41. Quint, A. and Cedar, H. (1981) *Nucleic Acids Res.* 9, pp. 633-646.
42. Razin, A., Urieli, S., Pollack, Y., Gruenbaum, Y., and Glaser, G. (1980) *Nucleic Acids Res.* 8, pp. 1783-1792.
43. Roberts, R.J. (1981) *Nucleic Acids Res.* 9, pp. 75-96.
44. Roize, G., Patillon, M. and Kovoov, A. (1977) *FEBS Lett.* 82, pp. 69-70.
45. Roy, P.H. and Smith, H.O. (1973) *J. Mol. Biol.* 81, pp. 445-459.
46. Rubin, R.A. and Modrich, P.J. (1977) *J. Biol. Chem.* 252, pp. 7265-7272.
47. Sato, S., Nakazawa, K., and Shinomiya, T. (1980) *J. Biochem.* 88, pp. 737-747.
48. Schlagman, S., Hattman, S., May, M.S., and Berger, L. (1976) *J. Bacterial* 126, pp. 990-996.
49. Sciaky, D. and Roberts, R.J. Unpublished observations.
50. Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* 73, pp. 804-808.
51. Smith, H.O. and Nathans, D. (1973) *J. Mol. Biol.* 81, pp. 445-459.
52. Smith, H.O. (1979) *Science* 205, pp. 455-466.
53. Streeck, R.E. (1980) *Gene* 12, pp. 267-275.
54. Van Der Ploeg, L.H.T. and Flavell, R.A. (1980) *Cell* 19, pp. 947-958.
55. Van Montagu, M. Sciaky, D., Myers, P.A., and Roberts, R.J. Unpublished observations.
56. Van Ormondt, H., Lautenberger, J.A., Linn, S., and Dewaard, A. (1973) *FEBS Lett.* 33, pp. 177-180.
57. Vanyushim, B.F. and Dobritsa, A.P. (1975) *Biochem. Biophys. Acta* 407, pp. 61-71.
58. Vovis, G.F. and Lacks (1977) *J. Mol. Biol.* 115, pp. 525-538.
59. Waalwijk, C. and Flavell, R.A. (1978) *Nucleic Acids Res.* 5, p. 3231.
60. Youssoufion, H. and Mulder, C. (1981) *J. Mol. Biol.* 150, pp. 133-136.