The effect of site specific methylation on restriction endonuclease digestion

Michael McClelland and Michael Nelson[1]

Department of Human Genetics and Development, Columbia University, 701 W168th St., New York, NY 10032, and [1]New England Biolabs, 32 Tozer Rd., Beverly, MA 01915, USA

INTRODUCTION

There have been a number of important new developments in the field of site specific DNA methylation since the last compilation (48).

A method for producing restriction endonuclease cleavage at 8, 10 and potentially at 12 or 14 base pair sequences has been developed (49). The technique relies on site specific methylases and methylation dependent restriction. The only methylation dependent restriction systems known are Dpn I (GmATC) and its isoschizomers. However, DNA from E.coli, which is the usual substrate for assaying endonucleases, is unmethylated except at GmATC and CmC(A/T)GG. Thus, it is likely that there are other methylation dependent endonucleases which have not been detected with this substrate.

Nelson et al have investigated a large number of restriction modification methylases as blocking agents for restriction endonucleases (52). Methylase specificities which partly overlap and block cleavage at endonuclease recognition sequences give restriction specificities which are both novel and rarer. For example, a subset of Nae I (GCCGGC) sites, GGCCGGC and GCCGGCC, are blocked by M.Hae III (GGmCC). The specificity of Nae I is thus (A/C/T)GCCGGC(A/G/T) in M.Hae III methylated DNA.

A new type of modification methylase specificity has been found for M.Bcn I (CC(C/G)GG) which methylates at the N-4 position of cytosine (30). Bst NI (CC(A/T)GG) from a related species is known to cut DNA that is methylated at the C-5 position of cytosine at any combination of cytosines in its recognition sequence. It seems likely that M.Bst NI is also an

#denotes a known 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,
Z=G or C, D=A,G or T, H=A,C or T
mC= 5-methylcytosine, mA= 6-methyladenine
Nomenclature is that proposed by Smith and Nathans (65)

Restriction Enzyme	Recognition Sequence	Sites cut	Sites not cut	References
Alu I	AGCT	------	mAGCT AGmCT#	23,52,53
Bsu F I	CCGG	?	mCCGG#	33
Hap II	CCGG	?	CmCGG#	16,75
Hpa II	CCGG	------	CmCGG# mCCGG(a)	16,41,42,55,77
Msp I	CCGG	CmCGG(b)	mCCGG#(a)(c)	16,32,70,75,77
Tha I	CGCG	------	mCGCG CGmCG	67
Bst E III	GATC(d)	?	GmATC	50,57
Dpn I	GmATC(e)	GmATC only cuts methylated DNA		37,74
Fnu E I	GATC	GmATC	?	40,52
Mbo I	GATC(d)	GATmC	GmATC#	8a,20,44,57
Pfa I	GATC	GmATC	?	57,71
Sau 3A	GATC(d)	GmATC	GATmC(a)	13,16,46,57
Hha I	GCGC	------	GmCGC# GCGmC	16,43,66
Hin P I	GCGC	?	GmCGC	53
Bsu R I	GGCC	?	GGmCC#(a)	24
Hae III	GGCC	GGCmC	GGmCC#(a)	2,41,42
Ngo II	GGCC	?	GGmCC#	36a
Taq I	TCGA	TmCGA	TCGmA#	23,46,70
Tth I	TCGA	TmCGA	TCGmA#	62
Tfl I	TCGA	?	TCGmA#	62
Scr F I	CCNGG	mCCNGG	CmCNGG	52,53
Dde I	CTNAG	?	mCTNAG	52
Hinf I	GANTC		GmANTC GANTmC(f)	52,53a
Fnu 4H I	GCNGC	?	GmCNGC	69
Sau 96 I	GGNCC	------	GGNmCC GGCCmC	44,52,53a
Aac I	CCXGG	CmCXGG	?	8
Apy I	CCXGG	CmCXGG	mCCXGG(g)	11,45,56,57
Bst N I	CCXGG(h)	CmCXGG mCCXGG(j)	-------	23,45,57
Eco R II	CCXGG(h)	mCCXGG	CmCXGG#	6,44,45,51,57,63
Mph I	CCXGG(h)	?	CmCXGG	34,57
Taq X I	CCXGG	mCCXGG	CmCXGG	22
Bcn I	CCZGG	mCCZGG	CmCZGG#(k)	29,30
Nci I	CCZGG	?	CmCZGG(l)	8,45
Bbv I	GCXGC	?	GmCXGC#	12,25,73
Ava II	GGXCC	?	GGXCmC	2,43
Eco 47 I	GGXCC	?	GGXCmC	31

Restriction Enzyme	Recognition Sequence	Sites cut	Sites not cut	References
Eco P I	AGACY(m)	?	AGmACY#	1,26
Fok I	CATCC	?	CATCmC	53
Mbo II	GAAGA	?	GAAGmA# 5mC	2,23,52,53
Hga I	GACGC	?	GACGmC	53
Sfa N I	GATGC	GATGmC	?	53
Hph I	TCACC	?	TmCACC# GGTGmA	52,53
Bsp I 1286	GDGCHC	?	GDGmCHC	52
Ava I	CYCGRG	CmCCGGG	CYmCGRG CTCGmAG	5,16,33,35,45,52
Aos II	GRCGYC	?	GRmCGYC	16,23,70
Aha II	GRCGYC	------	GRmCGYC GRCGYmC	52
Ban II	GRGCYC	?	GRGmCYC	52
Acc I	GTJKAC	?	GTJKmAC	46
Hin C II	GTYRAC	GTYRAmC(n)	GTYRmAC#	23,60
Hgi A I	GXGCXC	?	GXGmCXC	52
Hae II	RGCGCY	?	RGmCGCY	16,23
Ngo I	RGCGCY	?	RGmCGCY#	36a
Xho II	RGATCY	RGmATCY	RGATmCY	8
Eae I	YGGCCR	?	YGGCmCR	78a
Hind III	AAGCTT	?	mAAGCTT# AAGmCTT	8,23,60
Mlu I	ACGCGT	mACGCGT	?	53
Bgl II	AGATCT	AGmATCT	AGATmCT	4,8,13,15,54
Stu I	AGGCCT	?	AGGmCCT	53
Cla I	ATCGAT	?	ATCGmAT#	46
Pvu II	CAGCTG	?	CAGmCTG	8,12,58
Nco I	CCATGG	?	mCCATGG	52
Sma I	CCCGGG	?	CCmCGGG(1)	8,16,19,55
Xma I	CCCGGG	CCmCGGG(p)	CmCCGGG	80,81
Sac II	CCGCGG	?	mCCGCGG	52
Pvu I	CGATCG	CGmATCG	CGATmCG	8
Xor II	CGATCG	CGmATCG	CGATmCG	8,16
Xma III	CGGCGG	?	CGGmCCG#	69
Bsu M I	CTCGAG	?	CTmCGAG#	33
Pae R7	CTCGAG	?	CTCGmAG#	20a
Xho I	CTCGAG	?	CTmCGAG CTCGmAG	8,16,46,70
Pst I	CTGCAG	?	CTCGmAG# mCTGCAG	12,23,52,53,76
Sfl I	CTGCAG	?	CTGCmAG	8
Eco R I	GAATTC	GmAATC	GAmATTC# GAATTmC	14,17,21,52,61
Sac I	GAGCTC	GmAGCTC	GAGmCTC	53
Sst I	GAGCTC	?	GAGmCTC	8,58
Eco R V	GATATC	?	GmATATC	52
Nae I	GCCGGC	?	GmCCGGC GCCGGmC	52,53

Restriction Enzyme	Recognition Sequence	Sites cut	Sites not cut	References
Sph I	GCATGC	GCATGmC	?	52
Nhe I	GCTAGC	?	5mC	52
Bam H I	GGATCC	GGATCmC GGmATCC	GGATmCC	8,13,25,42
Nar I	GGCGCC	GGCGCmC	?	53
Kpn I	GGTACC	GGTAmCC GGTACmC	?	53
Apa I	GGGCCC	?	GGGmCCC#	69
Sal I	GTCGAC	?	GTCGmAC GTmCGAC	8,16,46,70
Hpa I	GTTAAC	GTTAAmC	GTTAmAC#	8,23,79
Nru I	TCGCGA	?	TCGCGmA	52
Xba I	TCTAGA	?	TmCTAGA TCTAGmA	23,27,52
Atu C I	TGATCA	?	TGmATCA	57,64
Bcl I	TGATCA	TGATmCA	TGmATCA	2,4,8,57
Bst GI	TGATCA		TGmATCA	57
Cpe I	TGATCA	?	TGmATCA	18,57
Bal I	TGGCCA	?	TGGmCCA#	69
Bst X I	CCAN$_6$TGG	?	mCCAN$_6$TGG CCmAN$_6$TGG	52
Mst II	CCTNAGG	mCCTNAGG	?	53
Xmn I	GAAN$_4$TTC	GAmAN$_4$TTC	GmAAN$_4$TTC	52,53
Bgl I	GCCN$_5$GGC	GCmCN$_5$GGC	GmCCN$_5$GGC	52,53
Bst E II	GGTNACC	GGTNAmCmC	?	27
Eco K	AACN$_6$GTGC(q)	?	AmACN$_6$GmTGC#	3
Eco A	GAGN$_7$GTCA	?	GmAGN$_7$GmTCA#	3
Eco B	TGAN$_8$TGCT(q)	?	TGmAN$_8$mTGCT#	3,38,39
Not I	GCGGCCGC	GCGGCCGmC	GCGGmCCGC	53
Sfi I	GGCCN$_5$GGCC	GGmCCN$_5$GGmCC GGCCN$_5$GGCmC	?	53

N-4 cytosine methylase (17).

The first example of protection from cleavage by methylation at a partly redundant site in a recognition sequence has been found; methylation at A in the Ava I recognition sequence C(T/C)CG(A/G)G protects against a subset of Ava I sites (52). However, it should be noted that this is not a general phenomenon, for instance, Ava I will cut CmCCGGG (53).

Finally, Msp I normally cuts CmCGG but does not do so at the sequence GGCmCGG (9,36). The insensitivity to Msp I digestion of certain CCGG sequences in vertebrate DNA had been erroneously attributed to mCCGG. Although, for some restriction endonucleases certain recognition sites are cut more slowly than others, this is the first example of a sequence outside a restriction endonuclease recognition sequence affecting the ability of the restriction enzyme to cleave a methylated recognition sequence.

Notes

a) Nicking occurs in the unmethylated strand of the hemimethylated sequence. For Hpa II see (77), for Sau 3A see (2,68), for Msp I and Hae III see (25), for Bsu RI see (7).
b) Msp I fails to cut GGCmCGG (9,36).
c) An M.Msp I clone methylates mCCGG (77,78). However, there is a report that Msp chromosomal DNA is methylated at mCmCGG (32)
d) Mbo I isoschizomers that are sensitive to GmATC include Bss G II, Bsa P I, Bst X II, Bst E III, Cpa I, Dpn II, Fnu A II, Fnu C I, Mno III, Mos I, Nde II, Nfl I, Nla II, Nsu I and Sin M I (57).
   Sau 3A I isoschizomers that are insensitive to GmATC include Bsr P II, Cpf I, Fnu E I, Mth I, Nsi A I, Pfa I (57).
e) Isoschizomers of Dpn I include Cfu I (28), Nmu E I, Nmu D I and Nsu D I (10).
f) There is evidence that Hinf I cuts GANTmC (23).
g) Apy I may cut CCXGG more slowly than CmCXGG (56).
h) Isoschizomers of Eco R II that are sensitive to CmCXGG include Atu B I, Atu II, Bst G II, Bin S I, Cfr 5 I, Cfr II I, Ecl II, Eca II, Eco 27 I, Eco 38 I and Mph I (57).
   Bst NI isoschizomers that are insensitive to CmCXGG include Aor I, Apy I, Mva I and Taq XI (57).
j) Bst N I cuts CmCXGG, mCCXGG and mCmCXGG. M.Bst NI may be an N-4 cytosine methylase (30,17).
k) M.Bcn I is the first example of a N-4 specific cytosine methylase (30).
l) Sma I and Nci I may cut mCmCGG methylated DNA (8,32). Possibly the second methylation negates the effect of CmCGG.
m) Type III restriction endonuclease (1,26).
n) There is a report that Hin C II does not cut GTYRAmC (27).
p) There is a report that Xma I does not cut CCmCGGG (8).
q) Type I restriction endonuclease. mT represents a 6-methyladenine in the complementary strand.

References

1 Bachi B, Reiser J, and Pirrotta V (1979) J. Mol. Biol. 128:143-163
2 Backman K (1980) Gene 11:167-171
3 Bickle TA (1980) in Nucleases, Linn SM and Roberts RJ, Eds., pp85-108, Cold Spring Harbor Lab, New York
4 Bingham AHA Atkinson T Sciaky D and Roberts RJ (1978) Nucleic Acids Res. 5:3457-3467
5 Bird AP Taggart MH and Smith BA (1979) Cell 17:889-902
6 Boyer HW Chow LT Dugaiczyk A Hedgpeth J and Goodman HM (1973) Nature New Biol. 244:40-43
7 Bron S Murry K and Trautner TA (1975) Mol. Gen. Genet. 143:13-23
8 Brooks JE and Roberts RJ (1982) Nucleic Acids Res. 10:913-934
9 Busslinger M deBoer E Wright S Grosveld FG and Flavell RA (1983) Nucleic Acids Res. 11:3559-3569
10 Camp R and Schildkraut I, unpublished results
11 DiLaurio R, unpublished results
12 Dobritsa AP and Dobritsa SV (1980) Gene 10:105-112
13 Dreiseikelman B Eichenlaub R and Wackernagel W (1979) Biochem. Biophys. Acta 562:418-428

14 Dugaiczyk A Hedgepeth J Boyer HW and Goodman HM (1974) Biochemistry 13:503-512
15 Dybvig K Swinton D Maniloff J and Hattman S (1982) J. Bacteriol. 151:1420-1424
16 Ehrlich M and Wang RYH (1981) Science 212:1350-1357
17 Ehrlich M, unpublished results
18 Fisherman J Gingeras TR and Roberts RJ (unpublished results)
19 Gautier R Bunemann H and Grotjahn L (1977) Eur. J. Biochem. 80:175-183
20 Gelinas RE Myers PA and Roberts RJ (1977) J. Mol. Biol. 114:169-180
20a Gingeras TR and Brooks JE (1983) Proc. Natl. Acad. Sci. USA. 80:402-406
21 Greene PJ Betlach MC Boyer HW and Goodman HM (1974) Methods Molec. Biol. 7:87-111
22 Grachev SA Mamaev SV Gurevich AL Igoshun AV Kolosov MN and Slyusarenko AG (1981) Bioorg. Khim. 7:628-630
23 Gruenbaum Y Cedar H and Razin A (1981) Nucleic Acids Res. 9:2509-2515
24 Gunthert U Storm K and Bald R (1978) Eur. J. Biochem. 90:581-583
25 Hattman S Keister T and Gottehrer A (1978) J. Mol. Biol. 124:701-711
26 Hattman S Brooks JE and Masurekar M (1978) J. Mol. Biol. 126:367-380
27 Huang L-H Farnet CM Ehrlich KC and Ehrlich M (1982) Nucleic Acids Res. 10:1579
28 Hurlin P and Schildkraut I, unpublished results
29 Januliatis A and Petrusyte MP (1982) Dokl. Akad. Nauk. USSR 257:749-750
30 Janulaitis A Klimasauskas S Petrusyte M (1983) FEBS Lett. 161:131-134
31 Janulaitis A Petrusyte M and Butkus V (1983) FEBS Lett. 161:213-216
32 Jentsch S Gunthert U and Trautner TA (1981) Nucleic Acids Res. 9:2753-2759
33 Jentsch S (1983) J. Bacteriol. 156:800-808
34 Jiang BD and Myers P, unpublished results
35 Kaput J and Sneider TW (1979) Nucleic Acids Res. 7:2303-2322
36 Keshet E and Cedar H (1983) Nucleic Acids Res. 11:3571-3580
36a Korch C Hagblom P and Normark S (1983) J. Bacteriol. 155:1324-1332
37 Lacks S and Greenberg (1977) J. Mol. Biol. 114:153-168
38 Lautenberger JA Kan NC Lackey D Linn S Edgell MH and Hutchison III CA (1978) Proc. Natl. Acad. Sci. USA 75:2271-2275
39 Lautenberger JA and Linn S (1972) J. Biol. Chem. 247:6176-6182
40 Lui APC McBride BC Vovis GF and Smith M (1979) Nucleic Acids Res. 6:1-15
41 Mann MB and Smith HO (1977) Nucleic Acids Res. 4:4211-4221
42 Mann MB and Smith HO (1979) Proc. of the Conference on Transmethylation, Eds. Usdin E Borchardt RT and Greveling CR Elsevier/North Holland NY:483-492
43 Mann MB and Smith HO, unpublished results
44 May MS and Hattman S (1975) J. Bacteriol. 122:129-138
45 McClelland M, unpublished results
46 McClelland M (1981) Nucleic Acids Res. 9:6795-6804
47 McClelland M (1981) Nucleic Acids Res. 9:5859-5866
48 McClelland M (1983) Nucleic Acids Res. 10:r169-r173
49 McClelland M Kessler L and Bittner M (1984) Proc. Natl. Acad. Sci. USA 81:983-987
50 Myers PA and Roberts RJ, unpublished results
51 Nathans D and Smith HO (1974) Ann. Rev. Biochem. 44:273-293
52 Nelson M Christ C and Schildkraut I (1984) Nucleic Acids Res. 12:5165-5173
53 Nelson M, unpublished results
53a Pech M Streeck RE and Zachau HG (1979) Cell 18:883-893
54 Pirrotta V (1976) Nucleic Acids Res. 3:1747-1760
55 Quint A and Cedar H (1981) Nucleic Acids Res. 9:633-646
56 Razin A Urieli S Pollack Y Gruenbaum Y and Glaser G (1980) Nucleic Acids Res. 8:1783-1792
57 Roberts RJ (1981) Nucleic Acids Res. 12:r167-r209
58 Roberts RJ, unpublished results
59 Roize G Patillon M and Kovoor A (1977) FEBS Lett 82:69-70
60 Roy PH and Smith HO (1973) J. Mol. Biol. 81:445-459
61 Rubin RA and Modrich PJ (1977) J. Biol. Chem. 252:7265-7272
62 Sato S Nakazawa K and Shinomiya T (1980) J. Biochem. 88:737-747
63 Schlagman S Hattman S May MS and Berger L (1976) J. Bacteriol. 126:990-996
64 Sciaky D and Roberts RJ, unpublished results
65 Smith HO and Nathans D (1973) J. Mol. Biol. 81:419-423
66 Smith HO (1979) Science 205:455-462
67 Strobl JS Thompson EB (1984) Nucleic Acids Res. 12:8073-8084
68 Streek RE (1980) Gene 12:267-275
69 Trautner TA, in Methylation of DNA, ed. Trautner TA, Current Topics in

Microbiology and Immunology 108:11-22
70 Van der Ploeg LHT and Flavell RA (1980) Cell 19:947-958
71 Van Montagu M Sciaky D Myers PA and Roberts RJ, unpublished results
72 Van Ormondt H Lautenberger JA Linn S de Waard RJ (1973) FEBS Lett 33:177-180
73 Vanyushin BF and Dobritsa AP (1975) Biochim. Biophys Acta 407:61-72
74 Vovis GF and Lacks S (1977) J. Mol. Biol. 115:525-538
75 Waalwijk C and Flavell RA (1978) Nucleic Acids Res. 5:3231
76 Walder RY, personal communication
77 Walder RY Langtimm CJ Chatterjee R and Walder JA (1983) J. Biol. Chem. 258:1235-1241
78 Walder R (1982) Fed. Proc. 41:A5425
78a Whitehead PR ad Brown NL (1983) FEBS lett. 155:97-101
79 Yoo OJ Dwyer-Hallquist P and Agarwal KL (1982) Nucleic Acids Res. 10:6511-652(
80 Youssoufian H and Mulder C (1981) J. Mol. Biol. 150:133-136
81 Youssoufian H Hammer SM Hirsch MS and Mulder C (1982) Proc. Natl. Acad. Sci. USA 79:2207-2210