A NEW SITE-SPECIFIC ENDONUCLEASE FROM *NEISSERIA CINEREA*

Robert WATSON, Michael ZUKER, Stan M. MARTIN, and Louis Peter VISENTIN
Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, K1A OR6, Canada

Received 16 June 1980

1. Introduction

Sequence-specific deoxyribonucleases have greatly facilitated the analysis and in vitro manipulation of DNA. Most of the known type II restriction enzymes recognize palindromic DNA sequences. Although a large number of these enzymes have been characterized, many of the possible palindromic DNA sequences are not recognized by any known enzyme (reviewed [1]). New restriction enzymes with unique recognition sites are desirable, as they increase the flexibility of recombinant DNA techniques.

We have screened a number of species of the genus *Neisseria* for restriction endonucleases, and report here the isolation and characterization of an enzyme from *Neisseria cinerea* which cleaves DNA at an unreported recognition sequence $5' \ldots CC(\frac{C}{G})GG \ldots 3'$. The identification of this sequence was assisted by a computer compilation of the number of each tetra-, penta- and hexa-palindromic base sequence in pBR322, ϕX174 and SV40 DNAs, which is presented here as an aid to other investigators.

2. Materials and methods

Restricted endonucleases *Bgl*I, *Hae*III, and *Hin*fI were obtained from Bethesda Research Labs. (MD) and DNA digestions with these enzymes were done using the suppliers suggested conditions. Simian virus 40 DNA, phage ϕX174 DNA, and adenovirus type 2 DNA were purchased from New England Biolabs. and Bethesda Research Labs. The derived plasmid pBR322 was prepared as in [2] from RR1 [pBR322GS] which was obtained from E. Lederberg, the Plasmid Reference Centre, Stanford (CA). Unmodified pBR322 was prepared from GM48 [pBR322] [3].

Address correspondence to L. P. V.

2.1. *Strain growth and enzyme purification*

Neisseria cinerea, NRCC strain 31006 was grown in BBL trypticase soy broth at 37°C for 31 h in a 7.5 l New Brunswick Scientific Microferm with aeration at 2.5 l/min, agitation at 300 rev./min with Antifoam (5 ml propylene glycol) added prior to sterilization. Cells were harvested in a Sharples centrifuge, mixed with an equal volume of 40% glycerol, 0.85 M NaCl, and frozen at −80°C until ready for use.

The enzyme *Nci*I was purified by a modified procedure of [4]. *N. cinerea* cells (5 g) were washed in 20 mM Tris–HCl (pH 7.5), 0.5 mM EDTA, and 6 mM 2-mercaptoethanol and resuspended in 10 ml same buffer. The cells were broken by five 30 s treatments with the small probe of a Branson sonifier Cell disruptor 185. The lysate was clarified by high speed centrifugation at 250 000 × *g* for 17 h and then applied to a heparin–agarose column prepared as in [4]. The column was developed with a 0–1.0 M NaCl gradient and fractions were assayed for the ability to cleave pBR322. The active fractions, which elute at 0.25 M NaCl, were dialyzed overnight against a buffer containing 5 mM KH_2PO_4, 5 mM 2-mercaptoethanol, 20 mM NaCl, 0.05 mM EDTA, 5 mM KCl, 5 mM $MgCl_2$ and 50% glycerol and stored at −20°C.

2.2. *Assay of* Nci*I activity*

Digestions with *Nci*I were done in 5–10 μl 6 mM Tris–HCl (pH 7.5), 6 mM $MgCl_2$, 6 mM NaCl, 6 mM 2-mercaptoethanol, and 100 μg/ml bovine serum albumin, containing 1.0 μg pBR322, SV40, phage ϕX174 or Ad-2 DNA and a 0.5–1.0 μl aliquot of enzyme fraction. After a 1 h incubation at 37°C digestion was terminated by incubation at 60°C for 15 min. One unit of the restriction endonuclease activity was defined as the amount sufficient to digest 1 μg pBR322 DNA in 1 h at 37°C. To digest DNA samples with a second restriction enzyme,

the *Nci*I digested DNA was first ethanol precipitated and resuspended in buffer appropriate for the second digestion.

2.3. *Gel electrophoresis*

Restriction fragments were separated on 1.0–1.5% agarose gels or 8% acrylamide gels containing 90 mM Tris, 90 mM boric acid, 25 mM EDTA (pH 8.3) by electrophoresis at 6–12 mA for 8–16 h. DNA bands were visualized as in [5] and their molecular weights determined by comparison of their mobilities with those of known pBR322 restriction fragments [6].

2.4. *Computer programs*

An IBM 3032 computer with an IBM TSS/370 time-sharing operating system was programmed with the complete nucleotide sequences of pBR322 [7], SV40 [8] and ϕX174 [9]. The ϕX174 sequence used was a corrected version prepared by Dr F. Sanger obtained from Dr R. Blakesley, Bethesda Research Labs. Details of the programs used to determine the numbers of recognition sites in these three DNAs (shown in table 1) or the cleavage coordinates and sizes of the fragments resulting from cleavage at these sites are available on request.

3. Results

Isolation of *Nci*I is relatively simple, involving a single heparin–agarose column fractionation (section 2). The enzyme elutes at 0.23–0.25 M NaCl and was judged to be free of contaminating non-specific nucleases as DNA samples digested with a 20-fold excess of the enzyme for 24 h gave sharp bands with no streaking on agarose gels. The yield of enzyme was $\sim 4 \times 10^4$ units/g wet wt cells, and it is active for up to 6 months stored at -20°C as judged by assay of its cleavage characteristics against plasmid pBR322 DNA. Maximal enzyme activity was observed at low ionic strength with inhibition of activity occurring at KCl and NaCl >0.15 M and >0.125 M, respectively. *Nci*I loses ~50% its activity on prolonged standing (24 h) at room temperature.

3.1. *Identification of the recognition site of* NciI

The pattern of DNA fragments observed on 8% acrylamide gels after electrophoresis of *Nci*I-digested pBR322 DNA is shown in fig.1(B,F). Although there is overlap of several bands, it is obvious that there

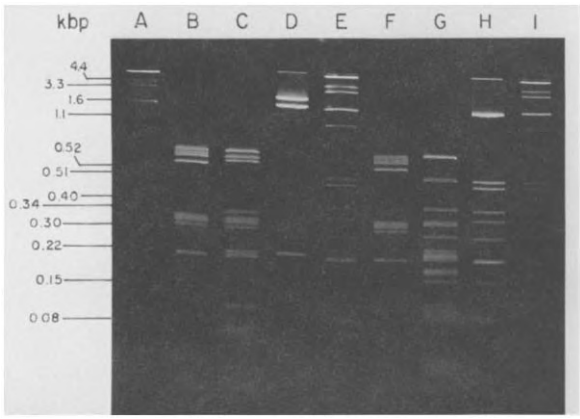

Fig.1. Polyacrylamide gel separation of the pBR322 fragments resulting from restriction endonuclease digestion with: (B,F) *Nci*I; (C) *Nci*I and *Bgl*I; (D) *Bgl*I; (G) *Nci*I and *Hin*fI; (H) *Hin*fI. The sizes of the molecular weight markers separated in (A), (E) and (I) are indicated at left in kilobase pairs (kbp). The markers are *Eco*RI, *Hin*cII and *Hin*fI fragments of pBR322, except the largest two fragments, which are linear pBR313 and ColEI-K30.

are ⩾8 *Nci*I cleavage sites in this molecule. As the identical pattern was observed following the same treatment of pBR322 purified from GM48[pBR322], *Nci*I is equally active against DNA modified by the adenine and cytosine methylase activities which are lacking in GM48 [2]. Digestion of ϕX174 DNA resulted in a conversion of its circular forms to a single linear molecule, indicating that it has one *Nci*I cleavage site. SV40 DNA was not susceptible to cleavage by this enzyme, and λ and Ad2 DNA were each cut into >15 pieces (not shown).

To assist in the determination of the nucleotide sequence which is recognized by *Nci*I, the complete nucleotide sequences of pBR322, ϕX174 and SV40 DNAs were scanned by computer to generate table 1. This table shows the number of cleavage sites expected in each of these species if they are cleaved at any of their simple 4–6 nucleotide palindromes, or at the recognition sequences of many known restriction enzymes which cleave at or near complex palindromic or non-palindromic sequences. This table is an extension of the data in [10,11] on the frequency of occurrence of nucleotide sequences in ϕX174 and SV40, although our search was limited to true palindromes.

Comparison of the number of cleavages produced by *Nci*I in pBR322, ϕX174 and SV40 DNA (>8, 1 and 0, respectively) with the data in table 1 shows

Table 1
Frequencies of tetra-, penta- and hexa-palindromic and related nucleotide sequences in pBR322, ϕX174 and SV40 DNAs[a]

Enzyme	Sequence	Frequency PBR322	PHX174	SV40	Enzyme	Sequence	Frequency PBR322	PHX174	SV40
	AATT	8	25	37		CASTG	11	6	22
	AANTT	10	31	33		ACATGT	1	0	0
	AARTT	4	17	18		CCATGG	0	0	3
	AASTT	6	14	15		GCATGC	1	0	2
	AAATTT	0	5	4		TCATGA	3	3	2
	CAATTG	0	1	3	HpaII	CCGG	26	5	1
EcoRI	GAATTC	1	0	1		CCNGG	16	3	16
	TAATTA	0	1	2	EcoRII	CCRGG	6	2	16
SacIII	ACGT	10	19	0	NciI	CCSGG	10	1	0
	ACNGT	14	15	25		ACCGGT	0	0	0
	ACRGT	8	7	25	SmaI	CCCGGG	0	0	0
	ACSGT	6	8	0		GCCGGC	4	0	1
	AACGTT	4	3	0		TCCGGA	1	0	0
	CACGTC	0	0	0	FnuDII	CGCG	22	14	0
	GACGTC	1	1	0		CGNCG	28	20	1
	TACGTA	0	0	0		CGRCG	9	8	0
AluI	AGCT	16	24	35		CGSCG	20	13	1
	AGNCT	9	20	19		ACGCGT	0	2	0
	AGRCT	2	6	3	SacII	CCGCGG	0	1	0
	AGSCT	7	14	16		GCGCGC	0	1	0
HindIII	AAGCTT	1	0	6		TCGCGA	1	2	0
PvuII	CAGCTG	1	0	3		CTAG	6	3	12
SacI	GAGCTC	0	0	0		CTNAG	8	14	19
	TAGCTA	0	0	2		CTRAG	1	4	5
	ATAT	8	21	22		CTSAG	7	10	14
	ATNAT	19	28	26		ACTAGT	0	0	0
	ATRAT	8	16	16	AvrII	CCTAGG	0	0	2
	ATSAT	12	14	13		GCTAGC	1	0	0
	AATATT	1	1	7	XbaI	TCTAGA	1	0	0
	CATATG	1	1	2	NboI	GATC	22	0	8
	GATATC	1	0	1	HinfI	GANTC	10	21	10
	TATATA	1	0	0		GARTC	6	11	5
	CATG	25	22	17		GASTC	4	10	5
	CANTG	18	16	31	BglII	AGATCT	0	0	0
	CARTG	7	10	9	PvuI	CGATCG	1	0	0
BamHI	GCATCC	1	0	1		TCNGA	15	18	9
BclI	TGATCA	0	0	1		TCRGA	8	14	9
NheI	GCGC	31	17	2		TCSGA	7	4	0
	GCNGC	35	30	22	ClaI	ATCGAT	1	0	0
BbvI	GCRGC	18	14	20	XhoI	CTCGAG	0	1	0
	GCSGC	19	16	2	SalI	GTCGAC	1	0	0
	AGCGCT	4	0	1		TTCGAA	0	0	0
	CGCGCG	1	0	0		TGCA	21	18	36
	GGCGCC	4	2	0		TGNCA	16	23	18
MstI	TGCGCA	4	1	0		TGRCA	7	10	10
HaeIII	GGCC	22	11	19		TGSCA	9	13	8
AsuI	GGNCC	15	2	10	AvaIII	ATGCAT	0	0	3
AvaII	GGRCC	8	1	6	PstI	CTGCAG	1	1	2
	GGSCC	7	1	4		GTGCAC	3	1	0
	AGGCCT	0	1	7		TTGCAA	1	2	3
	CGGCCG	1	0	0		TTAA	15	35	47
	GGGCCC	0	0	1		TTNAA	14	32	37
BalI	TGGCCA	1	0	0		TTRAA	6	16	28
	GTAC	3	11	11		TTSAA	8	16	9
	GTNAC	17	17	14		ATTAAT	1	2	4
	GTRAC	8	9	8		CTTAAG	0	2	1
	GTSAC	9	8	6	HpaI	GTTAAC	0	3	4
	AGTACT	1	0	0		TTTAAA	3	2	11
	CGTACG	0	2	0	AvaI	CYCGXG	1	1	0
KpnI	GGTACC	0	0	1	HaeII	XGCGCY	11	8	1
	TGTACA	0	0	2	HincII	GTYYAC	2	13	7
	TATA	7	11	18	HaeI	RGCGCR	7	6	11
	TANTA	8	19	12	AccI	GTMZAC	2	2	1
	TARTA	7	14	9	HgiAI	GRGCRC	8	3	0
	TASTA	1	5	4	MnlI	CCTC	26	34	51
	ATATAT	1	0	0	MboII	GAAGA	11	11	16
	CTATAG	0	1	2	HgaI	GACGC	11	14	0
	GTATAC	1	0	0	HphI	GGTGA	12	9	4
	TTATAA	0	1	3	BglI	GCC(N5)GGC	3	0	1
TaqI	TCGA	7	10	1	EcaI	GGTNACC	0	0	0
					AcyI	GXCGYC	6	7	0

[a] The 16 possible tetranucleotide palindromes are listed alphabetically under the 'sequence' column. These are underlined. Grouped under each underlined tetranucleotide are the penta and hexanucleotide palindromes derived as follows: Using GATC as an example, the pentanucleotides GARTC, GASTC and GANTC are derived by inclusion of a central nucleotide, where R = A or T, S = G or C and N is any nucleotide. The four hexanucleotide palindromes are produced by adding A and T, C and G, G and C or T and A to the ends. Known enzymes are named under the enzyme column. Enzymes recognizing non-palindromic sequences are included at the end of the table, where W = A or C, X = A or G, Y = C or T and Z = G or T. In the latter cases only the sequence from one DNA strand is shown

that 5′...CC(C/G)GG...3′ is the only candidate for the recognition sequence of the enzyme. This sequence occurs 10, 1 and 0 times in pBR322, ϕX174 and SV40 DNAs, respectively. Of course this finding does not exclude the possibility that the site is a complex palindrome or non-palindromic sequence.

To test the possibility that *Nci*I recognizes 5′...CC(C/G)GG...3′ the sizes of the pBR322 fragments which would be produced by cleavage at these sites alone or in combination with restriction endonucleases *Bgl*I or *Hin*fI were determined by computer. These data, together with the positions of the sites, is shown in table 2. The predicted fragment sizes were compared with those produced by *Nci*I digestion of pBR322 and by double digests of this plasmid with *Nci*I and *Bgl*I, and with *Nci*I and *Hin*fI (fig.1). Although several of the bands overlap in the separations of the *Nci*I-generated pBR322 fragments in fig.1(B,F), it is obvious that the fragment sizes given in table 2 are in close agreement with this experimental data. The fragments produced by the double digestion (fig.1(C,G)) also confirm that 5′...CC(C/G)GG...3′ is recognized by *Nci*I, as their sizes as those predicted in table 2. These data is especially convincing as it

Table 2
Computer generated digest patterns of plasmid pBR322 with restriction enzymes *Nci*I, *Hin*fI, *Bgl*I (size of fragments in kilobase pairs)

NciI*	NciI + BglI	BglI	HinfI + NciI	HinfI
		2.3190	0.6320	1.6310
		1.8090	0.5075	0.5170
0.7240	0.6990	0.2340	0.3940	0.5060
0.6990	0.6325		0.3630	0.9360
0.6960	0.6320		0.3510	0.3440
0.6320	0.3995		0.2885	0.2980
0.3630	0.3630		0.2535	0.2210
0.3510	0.3510		0.2200	0.2200
0.3280	0.3280		0.2185	0.1540
0.3080	0.3080		0.2175	0.0750
0.2260	0.2340		0.1885	
0.0350	0.2260		0.1815	
	0.0905		0.1540	
	0.0635		0.0965	
	0.0350		0.0905	
			0.0750	
			0.0445	
			0.0395	
			0.0350	
			0.0095	

Computer-predicted cleavage sizes of the pBR322 fragments resulting from *Nci*I digestion alone or in combination with *Bgl*I or *Hin*fI. Underlined sizes indicated new fragments generated by double digestion. **Nci*I cleavages in pBR322 were taken to be at the centre of the palindrome 5′...CC(C/G)GG...3′ recognition sequences at positions: 171.5; 534.5; 1258.5; 1484.5; 1812.5; 2120.5; 2155.5; 2854.5; 3550.5; 3901.5

relates the postulated locations of the 10 *Nci*I recognition sites within the pBR322 sequence to 13 sites whose locations are known.

One possibility which is not excluded by the above analysis is that the proposed recognition sequence is in some way related to a true recognition sequence. For example, if all of the 5′ . . . CC($\frac{C}{G}$)GG . . . 3′ sites in pBR322 and ϕX174 were 5′ . . . CCCGG . . . 3′ or 5′ . . . ACC($\frac{C}{G}$)GGT . . . 3′, then our conclusions could be in error. Obvious possibilities such as these examples have been eliminated by inspection of the nucleotide sequences in the regions of the proposed sites in pBR322 and ϕX174. In particular, both 5′ . . . CCCGG . . . 3′ and 5′ . . . CCGGG . . . 3′ are represented in these DNAs and we detect no similarities in these regions other than these pentanucleotide sequences.

4. Discussion

Computer analysis of known DNA sequences to predict the number and positions of putative restriction endonuclease sites and the expected sizes of restriction fragments is a significant improvement over conventional sequencing techniques as a means of determining sequence specificity of new restriction enzymes. This procedure is dependant upon the availability of pure DNA species of known nucleotide sequences and has been applied [10] for the identification of the *Ava*II recognition site using ϕX174 and SV40 DNAs. Since the plasmid pBR322 is easily purified and has been completely sequenced [7], we have extended our computer analysis to include this species. This addition allows the prediction of recognition sites which are poorly represented in ϕX174 and SV40 DNA yet are more abundant in pBR322, as is the case for those of *Nci*I. The data in table 2 should prove valuable for the identification of other new restriction enzyme recognition sites, and the pBR322 sequence marked with the positions of the sequences shown in table 2 is available upon request.

5′ . . . CC($\frac{C}{G}$)GG . . . 3′ is a new addition to the list of nucleotide sequences which are known to be recognized by restriction enzymes. *Nci*I should cleave DNA at all *Sma*I sites (5′ . . . CCCGGG . . . 3′), and should cleave a subset of *Hpa*II recognition sites (5′ . . . CCGG . . . 3′). The *Nci*I recognition site also resembles that of *Eco*RII (5′ . . . CC($\frac{A}{T}$)GG . . . 3′).

Acknowledgements

We thank Wally Rowsome, Gilles Masson, Roger Latta, Lynda Boucher and Pat Clay for their skilled technical assistance and Harry Turner for the excellent photographic contributions. The authors are indebted to Dr Robert Blakesley of Bethesda Research Labs. for his helpful suggestions. R. W. is a NRCC research associate.

References

[1] Roberts, R. J. (1980) Nucleic Acids Res. 8 r63–r80.
[2] Guerry, P., LeBlanc, D. J. and Falkow, S. (1973) J. Bacteriol. 116, 1064–1066.
[3] Marinus, M. G. and Morris, R. (1975) Mut. Res. 28, 15–26.
[4] Bickle, T. A., Pirrotta, V. and Imber, R. (1977) Nucleic Acids Res. 4, 2561–2572.
[5] Sharp, P. A., Sugden, B. and Sambrook, J. (1973) Biochemistry 12, 3055–3063.
[6] Sutcliffe, J. G. (1978) Nucleic Acid Res. 5, 2721–2728.
[7] Sutcliffe, J. G. (1979) Cold Spring Harbor Symp. 43, 77–90.
[8] Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Sain, B. S., Pan, J., Ghosh, P. K., Cellma, M. L. and Weissman, S. M. (1978) Science 200, 494–502.
[9] Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchinson, C. A., Slocombe, P. M. and Smith, M. (1977) Nature 265, 687–695.
[10] Fuchs, C., Rosenvold, E. C., Honigman, A. and Szybalski, W. (1978) Gene 4, 1–23.
[11] Blakesley, R. (1978) Focus 1 (4) Bethesda Research Labs., MD.