*Fse*I, a new type II restriction endonuclease that recognizes the octanucleotide sequence 5′ GGCCGGCC 3′

Janise Meyertons Nelson+, Sheila M.Miceli, Mary P.Lechevalier[1] and Richard J.Roberts*
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724 and [1]Waksman Institute, Rutgers University, Piscataway, NJ 08855, USA

Received December 28, 1989; Revised and Accepted February 26, 1990

ABSTRACT

A Type II restriction endonuclease, designated *Fse*I, has been partially purified from a *Frankia* species (NRRL 18528). This enzyme cleaves Adenovirus 2 DNA at three sites, but does not cleave the DNAs from bacteriophages lambda, T7, and φX174, the animal virus SV40, pUC18 and pBR322. *Fse*I recognizes the octanucleotide sequence 5′ GGCCGG↓CC 3′ and cleaves as indicated by the arrow. The frequency of occurrence of *Fse*I sites within sequenced regions of the human genome is similar to that for *Not*I sites.

INTRODUCTION

Many Type II restriction enzymes have been isolated from unusual non-streptomycete members of the Actinomycetales family (1). In general, bacteria within this family have a high G+C content in their DNA ranging between 60 and 75 mol% G+C. McClelland (2) has suggested that the G+C content of the recognition sequence of Type II restriction enzymes reflects the G+C content of the bacterial genome encoding them. Therefore enzymes with G+C rich recognition sequences of six or more nucleotides in length are generally found in bacterial species with G+C contents of at least 60% (2). In particular, the only two enzymes previously reported to recognize octanucleotide sequences were found in a *Nocardia* strain and a *Streptomyces* strain, both members of the high G+C family Actinomycetales (3,4). With this in mind, species from the actinomycete genus *Frankia* were screened for new Type II restriction enzymes. Microorganisms from the genus *Frankia* are dinitrogen-fixing, root-nodule symbionts of many nonleguminous plants, including *Myrica*, *Alnus*, *Casuarina* and other species (5,6). We now report the isolation from a *Frankia* species of a new Type II restriction enzyme, *Fse*I. It has been partially purified and shown to possess an octanucleotide recognition sequence.

MATERIALS AND METHODS

DNA, enzymes and chemicals

Adenovirus 2 (Ad2) DNA was prepared as described previously (7). Bacteriophage λ and φX174 DNAs were obtained from New England Biolabs; SV40 DNA was from Bethesda Research Laboratories; Bacteriophage T7 DNA and pUC18 and pBR322 plasmid DNAs were prepared by standard procedures. Oligonucleotides were synthesized in the Cold Spring Harbor oligonucleotide synthesis facility. Restriction endonucleases and T4 polynucleotide kinase were obtained from New England Biolabs. Calf intestinal alkaline phosphatase was obtained from Boehringer-Mannheim. The Klenow fragment of *E. coli* DNA polymerase I was obtained from Bethesda Research Laboratories. Enzymes were used according to the manufacturer's specifications. ^{35}S-α-dATP (>1000 Ci/mmole) was from New England Nuclear and ^{32}P-α-dATP (>2000 Ci/mmole) was from ICN. All other chemicals were of reagent grade quality.

Growth of *Frankia* species

Frankia species (NRRL 18528) was grown at 28°C (range 24°−33°C) under static conditions for 21−28 days in 100 ml B/2 broth contained in a 250 ml Erlenmeyer flask. The components of B/2 are 1 g NZ Amine type A (Sheffield Products, Kraft), 0.45 g Lab Lemco (Oxoid), 0.5 g yeast extract (Difco), 5 g dextrose (Mallinckrodt) and tap water to a final volume of one liter. The medium was adjusted to pH 7.3 with NaOH before autoclaving. The cell growth was dispersed prior to inoculation by passing the cell suspension through an 18 gauge needle. Approximately 10g wet weight of cells per liter were obtained.

The cell mass was harvested by centrifugation, washed once in 1M NaCl in 10 mM Tris-HCl, pH 7.6, pelleted again, washed in 10 mM Tris-HCl, pH 7.6 and resuspended in 50% glycerol, 10 mM Tris-HCl, 1 mM Na_2EDTA pH 8.0. The cells were stored frozen at −70°C.

Purification of *Fse*I

The enzyme isolation procedure was based on a method outlined in detail previously (8). Briefly, frozen cells (10−13 g) were thawed at room temperature and collected by centrifugation at 4°C. The cells were maintained at 4°C for the remainder of the enzyme isolation procedure. The cell mass was resuspended in Buffer A (20 mM Tris-HCl, pH 7.6, 1 mM Na_2EDTA, pH 8.0, 10% (v/v) glycerol and 10 mM β-mercaptoethanol). The cells were disrupted after two passages through a French pressure cell maintained at 1300−1500 psi. A solution of 20% streptomycin

* To whom correspondence should be addressed

+ Present address: Abbott Laboratories, Department 451, Building R5, Sheridan Road, N. Chicago, IL 60064, USA

sulfate was then added to the supernatant to a final concentration of 2% to precipitate RNA and DNA. The supernatant was clarified by centrifugation at 12,000×g and dialyzed against a 1000-fold excess volume of Buffer B (10 mM potassium phosphate, pH 7.4, 100 mM Na_2EDTA, 10% (v/v) glycerol, 10 mM β-mercaptoethanol).

The crude extract was loaded onto a 1.0 cm×10 cm DEAE-cellulose (Whatman DE52) column equilibrated with Buffer B. The restriction enzyme was eluted from the column with a 30 ml linear gradient of 0−1.0 M KCl in Buffer B. One ml fractions were collected and assayed for activity by observing digestion of Ad2 DNA.

The active fractions from the DEAE-cellulose column were pooled and dialyzed against Buffer A. Further purification was achieved using the Mono-Q anion exchange column on the Pharmacia FPLC system. Active enzyme was eluted from the column using a 30 ml linear gradient of 0−1.0 M KCl in Buffer A. One ml fractions were collected and assayed for activity by observing digestion of Ad2 DNA. The active fractions were pooled, mixed to a final concentration of 50% (v/v) glycerol and stored at −20°C.

Assay Conditions for *FseI*

Suitable dilutions of the crude extract or column purified enzyme were incubated with Ad2 DNA in medium salt restriction buffer (10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 50 mM NaCl, 10 mM β-mercaptoethanol, 100 μg/ml BSA) or 1×KGB (9) at 26°−28°C from one hour to overnight. These conditions were found to be optimal among a wide variety of salts, pH and temperatures that were tested. The reactions were terminated by the addition of loading dye (0.25% Bromophenol blue, 15% Ficoll type 400, 50 mM Na_2EDTA) and by heating the reaction to 65°C for five minutes. The DNA fragments were separated by electrophoresis at 80 volts for several hours in a 0.8% agarose gel containing 1 μg/ml ethidium bromide.

Characterization of the *FseI* cleavage site

The primed-synthesis reaction was used to characterize the *FseI* cleavage site (10). An M13 clone containing an *FseI* recognition site from a segment of the Ad2 genome (clone 1072, nucleotides 10925 to 11377) (11) was used as the template. The M13 single-stranded DNA template was incubated with γ-^{32}P end-labelled universal sequencing primer (GTTTTCCCAGTCACGAC, New England Biolabs), the four deoxynucleotides and modified T7 DNA Polymerase (Sequenase, version 2.0 kit, United States Biochemicals) for 10 minutes to extend the primer beyond the recognition site. The polymerase reaction was inactivated by heat treatment at 70°C, and then the reaction was incubated with the restriction enzyme *FseI*. This reaction was then divided in two; one half of the sample was incubated further with DNA polymerase I Klenow Fragment (New England Biolabs) plus the four deoxynucleoside triphosphates and the other half received no treatment. The reactions were electrophoresed on an 8% denaturing DNA sequencing gel adjacent to the dideoxynucleotide DNA sequencing reactions of the template.

RESULTS

Crude extracts of the *Frankia* species revealed the presence of an endonucleolytic activity that could degrade Ad2 DNA but was inactive on all other small DNAs tested. The activity was purified extensively by DEAE-cellulose and FPLC chromatography.

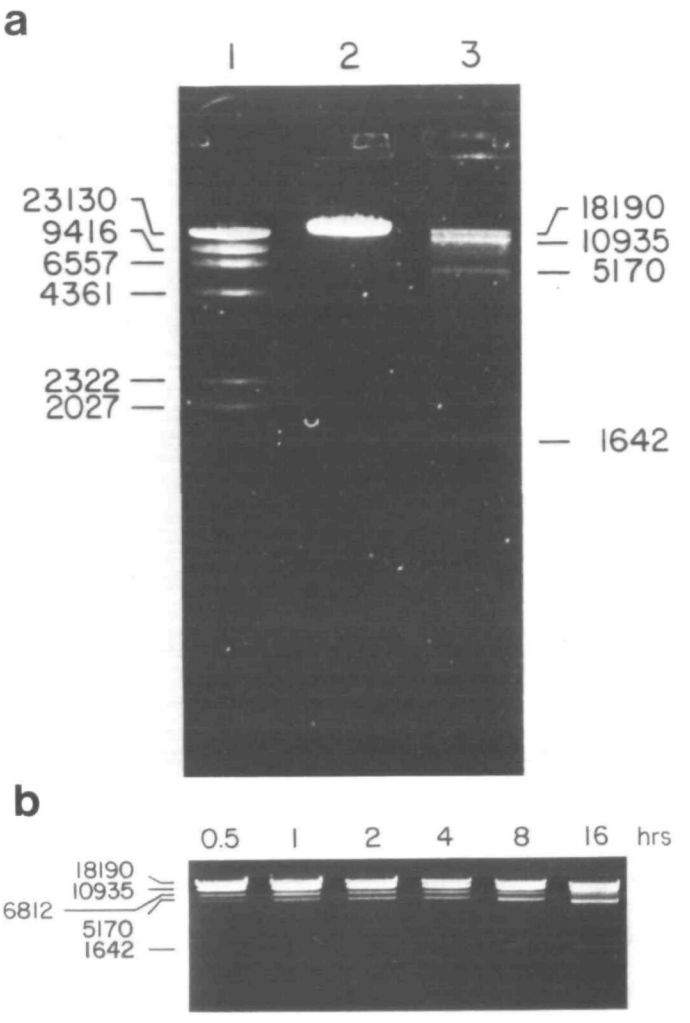

Figure 1: a) *FseI* recognizes three sites on the Ad2 genome producing fragments of 18190, 10935, 5170 and 1642 basepairs in length. Lane 1: Size Markers. Bacteriophage λ DNA digested with *Hin*dIII. Lane 2: Uncut Ad2 DNA (35937 basepairs total length). Lane 3: Ad2 DNA digested with *FseI*. b) Time course of *FseI* digestion. 0.5 μg Ad2 DNA was digested for the times indicated above each lane. The partial digestion product of 6812 nucleotides that results when the site at 12577 remains uncleaved is indicated.

Approximately 250 units of *FseI* activity were recovered per gram wet weight of cells. One unit of activity is defined as the amount necessary to digest one μg of Ad2 DNA to completion in one hour. Digestion was linear for at least 16 hrs and so usually digests were carried out for several hours to conserve enzyme. The purified enzyme was substantially free of contaminating non-specific nucleases as judged by the stability of the digestion pattern after 10-fold overdigestion with *FseI*.

Figure 1a shows an *FseI* digest of Ad2 DNA. *FseI* recognized three sites on the Ad2 genome producing four fragments of 18190, 10935, 5170 and 1642 nucleotide pairs in length. Preliminary mapping experiments with *FseI* on Ad2 and double-digestion of Ad2 with *FseI* and *Bam*HI, *Hpa*I, *Kpn*I and *Not*I showed that the three *FseI* cleavage sites were localized approximately to nucleotide positions 10950, 12600 and 17750 on the Ad2 genome. It should be noted that the cleavage site in Ad2 DNA at nucleotide 12600 is kinetically very slow (Figure 1b). At least a 10-fold excess of *FseI* is required to completely digest this site as compared with the other two sites. Examination

```
10935    GACCCCCGGTTCGAGTCTCG GGCCGGCC GGACTGCGGCGAACGGGGGT
12577    AGGGCCATCCGGCCCGATGA GGCCGGCC TGGTCTACGACGCGCTGCTT
17747    ATGCACCGTAGGAGGGGCAT GGCCGGCC ACGGCCTGACGGGCGGCATG
```

Figure 2: The three *Fse*I sites in Ad2 DNA. The sequences are taken from reference 15.

of the sequences of the Ad2 genome in the vicinity of these coordinates revealed that the octanucleotide sequence GGCCGGCC was present at all three positions (Figure 2). These were the only occurrences of this sequence within the Ad2 genome. Furthermore the flanking sequences in Ad2 showed no further similarities that might be consistent with an even longer recognition sequence. Inspection of the sequences of bacteriophages lambda and T7 DNAs showed that GGCCGGCC was not present, consistent with the finding that *Fse*I does not cut either of these DNAs. Similarly GGCCGGCC was not found in the DNA sequences of SV40, pBR322, pUC19, or ϕX174, all of which were refractory to cleavage by *Fse*I.

Based on the above mapping and computer analysis experiments we considered that the recognition sequence for *Fse*I was likely to be GGCCGGCC. However, formally some degenerate version of this sequence, such as GGCCGGCY, might also be a possibility. We therefore examined the sequences of each DNA tested above for the presence of sequences that differed from GGCCGGCC at a single position. With one exception, each possibility occurs at least once within one or more of these sequences. The exception is the sequence CGCCGGCC (or its complement GGCCGGCG). If this sequence were to be a recognition site, then the general form of the recognition sequence would be SGCCGGCC (S is the IUPAC degeneracy code for G or C). Such a sequence would be an unlikely candidate for a restriction enzyme recognition site, based on the patterns known to be recognized (1). The more plausible SGCCGGCS, in which the symmetry is maintained, can be excluded as a site because the specific sub-sequence CGCCGGCG occurs at positions 809 and 26429 in the Ad2 genome. No cleavage could be detected at these positions. Because there is no obvious commonality within the flanking sequences surrounding the three known sites in Ad2 DNA (Figure 2) we conclude that the true recognition sequence for *Fse*I is GGCCGGCC.

To characterize the precise site of cleavage within the recognition sequence we took advantage of an M13 clone, 1072, that had been isolated previously during our determination of the sequence of the Ad2 genome (11). This clone contains nucleotides 10925 to 11377 from the Ad2 genome and includes an *Fse*I recognition site. The autoradiograph of the primed-synthesis reactions used to characterize the cleavage site for *Fse*I is shown in Figure 3. Lane 1 shows the results of a primed-synthesis reaction cleaved with *Fse*I. This sample produced a single band. When compared with the sequencing lanes this band can be seen to comigrate with the sixth nucleotide in the recognition sequence 5′ GGCCGG|CC 3′. This result indicates that cleavage of the DNA, within the newly synthesized strand, occurred as shown by the arrow. Lane 2 shows the result obtained when the primed-synthesis reaction from Lane 1 is further incubated with the Klenow fragment of DNA Polymerase I. During this treatment the 3′-terminal extension present on the newly synthesized strand is resected by the exonuclease action of the polymerase until a blunt end is formed. From the resulting product, the position of cleavage of the template strand can be inferred. This sample

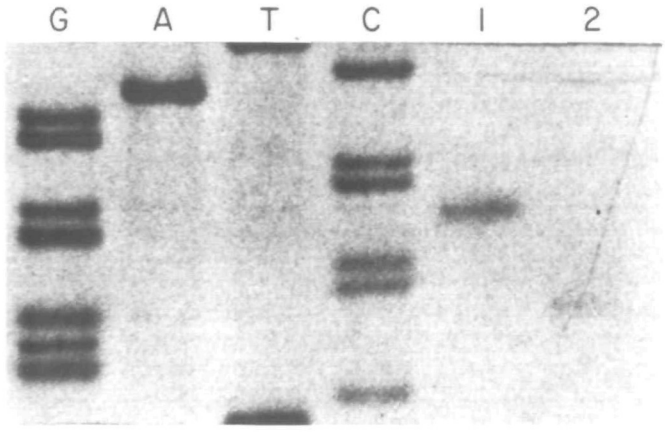

Figure 3: Characterization of the *Fse*I cleavage site. Shown is the autoradiograph of the primed-synthesis reaction used to characterize the cleavage site for *Fse*I. Lanes G, A, T and C contain the standard sequencing reactions through the *Fse*I recognition sequence, using the chain termination method (16). Lane 1: The primed-synthesis reaction was cleaved with *Fse*I. The resulting single-band indicated that DNA cleavage occurred within the recognition site 5′ GGCCGG^CC 3′, as indicated by the arrow. Lane 2: The primed-synthesis reaction from Lane 1 (cleaved with *Fse*I) was incubated with the Klenow fragment of DNA Polymerase I. The result indicated that *Fse*I cleaved symmetrically to produce a four-base 3′ extension.

produced a band that comigrated with the second base of the recognition sequence. This result indicates that *Fse*I cleaved the DNA to produce a four-base 3′ extension. The recognition sequence and cleavage site are thus:

5′ G G C C G G|C C 3′
3′ C C|G G C C G G 5′

We tested the ability of *Fse*I to cleave substrates in which various cytosine residues within the recognition sequence were modified to 5-methylcytosine. For these experiments we used both pFse945, a pBR322-based plasmid constructed to contain the *Fse*I recognition site, and a double-stranded oligonucleotide: 5′ TATTTTGGCCGGCCTTAGTT 3′. The results of these experiments showed that methylation by M.*Msp*I, which would produce the methylated sequence 5′ GGmCCGGCC 3′ (12), M.*Hae*III, which would produce 5′ GGmCCGGmCC 3′ (13) and M.*Hpa*II, which would produce 5′ GGCmCGGCC 3′ (13) inhibited cleavage by *Fse*I (data not shown).

DISCUSSION

We have isolated and characterized an enzyme from a *Frankia* species which recognizes the octanucleotide sequence 5′ GGCCGGCC 3′. This enzyme has been named *Fse*I. This is only the third among more than 1300 Type II restriction enzymes isolated thus far that recognizes an octanucleotide sequence (1).

The other two enzymes that recognize octanucleotide sequences are *Not*I from *Nocardia otitidis-caviarum*, recognition sequence GC↓GGCCGC (3) and *Sfi*I from *Streptomyces fimbriatus*, recognition sequence GGCCNNNN↓NGGCC (3,4). It is curious that in all three cases the recognition sequences are composed entirely of guanosine and cytosine residues. This would accord well with the hypothesis of McClelland (2), that such recognition sequences are only expected in organisms with a high G+C content. Nevertheless it is surprising that the first three enzymes isolated with these long recognition sequences are not more diverse in character. In none of the three cases so far known has it been established whether these enzymes are involved in restriction-modification *in vivo*. It will be of great interest to do so.

*Not*I and *Sfi*I have recently found great use in the mapping of large genomes because they produce very large fragments that can readily be resolved by pulsed-field gel electrophoresis (14). *Fse*I potentially represents a valuable addition to the repertoire of enzymes available for mapping large genomes. 157 sites for *Fse*I occur within the human genomic sequences present in GenBank version 61. This number is comparable to the 182 sites for *Not*I and is considerably less than the 285 sites for *Sfi*I. Unfortunately at present the yields of *Fse*I are rather poor and so quantities of the enzyme are limited. It will thus be very important that the gene for the enzyme is cloned so as to facilitate its production in large quantities. Such experiments are in progress.

ACKNOWLEDGEMENTS

We thank J. Duffy and P. Renna for help with the artwork and photography. This work was supported by a grant from the National Institutes of Health (GM40537 to RJR) and the Charles and Johanna Busch Fund (MPL).

REFERENCES

1. Roberts, R.J. (1989) Nucl. Acids Res. 17: r347–r387.
2. McClelland, M. (1988) Nucl. Acids Res. 16: 2283–2294.
3. Qiang, B.Q. and Schildkraut, I. (1987) Methods Enzymol. 155: 15–21.
4. Qiang, B.Q. and Schildkraut, I. (1984) Nucl. Acids Res. 12: 4507–4515.
5. Becking, J.-H. (1981) in *The Prokaryotes*, Vol. II (M. Starr, H. Stolp, H.G. Truper, A. Balows, H.G. Schlegel, eds.) Springer-Verlag pp. 1991–2003.
6. Lechevalier, M. P. (1983). Can. J. Botany 61: 2964–2967.
7. Pettersson, U. and Sambrook, J. (1973) J. Mol. Biol. 73: 125–130.
8. Meyertons, J.L., Tilley, B.C., Lechevalier, M.P. and Lechevalier, H.A. (1987). J. Ind. Microbiol. 2: 293–303.
9. McClelland, M., Hanish, J., Nelson, M. and Patel, Y. (1988). Nucl. Acids Res. 16:364.
10. Brown, N.L., Hutchison, C.A. and Smith, M. (1980) J. Mol. Biol. 140: 143–148.
11. Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.O. and Roberts, R.J. (1982) J. Biol. Chem. 257: 13475–13491.
12. Walder, R.Y., Langtimm, C.J., Catterjee, R. and Walder, J.A. (1983) J. Biol. Chem. 258: 1235–1241.
13. Mann, M.B. and Smith, H.O. (1977) Nucl. Acids Res. 4: 4211–4221.
14. Schwartz, D.C. and Cantor, C.R. (1984) Cell 37: 67–75.
15. Roberts, R.J., Akusjarvi, G., Alestrom, P., Gelinas, R.E., Gingeras, T.R., Sciaky D. and Pettersson, U. (1986) in 'Adenovirus DNA. The Viral Genome and Its Expression', ed. W. Doerfler, Martinus Nijhoff, Boston, Mass. pp 1–51.
16. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74: 5463–5467.