StyD4I Restriction-Modification System of *Salmonella* Typhi D4: Cloning and Sequence Analysis

Michiko MIYAHARA,^{*,a} Naoko ISHIWATA,^b and Yoko YOSHIDA^c

National Institute of Health Sciences,^a 1–18–1 Kamiyoga, Setagaya-ku, Tokyo 158, Japan, Showa Woman's University,^b 1–7 Taishido, Setagaya-ku, Tokyo 154, Japan and The Institute of Public Health,^c 4–6–1 Shirokane-dai, Minato-ku, Tokyo 108, Japan. Received July 22, 1996; accepted October 14, 1996

A plasmid (5.4 kbp) from *Salmonella* Typhi D4 has been identified as encoding a restriction and modification (R-M) system. DNA fragments (2537 bp) that carried the genes for restriction endonuclease and methyltransferase encoded on the plasmid were sequenced. Two divergently arranged open reading frames of 957 bp for the restriction endonuclease consisting of 318 aa (amino acids) and 1140 bp for the DNA methyltransferase consisting of 379 aa were identified. These sequences were similar to the sequences of the *SsoII* R-M system, including the interspace between the two genes.

Key words Restriction-modification system; sequence analysis; *StyD4I*; pSTd4; *Salmonella* Typhi D4; deletion mutant

Previously, we reported on *StyD4I* restriction endonuclease (ENase) of *Salmonella* Typhi D4 (*S. Typhi* D4).^{1–2)} *S. Typhi* D4 has a plasmid, named pSTd4 (5.4 kbp), which encodes ENase recognizing the CCNGG DNA sequence and the related modification methyltransferase (MTase). The methylase protect DNA against the attack of *StyD4I* ENase. The *StyD4I* methylase is classified as a DNA (cytosine-5)methyltransferase (m5C-MTase). The m5C-MTases methylate the 5-carbon of the pyrimidine ring of cytosine. They share a large set of well-conserved blocks of amino acid sequences, which are called motifs.³⁾ Many isoschizomers have been reported.⁴⁾ Of these, *DsaV*, *BssKI* and *SsoII* were reported to have the same cleavage site as *StyD4I*. The cleavage site of *StyD4I* had ↓ CCNGG the same as *DsaV*⁵⁾ and *SsoII*,⁶⁾ but not the same as the CC ↓ NGG of *ScrFI*.⁷⁾ For comparison with the sequences of other isoschizomers, the cloning and sequence analysis of the *StyD4I* restriction-modification (R-M) system have been carried out.

MATERIALS AND METHODS

Previously, a restriction map of pSTd4 (5400 base pairs (bp)) was reported.²⁾ According to the map, *EcoRI* fragments were selected for cloning.

We obtained five kinds of *EcoRI* fragments. These fragments of pSTd4 were inserted into *EcoRI*-digested dephosphorylated pUC118, with T4 DNA ligase. The ligated DNA was used for the transformation of *E. coli* WA803, and ampicillin-resistant(Ap^R) transformants were obtained. pUC118 has twelve *StyD4I* sites. We tested the activity of ENase and MTase. The activity of ENase was measured using the methods described before.¹⁾ Also, a cross-streak method (as follows) was used as a highly sensitive method. On a broth plate, a phage solution (λ clts 857 or T3) was painted in a straight belt. Across the phage belt, bacteria solution was also painted. After culture overnight, bacteria having ENase grew but bacteria without ENase did not at the site across the phage belt. MTase activity was indirectly detected. After isolation of the plasmid, each plasmid was cleaved by *StyD4I*.

* To whom correspondence should be addressed.

Plasmid which resisted cleavage by *StyD4I* was estimated to encode MTase. Plasmid DNA was isolated by the alkaline lysis method.⁸⁾ The location of ENase and MTase genes on pSTd4 was estimated.

Deletion mutants were made by the Deletion Kit for Kilo-Sequencing (Takara). The cleaving of a *BamHI*-*PstI* fragment from the starting DNA started the deletion reaction. Sequence deletion was constructed by Exonuclease III according to the manufacturer's instruction. Many mutants, with different sized plasmid, were tested for their activity of ENase and MTase. *E. coli* WA803 was used as the host. These deletion mutants were also used for sequencing.

Insertion mutants were used for the search for the ENase gene on pSTd4. pKDEC100 (pHSG415CM::Tn3)^{9,10)} was used as a transposon-carrying and temperature-sensitive plasmid. Competent cells of WA803 carrying pSTd4 were transformed by the plasmid. These transformants were cultured on the ampicillin (100 μ g/ml) and chloramphenicol (25 μ g/ml)-containing plates at 30 °C overnight. After this selection, further culture was accomplished on the ampicillin-containing (at the first time 200 and at the second 100 μ g/ml) plate at 42 °C overnight. Transposon Tn3 has an ampicillin resistant gene. By these treatments, cells having a Tn3 inserted genome or plasmid can survive on the ampicillin-containing plate.

The molecular weight of ENase, *StyD4I* was measured by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and gel filtration on a Superose 12 (Pharmacia, HR 10/30, FPLC).

For the sequencing work, *E. coli* MV1184 was used with a helper phage, M13K07. Single DNA was obtained from a transformant of MV1184. Of all the sequences of pSTd4, 2537 sequences were determined by the dideoxynucleotide chain-termination method.¹¹⁾

Search homology was accomplished by GENETYX-MAC (Software Development Co., Ltd., Tokyo).

RESULTS AND DISCUSSION

EcoRI fragment-inserted clones lost their ENase activities. The ENase gene was recognized to be over the

Fig. 1. ENase and MTase Activities of Deletion Derivatives

Activities were assayed by the methods described in Materials and Methods. +, detectable level; -, no detectable activity. The ligation site was shown as a line in the figure of the clone (5400-1 and 5400-2). The big column shows the estimated location of the MTase gene on pSTd4.

Fig. 2. ENase and MTase Activities of Insertion Mutants of pSTd4

Small arrows show the Tn3-inserted site on pSTd4. The big column shows the estimated location of the ENase gene on pSTd4 and the sequencing area.

EcoRI site of pSTd4. MTase activities remained on the 1—1736 bp position on the clone (5400-1 *etc.*), as shown in Fig. 1. Analysis on the insertion of mutants also showed these data (Fig. 2). Three insertion mutants (which showed negative activities of ENase) were obtained. The ENase gene covers the *EcoRI* sites (2 and 3). The gene area for the *StyD4I* R-M system seemed to be in the 2537 bp of pSTd4. After sequencing, the 2537 bp area was searched for its amino acid sequences. Two divergently arranged open reading frames were found, as shown in Fig. 3. These DNA and amino acid sequence data (*StyD4I* R-M

system) have been deposited in a GenBank database under accession No. D73442. By the gel filtration method, the molecular weight (MW) of *StyD4I* was estimated to be 70000 Da. On the other hand, on SDS-PAGE we estimated the MW to be 35000 Da. From these data, we concluded that *StyD4I* might act as a dimer and the MW of *StyD4I* (its monomer MW weight) might be 35000 Da. These data coincided with the MW of ENase calculated on open reading frames.

The 2537 bp DNA sequences show 98.6% homology to the area (2648 bp) for the gene of the *SsoII* R-M

10	20	30	40	50	60
CGCTTCTTGC	TGGATGCTG	AAACATGAA	AAATGGGAC	TTGATGGAT	TAAAGAGGGT
GGCAAGAGG	ACCGTACAC	TTTGTACTT	TTTACGCTG	AATGACGTA	ATTCTCCCA
170	80	90	100	110	120
TATGTGGGCG	TTGGCATGAT	GGTTCATAG	GTATCTGAC	GTATTAATAT	TATCAGGGCG
ATTACAGGCG	ACAGGTATG	CGAGATATG	CATATGATG	CGATTTTAA	TATATCGGCG
130	140	150	160	170	180
GATCGGGGCA	GTITTTGCGG	TGGTTTGTG	CCATTTTATC	CTGTCTGCTG	CGTATGATG
CTAGCGCGGT	CAAAAAGCG	ACCAAGAAC	GGTAATAAG	GACAGACAC	GGCATTGCA
190	200	210	220	230	240
CGTACAGCGG	TTTATGCGGT	GCGGTATG	CGAGTTATG	GTAAATCAT	ATGCGGCGG
CGAGTGGCG	AAAATGCGC	CGCGATTA	CGTCAATC	CATTATAGTA	TACGCGCGG
250	260	270	280	290	300
AATGGATG	TGAGCGCAT	TGTGTCTTA	TGATTAAT	TATGATATA	AGGTGATAT
TTACCTATG	ATCGCGGTA	ACAGAGGAT	ACTGATAGT	ACTGATAGT	TGACGATAT
310	320	330	340	350	360
GATTTTTCT	CGTATGAT	TGATACAGG	GATCGGCT	GATTCGCA	ATTGCTATAT
TATGAGGAT	CAATAAGAG	ACCGGATG	ACTATGCTG	CTGAGCTG	CTATAGGAT
TTGATGAT	CTATATGAT	TATGATGAT	TTGAGGAT	CTATAGGAT	TATGATGAT
370	380	390	400	410	420
AGGCTGATG	TGATACAGG	GGATATGAA	ATCATGATG	AGGCTGATG	GTGCTGAGG
TGGAGCAT	AGGCTGATG	CTATATGAT	TATGATGAT	TTGAGGAT	CTATAGGAT
430	440	450	460	470	480
TTGCGGCGGA	TTATTTTTCT	TTGGGATGAT	TGCTTTTTCT	TGCTTTTTCT	TTGCGGCGG
AAAGGCGGCT	CAATAAGAG	ACCGGATG	ACTATGCTG	CTGAGCTG	CTATAGGAT
490	500	510	520	530	540
ATGTTATG	TGCTGCGG	TTATGATG	AGTATGATG	CTATTTTTCT	TATGATGAT
TGATATGAT	ATGAGGATG	ATCATGATG	TGATATGAT	GATGATGAT	ATTATTTTTCT
550	560	570	580	590	600
ATACGAGG	CTTTACGCG	CTGCTGAT	TAGCATGAT	CTGCTGAT	GTGCTGAT
TTGCGGCGGA	TTATTTTTCT	TTGGGATGAT	TGCTTTTTCT	TGCTTTTTCT	TTGCGGCGG
610	620	630	640	650	660
GATGCGGAT	GATGCGGAT	ATTATTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TTGCGGCGG
670	680	690	700	710	720
TGTTTTTCT	TTTATGCGG	TTGAGGAT	AAATGATG	TTGATTTTTCT	TTTATTTTTCT
730	740	750	760	770	780
ATTATGAT	TTTATGCGG	TTGAGGAT	AAATGATG	TTGATTTTTCT	TTTATTTTTCT
790	800	810	820	830	840
GATGATGAT	TATGATGAT	TGATTTTTCT	TGATTTTTCT	TGATTTTTCT	TGATTTTTCT
850	860	870	880	890	900
ATGTTATG	TGCTGCGG	TTATGATG	AGTATGATG	CTATTTTTCT	TATGATGAT
910	920	930	940	950	960
TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT
970	980	990	1000	1010	1020
TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT
1030	1040	1050	1060	1070	1080
ATGATGATG	ATGATGATG	ATGATGATG	ATGATGATG	ATGATGATG	ATGATGATG
1090	1100	1110	1120	1130	1140
ATTGATGAT	TATGATGAT	TATGATGAT	TATGATGAT	TATGATGAT	TATGATGAT
1150	1160	1170	1180	1190	1200
AGGCTGAT	TGATGATG	TGATGATG	TGATGATG	TGATGATG	TGATGATG
1210	1220	1230	1240	1250	1260
GATATGAT	CTATGATG	CTATGATG	CTATGATG	CTATGATG	CTATGATG
1270	1280	1290	1300	1310	1320
TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT	TTTATTTTTCT
1330	1340	1350	1360	1370	1380
TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT
1390	1400	1410	1420	1430	1440
TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT	TGCTTTTTCT
1450	1460	1470	1480	1490	1500
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1510	1520	1530	1540	1550	1560
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1570	1580	1590	1600	1610	1620
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1630	1640	1650	1660	1670	1680
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1690	1700	1710	1720	1730	1740
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1750	1760	1770	1780	1790	1800
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1810	1820	1830	1840	1850	1860
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1870	1880	1890	1900	1910	1920
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1930	1940	1950	1960	1970	1980
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
1990	2000	2010	2020	2030	2040
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2050	2060	2070	2080	2090	2100
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2110	2120	2130	2140	2150	2160
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2170	2180	2190	2200	2210	2220
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2230	2240	2250	2260	2270	2280
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2290	2300	2310	2320	2330	2340
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2350	2360	2370	2380	2390	2400
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2410	2420	2430	2440	2450	2460
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2470	2480	2490	2500	2510	2520
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2530	2540	2550	2560	2570	2580
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2590	2600	2610	2620	2630	2640
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2650	2660	2670	2680	2690	2700
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2710	2720	2730	2740	2750	2760
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2770	2780	2790	2800	2810	2820
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2830	2840	2850	2860	2870	2880
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2890	2900	2910	2920	2930	2940
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT
2950	2960	2970	2980	2990	3000
TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT	TTGCTTTTTCT

Fig. 3. Nucleotide Sequences of the ENase and MTase Gene

The deduced aa sequences for ENase and MTase are shown below or above the gene sequences, respectively. U shows a stop codon in the amino sequences.

system.¹²⁾ We also compared the MTase gene of *DsaV*¹³⁾ or *ScrFI*¹⁴⁾ to the gene of *StyD4I* MTase. They showed 62.0% or 68.4% homology between two MTase genes,

respectively. These DNA sequences of ENases, *DsaV* and *ScrFI*, have not been reported. We concluded that these DNA sequences of the *StyD4I* R-M system are almost the same as that of the *SsoII* R-M system. PL4MSSOII, 2648 bp ds(double strand)-DNA, was registered in the Gene Bank as accession number M86545. These sequences belong to the plasmid, P4, of the *Salmonella sonnei* 47 strain. Plasmid P4 encodes *SsoII* ENase and MTase. These enzymes cleave at the same site, ↓ CCNGG. As shown in Bergey's manual of Systemic Bacteriology,¹⁵⁾ DNA relatedness between *Shigella* and *Salmonella* is 11—25%, approximately. However, our data demonstrate that *Shigella sonnei* 47 and *Salmonella* Typhi D4 have common sequences on P4 and pSTd4, respectively. There have been no comparisons between the DNA sequence data of the two R-M systems which work similarly. It is the first time to demonstrate this coincidence of DNA sequences.

The deduced amino acid sequences of the *StyD4I* R-M system were analyzed. All ten motifs were also conserved in *StyD4I* MTase. The MTases of *SsoII*, *DsaV* and *ScrFI* are all m5C-MTases and share such motifs.¹⁶⁾ Both amino acid sequences in all ten motifs of *StyD4I* and *SsoII* MTase were same. At the Motif I and X site, DNA sequences of several m5C-MTases were compared with each other (data not shown). As a result, it is recognized that the DNA sequences in these motifs are also conserved to some extent.

REFERENCES AND NOTES

- Miyahara M., Mise K., *Gene*, **129**, 83—86 (1993).
- Ishiwata N., Tanimura A., Miyahara M., Mise K., *J. Food Hyg. Soc. Jpn.*, **36**, 404—408 (1995).
- Roberts R. J., REBASE ver. 605, www.neb.com/rebase (1996).
- Posfai J., Bhagwat A. S., Posfai G., Roberts R. J., *Nucleic Acids Res.*, **17**, 2421—2435 (1989).
- Laue F., Evans L. R., Jarsch M., Brown N. L., Kessler C., *Gene*, **97**, 87—95 (1991).
- Karyagina A. S., Lunin V. G., Nikolskaya I. I., *Gene*, **87**, 113—118 (1990).
- Fitzgerald G. F., Daly C., Brown L. R., Gingeras T. R., *Nucleic Acids Res.*, **10**, 8171—8179 (1982).
- Sambrook J., Fritsch E. F., Maniatis T. (eds.), "Molecular Cloning. A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989, pp. 1.25—1.28.
- Hashimoto-Gotoh T., Franklin F. C. H., Nordheim A., Timmis K. N., *Gene*, **16**, 227—235 (1981).
- Yoshida Y., Okumura N., Kato J., Watanabe H., *J. Gen. Microbiol.*, **137**, 867—874 (1991).
- Sanger F., Nicklen S., Coulson A. R., *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463—5467 (1977).
- Karyagina A. S., Lunin V. G., Degtyarenko K. N., Uvarov V. Y., Nikolskaya I. I., *Gene*, **124**, 13—19 (1993).
- Bhagwat A. S., Gopal J., Yebra M. J., GenBank, DSU10528, accession No. U10528 (1994).
- Davis R., van der Lelie D., Mercenier A., Daly C., Fitzgerald G. F., *Appl. Environ. Microbiol.*, **59**, 777—785 (1993).
- Brenner D. J., "Berger's Manual of Systematic Bacteriology," Vol. 1, ed. by Kreig N. R., Williams & Wilkins, Baltimore/London, 1984, p. 411.
- Kumar S., Cheng X., Klimasauskas S., Posfai J., Roberts R., Wilson J., *Nucleic Acids Res.*, **22**, 1—10 (1994).