

COMMUNICATION

Nucleotide Sequence of the Gene of *HgaI* Restriction Endonuclease

Hiroyuki SUGISAKI*

Received July 20, 1993

The nucleotide sequence reported in this paper has been submitted to the DDBJ, EMBL, and GenBank Data Bank with accession number D17388.

The abbreviations used are: bp, base pair(s); kb, kilobase pair(s); ORF, open reading frame; R-M system, restriction-modification system; *RHgaI*, *HgaI* restriction endonuclease; *MHgaI*, *HgaI* modification methylase; IPTG, isopropyl- β -D-thiogalactopyranoside

KEY WORDS: DNA cloning / *Haemophilus gallinarum* / Type IIS restriction-modification system / Expression in *E.coli*

The genes of a variety of the type II restriction-modification (R-M) systems that recognize symmetrical DNA sequences have been cloned and their gene organization and nucleotide sequences have been analyzed [see the reviews¹⁻³⁾]. As concerns the genes consisting of the type IIS R-M systems that recognize asymmetrical DNA sequences, however, only those of the *FokI*⁴⁾, *MboII*⁵⁾ and *StuI*⁶⁾ [the endonuclease recognizes the same DNA sequence as *FokI* endonuclease but cleaves at different positions⁷⁾] R-M systems have been reported so far.

The *HgaI* R-M system found in *Haemophilus gallinarum*, that belongs to this category, recognizes a 5 bp segment of DNA, consisting of 5'-GACGC-3' in one strand and 3'-CTGCG-5' in the other (*HgaI* recognition site). *HgaI* restriction endonuclease (*RHgaI*) introduces staggered cleavages 5 and 10 nucleotides downstream from the recognition sequences^{8,9)}. Experiments have been designed to clone the genes consisting of the system and a DNA fragment carrying the genes of two *HgaI* modification methylase (*MHgaI*) has been successfully isolated from a genomic DNA library of *H.gallinarum*¹⁰⁾. The two methylase genes are respectively responsible for modification of the internal cytosine residues of the CTGCG and GACGC strands and render DNA resistant to digestion by *RHgaI*. However, the *RHgaI* gene was not identified. Since the genes of restriction endonucleases are usually linked to those of the corresponding methylases on chromosome, truncated open reading frame 3 (ORF3) previously identified, which is located just downstream from the two *MHgaI* genes, is most likely to be the corresponding gene (see Fig.2 in Ref.10). The separation of the domains for DNA sequence recognition and cleavage makes type IIS restriction endonucleases interesting objects for studies of DNA-protein recognition. Therefore, an experiment was designed to clone the complete gene from the chromosomal DNA.

As to maintain the plasmid carrying a restriction endonuclease gene in *E.coli* cells, coexistence of the corresponding methylase gene is indispensable. First, the plasmid carrying the

*梶崎弘幸: Research Facility of Nucleic Acids, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611, Japan

two *MHgaI* genes was constructed. The 2.9 kb region of pKS318¹⁰⁾ (from the leftmost *HindIII* site to the *XbaI* site at nucleotide position 2909, see Fig. 2) was inserted into the *HindIII*-*XbaI* region in the multicloning sites of pUC19¹¹⁾. Deletion was introduced from the *XbaI* site and a clone carrying a DNA fragment of nucleotide positions 1 to 2497 was isolated. The coding region for the two *MHgaI* genes could be just generated as a 2.4 kb fragment from the plasmid construct by *RNsp*(7524) V-*REcoRI* digestion and inserted into the *EcoRV* site of pACYC184¹²⁾ by blunt end ligation. The resulting construct carrying the two *MHgaI* genes under control of the *tet* promoter was named pKS430 (see Fig. 2) and introduced into *E. coli* MCH1061 cells¹⁰⁾.

Next, a DNA fragment adjacent to the insert of pKS318 on the *H. gallinarum* chromosome was searched. When digests of the chromosomal DNA by various restriction endonucleases were subjected to the Southern blotting hybridization¹³⁾ using, as probe, the 390 bp *XbaI*-*HindIII* fragment (nucleotide positions 2,909 to 3,399, see Fig. 2) which is located in the coding region for ORF3, an approximately 4.5 kb *ClaI* fragment was identified (data not shown). The DNA fragment would contain the region adjacent at the right side to the insert and a sufficient coding capacity for complete ORF3 because the *ClaI* site is located just upstream from the gene (see Fig. 2). Complete *ClaI* digests of the chromosomal DNA were fractionated by electrophoresis on 0.7% agarose gel, and fractions predominantly containing the fragments of 3.5-5.5 kb were collected. The sized *ClaI* fragments were ligated to a phosphatase-treated *AclI* digest of pUC19 and introduced into *E. coli* MCH1061 carrying pKS430. Among the recombinants examined by colony hybridization¹³⁾ using the *XbaI*-*HindIII* fragment as probe, one carrying ORF3 under control of the *lac* promoter in a single 4.5 kb *ClaI* fragment was named pKS456. The *RHgaI* activity was actually detected in the extract from the *E. coli* cells by induction with isopropyl- β -D-thiogalactopyranoside (IPTG). I, therefore, concluded that the 4.5 kb *ClaI* fragment contains the *RHgaI* gene.

The nucleotide sequence of the region adjacent to the insert of pKS318 was determined using the dideoxy sequencing method^{14,15)}. The DNA sequence of the 1.8-kb leftmost *ClaI* to *HincII* region of pKS456 is shown in Fig. 1. Analysis of coding capacities indicated that ORF3 terminates at nucleotide position 3808.

To verify that ORF3 is indeed the *RHgaI* gene, the endonuclease activity was determined by the *in vitro* assay procedure. The leftmost 1.6 kb region of the 4.5 kb *ClaI* fragment (from the *BamHI* site on the pUC19 moiety to the *EcoT22I* site, see Fig. 2) which encompasses entire ORF3 was inserted into the *BamHI*-*PstI* region in the multicloning sites of the expression vector pKK223-3 (Pharmacia P-L Biochemicals Inc.). The resulting construct carrying ORF3 under control of the *tac* promoter was named pKS475. The plasmid construct was introduced into *E. coli* MCH1061 cells carrying pKS430 and after induction with IPTG, protein expressed from the gene was purified by phosphocellulose (Whatman), heparin-agarose (Bethesda Research Laboratories), and hydroxyapatite (Clarkson Chemical Co.) columns. When the various DNAs were digested with the purified protein fraction and authentic *RHgaI* isolated from *H. gallinarum* cells, the same restriction patterns were obtained (data not shown). The results clearly indicated that ORF3 was the *RHgaI* gene. In ORF3, the ATG codon occurred at nucleotide positions 2345 and 2360. The predicted molecular weights of proteins beginning with ATG at nucleotide positions 2345 and 2360 were 56,734 and 56,119, respectively. I assigned the ATG codon at 2360 to the initiation site for translation of the endonuclease for only this one was accompanied by a putative ribosomal-binding sequence, 5'-GGAG-3',

HgaI Restriction Endonuclease Gene

TATATTAGAACTATTATTGGTGAGGCTATTTCCACCCAAATTACTTTTCAGCAATATGTTTT	2280
CCTGATGGAGAAAATATTAAATGTTAAATGATGATCAAAATTTGGATATTTAAAAACATAC	2340
Met Glu Lys Ile Leu Met Leu Asn Asp Asp Gln Ile Trp Ile Phe Lys Lys His Thr	2400
TAATAATATTCAGCTATTAATTGAAGTTGCTTTGTATTTAAAAAGCAATAAATCTTCTGT	19
Asn Asn Ile Gln Leu Leu Ile Glu Val Ala Leu Tyr Leu Lys Ser Asn Lys Ser Ser Val	2460
TTCAAAAAAGATAAAGATGCGATGTATGATATATTCAGTGAATCTGAATTATATAATCC	39
Ser Lys Lys Asp Lys Asp Ala Met Tyr Asp Ile Phe Ser Glu Ser Glu Leu Tyr Asn Pro	2520
AAGAGAATCTTTACGTGATAAACCTTTGGACACAATTAATCATAAATTAGATGGGTATC	59
Arg Glu Ser Leu Arg Asp Lys Pro Leu Asp Thr Ile Asn His Lys Leu Asp Gly Leu Ser	2580
TTATTTTATGTTTGGGTACTCAGATAGAATAAATGATGAAAAATAAATTTATTTTATGTC	79
Tyr Phe Met Phe Gly Tyr Ser Asp Arg Ile Asn Asp Glu Asn Lys Phe Ile Phe Ser Pro	2640
ATTAGGAAATCTTTTCTGAAATATCTACATGATAAAGATAAATTAAGCAAAATTTTTTCT	99
Leu Gly Asn Leu Phe Leu Lys Tyr Leu His Asp Lys Asp Lys Leu Ser Lys Ile Phe Ser	2700
TTGTGCTTATTAGTATGCAATTTCTCATCCATAGTAAGCCTAGTGAATGCTTTTT	119
Cys Met Leu Ile Ser Met Gln Phe Pro His Pro Tyr Ser Lys Pro Ser Glu Cys Phe Leu	2760
ACTATATCCATTAGATTAATTTTTAAACTGCTTTTAGATAAACGCCCTACAGGTAAGGCT	139
Leu Tyr Pro Phe Arg Glu Tyr Ser Lys Leu Leu Asp Lys Arg Leu Gln Tyr Leu	2820
ATATCATTATGAAGTATATAAAATTATTATTCATACAATATCTATAGATGAAGCCAAATA	159
Gly Tyr Val Glu Tyr Lys Ile Ile His Thr Ile Ser Ile Asp Glu Ala Lys Tyr	2880
TGAATTTCTAGTCAAGAGTATATTAATTTCTAGAAAAAATCTTGGAAATGAGAACTTAA	179
Glu Phe Leu Val Lys Ser Ile Leu Asn Ser Arg Lys Lys Ser Trp Asn Glu Lys Leu	2940
TGAATTATCAGAAATACAACATAAAGTTGTTAAATCTGTATATGAGTGGCAATACTACAT	199
Glu Leu Ser Glu Ile Gln His Lys Val Val Lys Ser Val Tyr Glu Trp Gln Tyr Ile	3000
TGTGCCATTGTTAGGTACCTTACACATTTTAAATTAATAATGGAGATATAGAGCAAAA	219
Val Pro Leu Leu Gly Ser Leu His Ile Phe Lys Ile Asn Asn Gly Asp Ile Glu Gln Lys	3060
ACTTTATCATCTCAAAAGATGGAAGTAAATCACCTCCAACAGCAAGAAAGGCAAAATAA	239
Leu Tyr His Pro Gln Lys Asp Gly Lys Ser Pro Thr Ala Arg Lys Asn Asn	3120
TGGCTATGTAGAAATTAATGATAATTTGACTAATTTTATTGATAAATTATTGAACAAGTA	259
Gly Tyr Glu Ile Asn Asp Asn Leu Thr Asn Phe Ile Asp Lys Leu Leu Asn Lys Tyr	3180
TTCTTTTTTAGATACCCCTATCTTATTATCTGATAGTCAACGAAAATCAAATGATGTGAC	279
Ser Phe Leu Asp Thr Pro Ile Leu Leu Ser Asp Ser Gln Arg Lys Ser Asn Asp Val Thr	3240
TAAGGAAATTTATCTTTTACCCAGAGCTTCTTTAGCAGAAATGGCGAAACAAATCTC	299
Lys Glu Ile Tyr Ser Phe Tyr Pro Glu Leu Leu Leu Ala Glu Ile Gly Glu Thr Ile Ser	3300
TTTGAATCTCATATTTTAAATTTCTAAATTAATCACAGAATATTTCCAAGAATCCAGA	319
Phe Glu Ser His Ile Leu Asn Ile Pro Lys Leu Ile Thr Glu Tyr Ser Lys Asn Pro Asp	3360
TAATAGCACATCTGGCAAAATTTGAGAAAATTTCTTGAGGAAGCTTTTAAATTTATTCATTGA	339
Asn Ser Thr Ser Gly Lys Phe Glu Lys Ile Leu Glu Glu Ala Phe Asn Leu Ile Asp	3420
TGTAGAACTCAATGGTTAGCAGGAGCAGGCAGAACAGATATTGAATGTATGTATTTGCC	359
Val Glu Ala Gln Tyr Leu Ala Gly Ala Gly Arg Thr Asp Ile Glu Cys Met Lys Pro	3480
AATAAATGAAAAATTTCTATTGAAGCTAATCAACAAAAATAAATTAAGTATGATTAA	379
Ile Asn Glu Lys Phe Ser Ile Glu Ala Lys Ser Lys Asn Lys Leu Ser Met Ile Asn	3540
TTCAGGTAGATTAAACGCCACAGAACTTAAATAGTGCAAAATTAATACTATTGTTATTAC	399
Ser Gly Arg Leu Lys Arg His Arg Thr Leu Ile Ser Ala Asn Tyr Thr Ile Val Ile Thr	399
TCCAAGATATGTACCTAGTGTACGTTATGATATTGAAGCACAAGATATAGTTCTCATAAC	3600
Pro Arg Tyr Val Pro Ser Val Arg Tyr Asp Ile Glu Ala Gln Asp Ile Val Leu Ile Thr	419
AGCCGATACATTAGCAGAGTATTTATACAATAATATTATTTCTAATAATCGAGATATTC	3660
Ala Asp Thr Leu Ala Glu Tyr Leu Tyr Asn Asn Ile Ile Ser Asn Asn Arg Asp Ile Ser	439
TTATGCTGATATACAAGCTATCATTGTTGCTAATTTGGGAAAAGATATTAGTACTCAAAT	3720
Tyr Ala Asp Ile Gln Ala Ile Ile Val Ala Asn Leu Gly Lys Asp Ile Ser Thr Gln Ile	459
TTCAAATTTAACACTTTCCAAATTTGGTTGATGGATAAATGAATATGCATAAGAACATTA	3780
Ser Asn Leu Thr Leu Ser Lys Phe Gly	479
CAGAAAAAATCTTTGATGATAGCAATCACAAAATTTGTGATCTACATACCTCATAACTTA	3840
TACAGTTTTAAAGGGTTAAATCGCCATTCTATATTCGAGTGGGCGATTTTTTATGTCT	488
AGCGTTTACGCTTTGAGTACCTTGACACACAAGGAATCTTATGGATCATTACTCTTT	3900
CAAGCAAAATCTAAACATATCCCTTAAACATATCTTCGCTCTAAGTGAAGATGAAGCCT	3960
TCCAACCTTTGAAAGCAAAACCGTTGGGAAAATCCTGATAAAATCCTCCAAAGTAGTACGT	4020
CTAAATTTATAGAATAAGCCACTTAAGATAATCACTCAAGGTGTTTAAATATCATCAATT	4080
AAACTTGTCAAAAAATTTCTCTTGACAGAATAAGACCTTTTTGGCACTCGCTATATTTCA	4140
AAAAATGCAGAAAATTTCAAAAAGCACTTGACAAAAGAAACCGCTTGACAAGCACCCCTATA	4200
AAATAATATCTTTTGGCCACCGAGAATAGTCCCTTCATTATGACTTTGTTTTTAAAGCC	4260
TAAAAAGGCTTGGAAATTTTCAGAGTTTTAATGAAGGGATTTTTGTTTTCTATGGCACAACA	4320
TTTTCGGTTATCAGCGACAGCAGCCAGCTATCTATGGATAGCTTGGTGAGCTTGTCTGA	4380
TGACGAGATTTTTTCAGTATTGAAACAAGCTCGATGGGGTAATGTAGATGAAATAAACGA	4440
TGTTATTTTGTCCGCAATGCCATACTGCCACCACGCCTATTTTCTACTCGTAAACA	4500
ATGGCAGTGCAAAATTTGCCAACATCGTTTTTCTATTACAGCAGGCACTATTTTCAAGG	4560
TGCAAAAATCTTTTGGGTAAAATTTGTTGGCAATTTTCTATTTTTCGACTGAAAGCAA	4620
AGGATATCTGCTATTACACTATCACACAAGCTCAACGTGCAATATAAAACAGCGTGGGT	4680
ATTGCTACATAAATTTTCGTGAATCGTTGATAAAGCCAAAGATTTAACACATTTAAGTGG	4740
CGAAGTTTATTTGATGGCGCTTACATGAACAACATTCGTTCTCTAAAAACTTTCTCCA	4800
TAAACGTTATGATAGATGAAAGAAACGTTATCAGCGTGCAGATAAATCTTGTGATTGGT	4860
ATTCCGTC	4920
	4980

Fig.1 Nucleotide sequence of the leftmost *Clal* to *HincII* region of the 4.5 kb insert in pKS456. The amino acid sequence assigned for *RHgaI* are given below the nucleotide sequence. The nucleotide sequence is numbered from the leftmost end of pKS318⁹⁾, and the amino acid sequences from the first initiation codon, respectively.

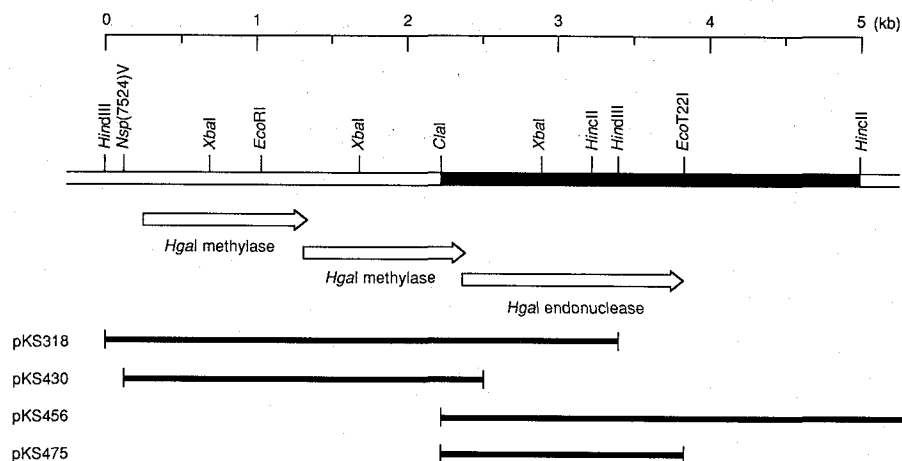


Fig.2. Gene organization of the *HgaI* R-M system. The filled line represents the region of which the DNA sequence was determined as shown in Fig.1 and open line the other region for *H.gallinarum* chromosomal DNA. Below the physical map on which the major restriction sites are indicated, the two methylase and one endonuclease genes are shown by open arrows, and the region inserted into the plasmid construct in this experiment are indicated by bold lines.

in the appropriate position. To confirm this assignment, I have a plan to analyze the N-terminal amino acid sequence.

The gene organization of the *HgaI* R-M system derived from Ref.10 and this paper is shown in Fig.2. This R-M system is consisted of two cytosine methylase genes responsible for modification of different DNA strands in the target DNA and a endonuclease gene for cleavage of both DNA strands.

Acknowledgments

I wish to thank Dr. M. Takanami for his instructive advice and criticism during the course of this work.

References

- (1) G.G. Wilson, *Nucleic Acids Res.*, **19**, 2539-2566(1991).
- (2) G.G. Wilson, and N.E. Murray, *Ann. Rev. Genetics*, **25**, 585-627(1991).
- (3) T.A. Bickle, and D.H. Krüger, *Ann. Rev. Microbiol.*, **57**, 434-450(1993).
- (4) K. Kita, H. Kotani, H. Sugisaki, and M. Takanami, *J. Biol. Chem.*, **264**, 5751-5756(1989).
- (5) H. Bocklage, K. Heeger, and B. Müller-Hill, *Nucleic Acids Res.*, **19**, 1007-1013(1991).
- (6) K. Kita, M. Suisha, H. Kotani, H. Yanane, and N. Kato, *Nucleic Acids Res.*, **20**, 4167-4172(1992).
- (7) K. Kita, H. Kotani, H. Ohta, H. Yanase, and N. Kato, *Nucleic Acids Res.*, **20**, 618(1992).
- (8) N.L. Brown, and M. Smith, *Proc. Natl. Acad. Sci. USA*, **74**, 3213-3216(1977).

HgaI Restriction Endonuclease Gene

- (9) H. Sugisaki, *Gene (Amst.)*, **3**, 17-28(1978).
- (10) H. Sugisaki, K. Yamamoto, and M. Takanami, *J. Biol. Chem.*, **266**, 13952-13957(1991).
- (11) C. Yanisch-Perron, J. Vieira, and J. Messing. *Gene (Amst.)*, **33**, 103-119(1985).
- (12) A.C.Y. Chang, and S.N. Cohen, *J. Bacteriol.*, **134**, 1141-1156(1987).
- (13) J. Sambrook, E.F. Fritsch, and T. Maniatis, "Molecular cloning: A Laboratory Manual" 2nd Ed, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- (14) F. Sanger, S. Nicklen, and A.R. Coulson, *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467(1977)
- (15) J. Messing, *Methods in Enzymol.*, **101**, 20-78(1983).