Isolation and genetic structure of the *Avall* isoschizomeric restriction-modification system *Hgi*BI from *Herpetosiphon giganteus* Hpg5: *M.Hgi*BI reveals high homology to *M.Ban*I

Andreas Düsterhöft, Dirk Erdmann and Manfred Kröger*

Institut für Mikrobiologie und Molekularbiologie der Justus-Liebig-Universität Giessen, FB15 Frankfurter Straße 107, D-6300 Gießen, FRG

Received April 22, 1991; Accepted May 29, 1991

EMBL accession no. X55137

ABSTRACT

The complete type II restriction-modification system *Hgi*BI of *Herpetosiphon giganteus* strain Hpg5 recognizing the *Avall* specific DNA sequence GGWCC has been cloned and expressed functionally active in *Escherichia coli*. A considerable acceleration in cloning could be achieved by preparing a size restricted library after application of a related hybridization probe. Both methyltransferase (437 codons) and restriction endonuclease gene (274 codons) were found to be encoded on a 3.6 kilobases *Clal*/*HincII* fragment in the same transcriptional orientation separated by one triplet only. Protein sequence comparisons revealed a close resemblance of *M.Hgi*BI to the group of m⁵C-methyltransferases, especially to *M.Ban*I from *Bacillus aneurinolyticus* with the related recognition sequence GGYRCC. In contrast, no significant similarities have been observed for the associated endonuclease *R.Hgi*BI with any other restriction enzyme described so far, even not with the isoschizomeric *R.Sin*I from *Salmonella infantis*, or with *R.Ban*I.

INTRODUCTION

Type II restriction-modification systems seem to become one of the most intensively studied group of enzymes (1,2). This is not only because of their commercial interest, but also because they may serve as an especially useful model system for DNA protein interaction. Since they are found in more than 1117 different microorganisms (3), and since they can only use a limited number of specificities, each addition of a newly characterized system may help to increase the chance to predict structure function relationships between conserved amino acids and the limited number of nucleic acid residues involved. Until now only two X-ray structures for restriction enzymes [*Eco*RI and *Eco*RV (4)], but none for a methyltransferase are known. On the other hand, a fair number of restriction-modification systems are cloned and sequenced, and especially the m⁵C-methyltransferases can

already be described as a large group of enzymes with highly conserved amino acids seemingly involved in common functions (5,6,7). But even in this group no general rules for DNA recognition could be found.

Rather than searching for new, but unrelated specificities our laboratory concentrates on one selected organism, to study these rules in a more evolutionarily orientated approach (8,9). Here we describe the cloning and analysis of the restriction-modification system *Hgi*BI (GGWCC) from the gliding bacterium *Herpetosiphon giganteus* Hpg5 (9) and the particular relationship of the primary structure of its methyltransferase to the recently published sequence of *M.Ban*I from *Bacillus aneurinolyticus* modifying the related sequence GGYRCC (10).

MATERIALS AND METHODS

Bacterial strains and plasmids

Herpetosiphon giganteus Hpg5 was a gift of H. Reichenbach (GBF, Braunschweig). *E.coli* HB101 (*hsdR_B⁻ hsdM_B⁻, mcrA⁺ mcrB⁻*) was taken from our strain collection. Plasmid pBluescriptIISK(+) was originally obtained from Stratagene (Heidelberg).

Media

Herpetosiphon giganteus cultivation was performed as described by Mayer und Reichenbach (11) in *Hp* 74 medium at 28°C and on Myx agar plates. *E.coli* strains were grown in Luria Bertani (LB) medium and on LB agar plates according to standard procedures (12).

Enzymes and chemicals

Restriction enzymes, T4 DNA Ligase and Klenow polymerase were obtained from various distributors (AGS, Heidelberg; Boehringer, Mannheim; New England Biolabs; Schwalbach). Chain termination sequencing was done using T7 DNA polymerase according to the manufacturers protocol

* To whom correspondence should be addressed

(Pharmacia/LKB, Freiburg). [^{35}S - α]dATP (> 650 Ci/mmol) was obtained from Amersham (Braunschweig).

DNA preparation

Total DNA of *H. giganteus* cells was purified by SDS/proteinase K lysis followed by a N-cetyl-N,N,N-trimethylammonium-bromide/NaCl and phenol/chloroform extraction as described in reference 13. Plasmid DNA isolation was performed by mini-prep alkaline-SDS procedure (12).

Transformation

E. coli transformations were carried out according to the Hanahan protocol (14).

Isolation and identification of genomic DNA fragments encoding *hgiBIRM*

H. giganteus Hpg5 DNA prepared from 2 g of cells was digested with various restriction enzymes and fragments were separated on agarose gels. Southern-transfers were performed according to standard procedures (12), except that the nonradioactive Digoxigenin-dUTP labelling kit (Boehringer, Mannheim) was used for signal detection. As DNA probe the previously cloned isomethylomeric gene *hgiCIIM* of *H. giganteus* Hpg9 was used for immunochemical signal detection. Different DNA fragments hybridizing with the methyltransferase *hgiCIIM* insert were identified. A suitable *Clal/HincII* generated fragment region in the range of 3.3 to 3.9 kb was isolated from an agarose gel to form a size restricted genomic library.

Construction of a partial library from *H. giganteus* genomic DNA

About 0.1 μg of the isolated *H. giganteus* Hpg5 DNA fragment region was ligated with T4 DNA Ligase into 0.3 μg *Clal/HincII* cleaved pBluescriptISK(+) vector DNA at 17°C for 16 hr in 20 μl reaction volume. The ligation mixture was transformed into competent *E. coli* HB101, cultured in LB medium for 1 hr and then plated onto LB plates containing 100 $\mu\text{g}/\text{ml}$ ampicillin (Ap). Following an overnight growth period about 500 colonies were scraped together in 5 ml of 10 mM Tris.HCl pH 7.5, 10 mM MgCl_2 . 1 ml of this library was used to inoculate 20 ml of Ap-LB medium for the isolation of plasmid DNA.

Selection of methyltransferase clones from a partial Hpg5 library

0.3 μg DNA of the partial library was completely digested with 36 units of the commercially available *HgiBI* isoschizomer *AvaII* (New England Biolabs). After transformation into competent *E. coli* HB101, plating onto Ap-LB agar plates and overnight incubation at 37°C, mini-prep plasmid DNAs of individual surviving colonies were isolated and analyzed for *in vitro* *AvaII* endonuclease resistance.

Detection of restriction endonuclease R.*HgiBI* activity *in vitro*

A 20 ml liquid culture of *E. coli* HB101 harbouring the respective derivative of plasmid pRMB1-1 was grown to saturation. After centrifugation for 10 min at 5000 rpm the pellet was resuspended in 5 ml of 10 mM Tris.HCl pH 7.5, 1 mM Na_2EDTA , 10 mM 2-mercaptoethanol, 0.01 % Triton X-100, 1 mg/ml lysozyme.

Figure 1. Genetic organization of the R-M system *HgiBI*. The plasmid pRMB1-1 containing a *Clal/HincII* insert of 3.6 kb (indicated as 'cloned' area) was selected by *AvaII* overdigestion of a size restricted *H. giganteus* Hpg5 library in *E. coli* HB101. DNA sequencing was done on both strands of the *Clal/NheI* segment (indicated as 'sequenced'). Functional localization of *hgiBIM* and *hgiBIR* was carried out by *in vitro* analysis of subclones (shown below the restriction map). The enzymatic activities were correlated with the sequence data in Figure 2. Double headed arrows indicate orientation independent activity supported by the outside-in p_{lac} promoter. Single headed arrows indicate orientation dependent activity with only one viable construct. The failure to subclone the central *HindIII* segment in any orientation was interpreted as the self-destruction of a *hgiBIR*⁺*M*⁻ phenotype with R.*HgiBI* being expressed from its own promoter. * denotes an *HincII* site uncleaved in genomic DNA (see text).

Disruption of cell walls was achieved by sonication for 1 min. Cellular debris and insoluble material was removed by 20 min centrifugation at 18000 rpm. 1, 2.5 and 5 μ l of the supernatant were tested for their ability to cleave 0.3 μ g of suitable plasmid DNAs in 20 μ l reaction mixture containing 10 mM Tris.HCl pH 8.0, 10 mM MgCl₂, 1 mM 2-mercaptoethanol at 37°C for 1 hr.

DNA sequence determination

Sequence information was determined in both orientations using the chain termination reactions (15) either on certain subclones

with standard primers or on purpose synthesized oligonucleotides as specific primers. These oligonucleotides were kindly synthesized by Dr. J.Hegemann from our laboratory on an ABI Model 380B DNA synthesizer.

Computer analysis

Analysis of DNA and protein sequence data was carried out on 80386 AT-compatible computer systems using the DNASIS and PROSIS software (Version 5.02; Pharmacia/LKB) and the FLEXP program (16).

```

ATCGATCGCCGCCCAAGTGATGGTTGAAATTAACGGCATGCTCAACGATGGCTGCACCGCATTCCACGAAGCCAAGTGGTGAAGGCAACACGATT 100
AAAAATTGAAGTAACGACGATTCCGGCCAAAGATGCTATGTGTACCCAAAGAAATCTCACCATTAGCACTACGATTCAAGTTGATGCTCAGCTCCAACGAG 200
GTGAATATACGATTCTGGTTAACGATGTTGCTGAAGCGCTTAAATTTGAAATTTGCAGCAAGTTTTGATTTTCAGCGGGAGCTTCGGCTCTCCGTTGAT 300
TGTTTAATCTGAAATGTTAAGTGCACACATTAATTTGATCAATCTTCATTAACTGCTCACTTTGACAGATTGTCAAGCTGTTCTAAATAATCGAACAGAC 400

M.HgiBI:      M Q Q F R F I D L F A G I G G F R L G L E A V G G V      26
ATTAACCACTGGAAGATAATTTATGCAACAATTCGCTTTATAGATTTATTTGACGGGATTGGTGGGTTTCGGTTAGGCCCTGAGGCAGTTGAGGAGTGT 500
      *****
C V A S A E I D Q Q A I K V Y R Q N W P T D G V D H N L G D I T A I      60
TGTGTTGCCCTCAGCTGAAATTGATCAACAGCGATTAAAGTTTATCGGCAAAATGGCCTACCGATGGAGTAGATCATAATTTAGGTGATATTACAGCAA 600

      Q Q L P A H D V L V G G V P C Q P W S I A G K N Q A F D D P R G Q      93
TTCAACAATTGCCGCCCATGATGACTCGTTGGTGGCGTACCATGCCAACCTTGGTGCATTGCAGGCAAGAATCAAGCGTTGATGATCCACGTGGGCA 700

      L W A D V I R L V Q I N Q P K A F I F E N V K G L V D P R N R L C      126
ATTGTTGGCTGATGTTATTCGGCTGGTTCAAATTAATCAGCCTAAAGCCTTTATTTTGAAAATGTCAAAGGTTTAGTTGATCCACGAAATCGTTTATGT 800

      L E I I L D S F K D L G Y S V F Y K L L N S F D F G V A Q N R D R V      160
TTAGAAATCTGTTGATAGTTTAAAGATTAGGCTATAGCGTTTTTTATAGCTATTAATTTCTTTGATTTTGGGGTTGCCCAAAATCGTGATCGCG 900

      F I V G I Q Q K L D L N G F S F P E Y T E S E Q R L Y H I L D N L      193
TTTTTATGTTGGAATACAACAAAACCTTGACCTGAATGGCTTTTCATTCCAGAATATACTGAATCTGAACACCGGCTTTACCATTATTTAGATAACT 1000

      E V P E T K L E S I P I Q R N L F G E R I D V G Y N K L T P R G A      226
TGAAGTACCGAACTAAGCTTGAATCAATTCGAATTCAGCGCAATTTATTTGGTGAGCGAATTGATGTTGGCTATAATAAATAACTCTCGTGGTGCA 1100

      F N D F F I L N D I R N G P T S I H S W E I Y P T T E R E K Q I C M      260
TTTATGATTTTTTTTATCTTAATGATATTCGCAATGGCCCAACAAGTATTCATAGCTGGGAAATATATCTCAACTGAACGTGAAAGCAGCATTTGA 1200

      I I M R N R R N S R Y G D C D G N P M S Y Q D I A E L V A G L A E      293
TGATTATTATGCGTAATCGACGAATTTCTCGTTATGGCGACTGCGATGGCAACCCGATGAGCTATCAGGATATAGCTGAATTTGGTGTGCTGGTTAGCAG 1300

      K E L Q T L V E K R I L R Q Y P D G K Y E F F N R R L S G G I D G      326
AAAAGAAATACAAACCTTAGTAGAAAACGAATTTACGTCAATATCCTGATGGTAAGTATGAATTTTTTAATCGCGCTTTATCTGGTGGGATCGACGGA 1400

      T Y R I F L P N A R F F G T L T A R G M H D E I A E I S V S G A N A      360
ACATACCGAATTTTTTACCAATGCGCGCTTTTTTGGTACATTAACCGCAGTGGAATGCAGATGAAATGCTGAAATTAGTTAGTGGTGAATG 1500

      E E Y K H N F I Q Q V L I P K R Y R K I T V S E A A R L Q G F P G      393
CTGAGGAATAATAACATAATTTTATTCAGCAAGTGTAAATCCCAAAACGCTATCGGAAAATTACCGTAAGCGAAGCTGCTCGTTACAAGGCTTCTCTGG 1600

      S F Q F H S N Q S A N F R L I G N S V A P P V I V A L G K A L Q C      426
ATCATTCGAATTTCACTCCAATCAATCAGCAAAATTTTCGCTTGATTGGCAATAGTTGCTCCACCAGTTATGTTGGCTTGGGTAAGCGTTGCAATGT 1700

V K L F E Q E L C E V *
GTAATAATTTATGAGCAAGAGTTGTGTGAGGTTTGAATATGGCGATTAAATCCGATTACTCGTAATAAGATTAAGGATTATCTGAATAGTTTATTCAAC 437
R.HgiBI: ***** M A I N P I T R N K I K D Y L N S F I Q Q      21

AGCAATTATCAGTTTATTCGCAACGAAGTTTGCAGCAATTCAGGATGTTGATAGCTATCCGTCATCGCTATCAAAAGATGGTGATCTGAAGCCGTTTCA 1900
Q L S V Y S Q R S L R E F Q D V D S Y P S S L S K D G D L K P F H      54

TGCACTCACTTATTCCTGCCTCAATTTGCGCTTAAATCGCTTTGAGCGCAGCTTGAGCACAGGGCTTGGCTCAACATTTGAGGAATGTACACGCTTAAT 2000
A S L I P A S I M R L N R F E R S L S T G L G S T F E E C T R L I      87

GCGCTTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
A L D H H A V A L R N Y D I Q A A L D Q A Q W A A I D Q L I S T I D      121

ATCGTGGCTGAAGCACCAAACTCCATCACTTAATCAAAATGCTTGAGCAAAATCAATCAATCACTTAACCGGTAATCTAGAAACCAATATGTTTCGCGC 2200
R G L K H Q T P S L N Q M L E Q I Q S I P L T G I L E T H I V R A      154

AGATCTGTATATTCACGACATGATGGTAGCGAATTTGTTTTGAAATAAAATCGCCAAACCTAGTAAAGGCAATGCCTTGAGGTTATGCAACGCCTG 2300
D L Y I Q R H D G S E L F F E I K S P K P S K G Q C L E V M Q R L      187

CTGCGAATTTATACAATCAAAACAATCTGCTGTGCCAGTCAAGGCTTTTATGCAATGGCCTATAACCCCTTGGGGTATTAGCCGAGCTAGTTATCGCT 2400
L R I Y T I K Q Q S A V P V K A F Y A M A Y N P W G I S R A S Y R S      221

CCAGTATTACTAAAAATATACCGATTTTCCAATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2500
S I T K K Y T D F S N A V V I G Q E F W S L I G E P S T Y T E L L      254

GGAAATTTATCATGAAGTAGGGTTAGCAAAATCTGCTGAGATAACCCAAAACCTTTTGAATAGCCTCATAAAGGCTCGCAAAAGTGATAGATACGGCGA 2600
E I Y H E V G L A K S A E I T Q K L L Q *      274

AGTTGTTGGACGATATTGCGGCTCCATGGCCAGCTGAACCAACAGGATACAACATACACAACTCTTTCACGACAAATTAAGCAGCAAGCCCAACGCTT 2700
TGAGGAGAGTGCAACATCCACTGGCGATTGCGGCTTTGCTTGAGCTTTATTTAGCAACTGGCTATGAAAAAGTTGTGGAGAATTGTTGAAACATTAA 2800
TCGCAAAAATGCTGAATTTGGGGCTTGGCGCGCAGCATTCAAGGTTGTAAGGTAGAGCTTAGGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2900
TTGTTGATCTTCTACGCGCTTGATTGCGCGCTAGC      2938

```

Figure 2. Nucleotide sequence of the 2938 bp *Clal/NheI* segment encoding RM.HgiBI and predicted primary structures of both enzymes (M: position 423 to 1736; R: 1740 to 2564). Putative '-10' and '-35' promoter sequences are underlined within the DNA sequence. Potential Shine-Dalgarno sequences are marked by (*).

RESULTS AND DISCUSSION

The DNA sequence GGNC seems to be very much in favour as recognition sequence for prokaryotic defense systems using restriction and modification. According to Roberts (17) 39 isoschizomers of *AsuI* from 23 different organisms are characterized. In addition 52 isoschizomers with the slightly higher specific *AvaII* recognition sequence GGWCC are characterized from 28 different organisms. Consequently, this system provides an exceptional source for systematic analysis and comparison of restriction-modification systems with the same recognition sequence.

For each GGNC recognition system one representative has been sequenced, *Sau96AI* (18) and *SinI* (19). We isolated and characterized a second system of the GGWCC type, namely *HgiBI*, as described below.

Identification and cloning of the genes encoding the R-M system *HgiBI*

From *H. giganteus* *Hpg9* a DNA fragment was cloned in our laboratory, which coded for one of the relevant methyltransferase genes, *hgiCIIM* (to be published elsewhere). Plasmid DNA containing this gene was used as a probe in a Southern-blot analysis to look for possible related sequences belonging to the *hgiBIM* gene of *H. giganteus* *Hpg5*. Indeed, several signals could be detected using differently cleaved genomic *Hpg5* DNA (e.g. *EcoRI*, *ClaI*, *HincII*, or *ClaI/HincII*, data not shown). Among these hybridizing DNA fragments was a *ClaI/HincII* generated of about 3.6 kb, which seemed to be suitable for cloning.

Applying the well introduced selection for methyltransferase activity (20), this size reduction led to the immediate identification of a *M.HgiBI* containing clone, with the expected 3.6 kb *ClaI/HincII* insert. For selection the plasmid DNA from the size reduced library of *H. giganteus* *Hpg5* DNA was exhaustively digested with isoschizomeric *AvaII* endonuclease from commercial sources. After retransformation into *E. coli* HB101 8 out of 20 analyzed colonies contained plasmids mediating full resistance against *R.AvaII* cleavage, while other (e.g. *AhaII* and *BanI*) restriction activities remained unaffected (data not shown).

Crude extracts of these clones were able to split DNA from different plasmids into an *AvaII* specific pattern *in vitro* (data

not shown). Therefore, the cloned DNA fragment from *H. giganteus* *Hpg5* was concluded to encode the complete *HgiBI* R-M system and the respective plasmid was consequently named pRMB1-1. This immediate success is, however, in contrast to the *HgiCII* R-M system, which could not be cloned directly as a complete system (to be published elsewhere).

Functional localization of *hgiBIRM*

Functional characterization *in vitro* of various subclones derived from the pBluescriptISK(+) derivative pRMB-1 containing the entire *ClaI/HincII* insert allowed the preliminary localization of the genes coding for RM.*HgiBI* as given in Fig. 1. For *M.HgiBI* activity a *ClaI/NsiI* generated 1904 bp DNA segment spanning the left halve in Fig. 1 was sufficient. It exhibited methyltransferase activity orientation independently. Therefore a promoter must be located within this fragment. The *R.HgiBI* activity was eliminated from the central region of the insert using *NsiI*, *XbaI* or *BglIII* generated fragments. However, a central *NcoI* fragment of about 2.2 kb was the smallest part of the clone able to produce both R and M enzymes. Thus sequencing was performed only on this *NcoI* fragment and some flanking areas, but not on the adjacent *NheI/HincII* fragment (see Fig. 1). The final identification of the correct start codon of *M.HgiBI* was performed parallel to sequencing in a series of active and inactive deletion variants with a resolution of 68 bp (data not shown).

Deletion of the methyltransferase led to non-viable constructs. As shown in Fig. 1 the 2431 bp *HindIII* fragment was not obtained in either orientation, that means either transcribed or not transcribed from an outside vector *p_{lac}* promoter. Surprisingly enough, the 2218 bp *NcoI* fragment was clonable in one orientation, when the *p_{lac}* promoter was upstream, but not in the reverse orientation. Thus, the *NcoI* fragment must contain both R and M enzyme, with only an endogenous promoter for the restriction enzyme expression and not with the promoter preceding the entire system.

DNA sequence analysis and genetic organization of *hgiBIRM*

After nucleotide sequence determination of the *ClaI* to *NheI* fragment (2938 bp, indicated in Fig. 1) two open reading frames (orfs) could be found, exactly fitting the functional data (see Fig. 2): The coding sequence of the methyltransferase *M.HgiBI*

	region A	region B	C	region D	E	
M.HgiBI	MQQFRFDLPAFIGGFRGLGLEAVG-----GVCVASAEIDQAIKVYRQNWPTDGVHDNLGDTAITQQLPAHDVLVGGVPCQPMWSIAGKNQAFDDPRGLWADVIR					100
M.BanI	MRKIFVDLPAGIGGIRIGFERAAKRFELETECVLSSEIDKKACETYALNF---KEEPQGDIEHITSPPFDFLLAGFPCCPPSYAGKQGGFGDTRGTLFFEVER					111
	region F	G	region H			
M.HgiBI	LVQINQPKAFIFENVKGLVDPRNRLCLEIILDSFKDLGYSVFYKLLNSFDGVAQNRRDRVFIVGI---QQKLDL---NGFSFPEYTESEORLYHILDNLEVPET					198
M.BanI	VLRDNRPKAFLENNVRLVTHDKGRTLKTIIISKLEELGYGVSYLLNSSTFGVPQNRVRIYILGLILGSKPKLTLSNVGAADSHKYKNEQISLFD--ESYATVKD					204
M.HgiBI	KLESIPQIRNLGERIDVGYNKLTFRG-AFNDFILNDIRNGPTSIHSWE--IYPTTEREKQICM-IIMNRRNSRYGDC-DGNFMSYQDIAELVAGLAELKELQT					298
M.BanI	ILEDSPSEKYRCSDEFIQLSKVGNFELLHGYYRLIDYRGG-NSIHSWELGIKGDCTKEEIEFLNQLIANRRKKIYGTQDGKALTLEQIRTF---YNHDQLEV					306
M.HgiBI	LVEKRILRQYPDGKYEFFNRRLSGGIDGTYRIFLFPNARFFGTLTARGMHDEIAEISVSANAEYKHNFIQQVLI PKRYRKITVSEARLQGGFPFSQFHSNQSA			region I	I	403
M.BanI	IIKSLQKQYLRREENKFN-PVCGNM--SFEVF---KFL-----DPDSISITLTSSDAHKLGVVQNN-VP---RRITPRECARLQGGFPDFILHSNDNF					389
	region J					
M.HgiBI	NFRLIGNSVAPPVIVALGKALQCCKLFEQELCEV*					437
M.BanI	AYKQLGNSVTVKVEKVIDELFQNNVNELFQGMKLANVV*					428

Figure 3. Selected comparison of primary structures between the methyltransferases *M.HgiBI* (GGWCC) and *M.BanI* (GGYRCC). The initial alignment was carried out using the FASTP algorithm of the PROSIS software package (Pharmacia/LKB, Freiburg). Both enzymes share 32% identical residues (indicated by double points) plus a high degree of related residues (marked by points; according to PROSIS rules). The positions of the ten conserved boxes, which are observed in all known prokaryotic m³C-methyltransferases (7, 8), are underlined. The most striking feature was the number of significant similarities throughout the so-called variable region separating block H and I, especially the conserved hexapeptide 'SIHSWE'.

starts at an ATG codon at position 423, which is preceded by a putative Shine-Dalgarno sequence overlapping with an *NcoI* site at position 408, and has a length of 437 amino acids coding for a protein with a calculated MW of 49,625 D. Separated by only one triplet from the *hgiBIM* stop codon and also preceded by a suitable translational start sequence follows the ATG of a shorter orf with identical transcriptional orientation (see Fig. 2, position 1740 to 2564), functionally characterized as the desired endonuclease R.*HgiBI* (274 amino acids; MW 31,193 D).

Sequencing disclosed potential promoter sequences similar to the *E. coli* consensus sequence: TTGACA (putative '−35' region; position 362 to 367) and TAAATA (putative '−10' region; position 384 to 390). This promoter was called p_{MR} and functionally characterized by active and inactive subclones. The endogenous promoter p_R was found upstream of the endonuclease [TTGGGT for a putative '−35' region (position 1680 to 1685) and TAAAATT for a putative '−10' region (position 1702 to 1708)], which overlaps the C-terminal coding region of *hgiBIM*. In addition to the above described functional analysis, we have characterized p_R as a fairly weak promoter using an 450 bp *PmaCI/NsiI* fragment, separately (data not shown).

As observed for several other R and M genes from different organisms (see reference 8) and like the previously described *H. giganteus* *Hpa2* genes *hgiDIRM* (8) and *hgiDIIM* (21) high A+T contents of 62 % for *hgiBIM* and 60 % for *hgiBIR* were calculated from the sequence.

Presumed additional modification activities in *H. giganteus* *Hpg5*

At position 218 (GTTAAC; compare Fig. 2) sequencing disclosed an additional *HincII* (= *HpaI*) restriction site, which had not been observed at this position in any Southern-blot analysis of genomic *H. giganteus* *Hpg5* DNA. However, *in vitro* cleavage of the cloned insert DNA from *E. coli* HB101/pRMB1-1 could be performed using either *HincII* or *HpaI*. A new methyltransferase specifically modifying the *HpaI* subset (GTTAAC) of the *HincII* (GTYRAC) sites in *Hpg5* DNA might be responsible for this phenomenon. Attempts to clone this modification enzyme are currently carried out in our laboratory.

Protein sequence comparisons

Computerized comparisons of the *HgiBI* enzymes between each other and with all known restriction endonucleases and methyltransferases confirm the features generally observed for type II systems. Significant homologies were neither detected between M.*HgiBI* and R.*HgiBI*, nor between R.*HgiBI* and any other restriction enzyme, including R.*HgiDI* (GRCGYC) from strain *Hpa2* of *H. giganteus* (8), or endonucleases with related or even identical specificities, like R.*Sau96I* (GGNCC) from *S. aureus* (18) and R.*SinI* (GGWCC) from *S. infantis* (19).

Comparative studies of the methyltransferase M.*HgiBI* revealed a perfect resemblance to the pattern of ten conserved boxes (compare Fig. 3) described for all prokaryotic m⁵C-modification methylases (7, 8). Additional homologies in the so-called variable region separating block H and I (compare Fig. 3) were not observed with M.*SinI* (19) or M.*Sau96I* (18), but only with the non-isomethylomeric M.*BanI* (GGYRCC) from *B. aneurinolyticus* (10). It was most surprising to find an identical hexapeptide within M.*BanI* (SerIleHisSerTrpGlu; at position 242 in M.*HgiBI*, Fig. 3). This motif does not occur within any other

known R-M enzyme sequence. During database searches the identical motif was neither found in the PIR-NBRF protein data library (release 22.0) nor in the SWISS-PROT database (r12.0). Interestingly, the most similar sequence containing only one mismatch (IIHSWE) was found in a DNA polymerase of the yeast *Kluyveromyces lactis* (22), but in no other protein.

CONCLUSIONS

Since the above identified SIHSWE motif is located within the long variable region between block H and I (8), which corresponds to the domains of recognition specificity within multispecific m⁵C-methyltransferases from *B. subtilis* bacteriophages, we like to speculate, that this motif may be involved in sequence recognition as well (23). Our future work will focus on the question, whether these additional homologous regions play any role in recognition of the common GG- or CC-dinucleotides within the independent target sequences of *HgiBI* and *BanI*.

ACKNOWLEDGEMENTS

We like to thank Prof. Dr. H. Reichenbach (GBF Braunschweig) for his kind provision of the *Herpetosiphon giganteus* strain *Hpg5* and D. Möstl for technical assistance. This work is a partial fulfillment of the requirements for Ph.D. thesis of A. Düsterhöft (Justus-Liebig-Universität Giessen, 1990). It has been supported by the Deutsche Forschungsgemeinschaft (Grant Kr 591/2-2,3).

REFERENCES

1. Wilson, G.G. (1988) *Gene* **74**, 281–289.
2. Wilson, G.G. (1988) *Trends Genet.* **4**, 314–318.
3. Kessler, Ch. and Manta, V. (1990) *Gene* **92** 1–248
4. Rosenberg, J.M. (1991) *Current Opinion in Struct. Biol.* **1**, 104–113.
5. Chandrasegaram, S. and Smith, H.D. (1988) In Sarma, M.H. and Sarma, R.H. (eds.) *Structure and Expression Vol. 1, From Proteins to Ribosomes*, Adenine Press, Schenectady, NY, pp 149–156.
6. Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
7. Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
8. Düsterhöft, A., Erdmann, D., and Kröger, M. (1991) *Nucleic Acids Res.* **19**, 1049–1056.
9. Kröger, M., Hobom, G., Schütte, H. and Mayer, H. (1984) *Nucleic Acids Res.* **12**, 3127–3141.
10. Maekawa, Y., Yasukawa, H. and Kawakami, B. (1990) *J. Biochem. (Tokyo)* **107**, 645–649.
11. Mayer, H. and Reichenbach, H. (1978) *J. Bacteriol.* **136**, 708–713.
12. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
13. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
14. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Kröger, M. and Kröger-Block, A. (1984) *Nucleic Acids Res.* **12**, 113–120.
17. Roberts, R.J. (1990) *Nucleic Acids Res.* **18**, r2331–2365.
18. Szilak, L., Venetianer, P. and Kiss, A. (1990) *Nucleic Acids Res.* **18**, 4659–4664.
19. Karreman, C. and De Waard, A. (1988) *J. Bacteriol.* **170**, 2527–2532.
20. Szomolanyi, E., Kiss, A., and Venetianer, P. (1980) *Gene* **10**, 219–225.
21. Düsterhöft, A. and Kröger, M. (1991) *Gene*, submitted for publication.
22. Tommasino, M., Ricci, S., and Galeotti, C.L. (1988) *Nucleic Acids Res.* **16**, 5863–5878.
23. Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305–312.