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

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 HgiBl 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 ClaI/HincII fragment in the same transcriptional orientation separated by one tripeptid only. Protein sequence comparisons revealed a close resemblance of M.HgiBl to the group of m^5^C-methyltransferases, especially to M.Banl from *Bacillus aneurinolyticus* with the related recognition sequence GGYRCC. In contrast, no significant similarities have been observed for the associated endonuclease R.HgiBl with any other restriction enzyme described so far, even not with the isoschizomeric R.Sinl from *Salmonella infantis*, or with R.Banl.

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 [EcoRI and EcoRV (4)], but none for a methyltransferase are known. On the other hand, a fair number of restriction-modification system are cloned and sequenced, and especially the m^5^C-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 HgiBl (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.Banl 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 (hsdR1` hsdM1`, mcrA`, mcrB`) was taken from our strain collection. Plasmid pBluescriptII SK(+) 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). [\textsuperscript{35}S-\alpha]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-trimethylammoniumbromid/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 *hgiCHM* of *H. giganteus* *Hpg9* was used for immunochemical signal detection. Different DNA fragments hybridizing with the methyltransferase *hgiCHM* insert were identified. A suitable *ClaI/HinClI* 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 \(\mu\)g of the isolated *H. giganteus* *Hpg5* DNA fragment region was ligated with T4 DNA Ligase into 0.3 \(\mu\)g *ClaI/HinClI* cleaved pBluescriptII SK(+) vector DNA at 17\(^\circ\)C for 16 hr in 20 \(\mu\)l 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 \(\mu\)g/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 \(\mu\)g DNA of the partial library was completely digested with 36 units of the commercially available *HgiBl* isoschizomer *Avall* (New England Biolabs). After transformation into competent *E. coli* HB101, plating onto Ap-LB agar plates and overnight incubation at 37\(^\circ\)C, mini-prep plasmid DNAs of individual surviving colonies were isolated and analyzed for *in vitro* *Avall* endonuclease resistance.

**Detection of restriction endonuclease *R.HgiBl* 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\(_2\)EDTA, 10 mM 2-mercaptoethanol, 0.01 % Triton X-100, 1 mg/ml lysozyme.

---

**Figure 1. Genetic organization of the R-M system *HgiBl*.** The plasmid pRMB1-1 containing a *ClaI/HinClI* insert of 3.6 kb (indicated as ‘cloned’ area) was selected by *Avall* overdigestion of a size restricted *H. giganteus* *Hpg5* library in *E. coli* HB101. DNA sequencing was done on both strands of the *ClaI/HinClI* 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^MR* promoter. Single headed arrows indicate orientation dependent activity with only one viable construct. The failure to subclone the central *HinClI* segment in any orientation was interpreted as the self-destruction of a *hgiBIR* \(^+\) phenotype with *R.HgiBl* being expressed from its own promoter. \(*\) denotes an *HinClI* site uncleaved in genomic DNA (see text).
Disruption of cell walls was achieved by sonication for 1 min. Cellular debris and unsoluble material was removed by 20 min centrifugation at 18000 rpm. 1, 2.5 and 5 µl of the supernatant were tested for their ability to cleave 0.3 µg of suitable plasmid DNAs in 20 µl reaction mixture containing 10 mM Tris.HCl pH 8.0, 10 mM MgCl\(_2\), 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).

---

**Figure 2.** Nucleotide sequence of the 2938 bp ClaI/Hind segment encoding RM.HgiBl and predicted primary structures of both enzymes (M: position 423 to 1736; R: 1740 to 2564). Putative ‘—10’ and ‘—35’ promotor sequences are underlined within the DNA sequence. Potential Shine–Dalgarno sequences are marked by (*).
RESULTS AND DISCUSSION

The DNA sequence GGWCC 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 Alu from 23 different organisms are characterized. In addition 52 isoschizomers with the slightly higher specific Ava II 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 GGWCC recognition system one representative has been sequenced, Sau96Al (18) and SinI (19). We isolated and characterized a second system of the GGWCC type, namely HgiBl, as described below.

Identification and cloning of the genes encoding the R-M system HgiBl

From H.giganteus Hpg5 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, Clal, HincII, or Clal/HincII, data not shown). Among these hybridizing DNA fragments was a Clal/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.HgiBl containing clone, with the expected 3.6 kb Clal/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 retransformatino into E.coli HB101 8 out of 20 analyzed colonies contained plasmids mediating full resistance against R.Avail cleavage, while other (e.g. AhalII 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 HgiBl R-M system and the respective plasmid was consequently named pRMBI-1. This immediate success is, however, in contrast to the HgiCIIM 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 pl BluescriptISK(+) derivative pRMB-1 containing the entire Clal/HincII insert allowed the preliminary localization of the genes coding for RM. HgiBl as given in Fig. 1. For M.HgiBl activity a Clal/SinI generated 1904 bp DNA segment spanning the left half in Fig. 1 was sufficient. It exhibited methyltransferase activity orientation independently. Therefore a promoter must be located within this fragment. The R.HgiBl activity was eliminated from the central region of the insert using SinI, XbaI or BglII 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 Nhel/HincII fragment (see Fig. 1). The final identification of the correct start codon of M.HgiBl 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 pGem promoter. Surprisingly enough, the 2218 bp NcoI fragment was clonable in one orientation, when the pGem 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 Clal to Nhel 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.HgiBl region F and region G are located in the same strand, while region A and region B are located in opposite strands, respectively.
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.HgiB (274 amino acids; MW 31,193 D).

Sequence disclosed potential promoter sequences similar to the E.coli consensus sequence: TTGACA (putative ‘−35’ region; position 362 to 367) and TAAAATA (putative ‘−10’ region; position 384 to 390). This promoter was called pMR and functionally characterized by active and inactive subclones. The endogenous promoter pR 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 pR as a fairly weak promoter using an 450 bp PmaCl/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 hgiBIM (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 (GGTAAAC; compare Fig. 2) sequencing disclosed an additional HindII (= 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/pRMBl-1 could be performed using either HindII or HpaI. A new methyltransferase specifically modifying the HpaI subset (GGTAAAC) 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 HgiBl 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.HgiBl and R.HgiBl nor between R.HgiBl and any other restriction enzyme, including R.HgiBl (GGCGYC) from strain Hpa2 of H.giganteus (8), or endonucleases with related or even identical specificities, like R.SacI (GGGCC) from S. aureus (18) and R.SimI (GGWCC) from S. infantis (19).

Comparative studies of the methyltransferase M.HgiBl revealed a perfect resemblance to the pattern of ten conserved boxes (compare Fig. 3) described for all prokaryotic mC-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.SimI (19) or M.SacI (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.HgiBl, 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 (IHIHSE) 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 mC-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 HgiBl 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