Stepwise cloning and molecular characterization of the HgiDI restriction-modification system from Herpetosiphon giganteus Hpa2

Andreas Düsterhöft, Dirk Erdmann and Manfred Kröger*
Institut für Mikrobiologie und Molekularbiologie der Justus-Liebig-Universität Giessen, FB 15 Frankfurter Strasse 107, D-6300 Giessen, FRG

Received December 10, 1990; Revised and Accepted January 28, 1991

EMBL accession no. X55140

ABSTRACT

The restriction-modification system HgiDI from Herpetosiphon giganteus strain Hpa2 has been cloned in E. coli in a two-step procedure. Selection of the methyltransferase (M.HgiDI) gene in vitro was performed using the heterologous restriction endonuclease AhaW, an Isoschizomer of Acyl and HgiDI (GRCGYC). Cloning of the complete HgiDI endonuclease (R.HgiDI) gene could only be achieved in recipient cells harbouring a recombinant plasmid, which was expressing the corresponding methyltransferase and could thereby prevent the host from self-destruction of its genetic material. The HgiDI restriction-modification system was sequenced and functionally correlated with two open reading frames of 309 (M) and 359 (R) codons. In homology studies M.HgiDI showed significant similarities to 20 other m\(^5\)C-methyltransferases and turned out to be the most compact enzyme of this group described so far. Initial attempts for overexpression of M.HgiDI and partial purification of R.HgiDI have been successful.

We have previously characterized a group of restriction enzymes derived from the soil gliding bacterium Herpetosiphon giganteus (23). A compilation of the recognition sequences of almost all enzymes known from this organism led us to the model, that they might have evolved from a common ancestor (23). Working on the H. giganteus enzyme group was especially promising, since many of these systems showed degenerated, overlapping specificities. Therefore we have started an approach to clone some of the H. giganteus restriction-modification systems, to begin an analysis of the enzyme repertoire of one organism on the molecular basis.

In this paper we will report on cloning of RM.HgiDI (GPuCGPyC) as another restriction-modification system, which, similar to RM.Ddel (24) and RM.BamHI (25), could not be obtained in one cloning step. Sequence determination, gene expression data and the results of the comparison of the HgiDI restriction-modification system with other known type II systems will be presented.

INTRODUCTION

Type II restriction-modification systems are of increasing commercial and scientific interest, since they are both valuable tools and model systems to study the molecular basis of protein nucleic acids interaction. Several most interesting questions have been raised. An increasing number of different R-M systems has been cloned (1) and further analyzed in their primary structure (2–20), but it is, for example, still unresolved, how methyltransferase and corresponding restriction enzyme manage to recognize the identical nucleotide sequence, although only few homologies between their amino acid sequences have ever been detected (21). Similarly, it is still in question, how different methyltransferases recognize identical or related sequences with a minimum of sequence homology (22), and it is unknown, whether and how different specificities within one and within different organism might have evolved from each other.

MATERIALS AND METHODS

Bacterial strains, plasmids and phages

Herpetosiphon giganteus Hpa2 was a gift of H. Reichenbach (GBF, Braunschweig). E. coli HBl01 (hsdR\(^-\) hsdM\(^-\), mcrA\(^+\) mcrB\(^-\)) was taken from our strain collection. Standard M13mp18 and mp19 phages were grown on E. coli TG1, obtained from Amersham (Braunschweig). Lambda DNA was purchased from BRL (Eggenstein). Plasmid pOM8 is a multilinker variant of pUC18 kindly provided by I. Oberbäumer (26). pHK255 (amp\(^R\)), pHK274 (amp\(^R\)) and pHK278 (tet\(^R\)) are constructed in our laboratory as expression level controlled vectors with the lac\(^Q\) gene, the synthetic tac promotor and terminators derived from rmb of E. coli: pHK255 and pHK274 are multilinker variants of pIF118 (27), while pHK278 additionally carries the tetracycline resistance gene of pBR322 instead of the ampicillin resistance gene.

* To whom correspondence should be addressed
Media

Herpetosiphon giganteus cultivation was performed as described by Mayer und Reichenbach (28) 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 (29).

Enzymes and chemicals

Restriction enzymes, T4 DNA Ligase and Klenow polymerase were obtained from various distributors (AGS, Heidelberg; Boehringer, Mannheim; New England Biolabs; Schwalbach). The ‘Erase-a-base’ kit for exonuclease III deletion of plasmids (30) was purchased from Promega (Heidelberg). Chain termination sequencing was done using T7 DNA polymerase from Pharmacia/LKB (Freiburg). [32P-α]dATP (> 650 Ci/m mole) 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 31. Plasmid DNA isolation was performed by mini-prep alkaline-SDS procedure or as a maxi-prep followed by CsCl/ethidiumbromide ultracentrifugation (29).

Transformation

E. coli transformations were carried out according to the Hanahan protocol (32).

Construction of H. giganteus genomic DNA libraries

About 2 μg of purified H. giganteus Hpa2 DNA were digested with HindIII and were ligated with T4 DNA ligase to 0.5 μg HindIII cleaved, dephosphorylated pOM8 vector DNA at 17°C for 16 hr in 50 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 mg/ml ampicillin (Ap). Following an overnight growth period about 6000 colonies obtained were scraped together in 5 ml of 10 mM Tris.HCl pH 7.5, 10 mM MgCl₂, 1 mM EDTA, 5 mM 2-mercaptoethanol. 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.

DNA sequence determination

DNA sequencing was done from both directions either from fragments directionally deleted by exonuclease III/S1 nuclease (30) or using internal oligonucleotides as synthetic primers following the chain termination protocol (34), except that T7 DNA polymerase was used. All oligonucleotides were kindly synthesized by Dr. J. Hegemann from our laboratory on an ABI Model 380B DNA synthesizer.

Selection of methyltransferase clones

1 μg of the library DNA was completely digested with 30 units of the commercially available HgiDI isoschizomer AhaII (New England Biolabs). After transformation into competent E. coli HB101, plated onto Ap-LB agar plates and overnight incubation at 37°C, mini-prep plasmid DNAs of individual surviving colonies were purified and analyzed for AhaII endonuclease resistance.

Cloning of adjacent genomic DNA fragments

Southern-transfer of H. giganteus DNA, digested with various restriction enzymes and run on agarose gels, was performed according to standard procedures (29), except that the nonradioactive Digoxigenin-dUTP labelling kit (Boehringer, Mannheim) was used for the preparation of a DNA probe from the previously cloned fragment and for the immunochemically signal detection. Fragments overlapping with the 3 kb HindIII methyltransferase insert were identified and used to construct a partial genomic map. Certain fragment regions were isolated from agarose gels (29) and ligated into M13mp18/19 replicative form DNA at corresponding sites. After transfection of E. coli TG1 cells, plaques grown overnight were identified by nonradioactive in situ hybridization (33) and subsequently sequenced.

Detection of restriction endonuclease activity in vitro

20 ml liquid culture of a potentially endonuclease expressing clone were 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 Na₂EDTA, 10 mM 2-mercaptoethanol, 0.01% Triton X-100, 1 mg/ml lysozyme. Disruption of cell walls was achieved by sonication for 1 min. Nucleic acids and unsoluble material was removed by 20 min centrifugation at 18000 rpm.

Partial purification of endonuclease R.HgiDI

Cloned HgiDI restriction endonuclease was partially purified from E. coli HB101 carrying the expression plasmid pHK575. 10 ml Ap-LB medium were inoculated with 0.5 l of a pHK575/HB101 culture and grown to an OD₅₀₀ of 0.7. For tac promotor induction IPTG was added to a final concentration of 1 mM and incubation was continued to an OD₅₀₀ of 2.2. 33 g of wet cell mass were obtained after centrifugation. Cells were resuspended in an equal volume of buffer A (50 mM Tris.HCl pH 7.5, 5 mM EDTA, 1 mM 2-mercaptoethanol, 1 mM protease inhibitor phenylmethylsulfonylfluoride). Disruption was achieved by a French press and the soluble fraction was separated by ultracentrifugation for 40 min at 45000 rpm. Nucleic acids were removed by careful addition of streptomycin sulfate to a final concentration of about 5%, followed by 40 min centrifugation at 45000 rpm. The supernatant was loaded onto a Heparin-Sepharose column (30 ml bed volume, Pharmacia/LKB) and bound proteins were eluted with a linear gradient of 0 to 1 M KC1 in 300 ml of buffer A. Individual fractions were assayed for HgiDI specific endonuclease activity in vitro, analogously to the crude extracts from above. The mean endonuclease activity eluted in a range of 0.3 to 0.5 M KC1. Seven of these fractions were pooled, dialyzed against buffer B (50 mM Tris.HCl pH 8.0, 5 mM EDTA, 5 mM 2-mercaptoethanol) and loaded onto a Mono Q FPLC column (Pharmacia/LKB). R.HgiDI did not bind to the column, but could be detected in the flowthrough and was thereby separated from most of the contaminating proteins. After dialysis against storage buffer (10 mM Tris.HCl pH 7.4, 50 mM KC1, 1 mM EDTA, 5 mM 2-mercaptoethanol, 60% glycerol) 19.5 ml of enzyme volume could be stored at 4°C.

Endonuclease assays of purified R.HgiDI

The R.HgiDI activity in vitro was tested by the incubation of the enzyme with 1 μg lambda DNA at 37°C for 1 hr in 50 μl
Nucleic Acids Research, Vol. 19, No. 5

reaction mixture containing 10 mM Tris.HCl pH 8.0, 10 mM MgCl₂ and 1 mM 2-mercaptoethanol.

**Computer analysis**

Analysis of sequence data was carried out using the DNASIS and PROSIS software (Version 5.02; Pharmacia/LKB) and the FLEXP program (35).

**RESULTS AND DISCUSSION**

The cloning strategy for hgiDIM/hgiDIR was based on the in vitro selection protocol of specifically self-methylating plasmids (36). In many cases this method turned out to be useful both for cloning of the methyltransferases and for the search of the appropriate endonucleases, which are mostly encoded in neighbouring genes (37). We modified the procedure by using a heterologous restriction enzyme of the same specificity as R.HgiDl rather than the native H. giganteus enzyme for the selection step, as the native enzyme could not directly be purified in sufficient concentration and quality.

**Cloning of the gene coding for M.HgiDl**

A genomic HindIII generated library of H. giganteus Hpa2 DNA was constructed in E. coli HB101, paying attention to the mcrB⁻ phenotype of this strain (38). The vector pOM8 (26) contained three restriction sites for R.HgiDl. Plasmid DNA,
isolated from this library, was completely digested with the R.HgiDl isochizomeric endonuclease AhaU. Following retransformation into E. coli HB101, the analysis of 48 transformants resulted in the identification of two clones, which contained recombinant plasmids, specifically resistant both against AhaII (GR/CGYC) and NarI (GG/CGYC) endonucleases. AhaU (GR/CGYC) and HgiDl show any endonuclease activity cultures as described in the METHODS section. They did not influence the plasmid pMDl-2 (also Fig. 1A) did not show any endonuclease activity. Using M13mpl8 as cloning vector in fact led to the identification and isolation of the desired fragment within this 4.5 kb fragment.

Therefore, we decided to clone ORF A in a truncated form (ampR, hgiDIR+ M+). DNA sequencing and functional localization of hgiDIR and surrounding genes within pMD1-1

The construction of exonuclease III deletion variants from both orientations of the 3 kb HindIII insert allowed functional characterization parallel to sequence determination of individual subfragments. Two incomplete open reading frames (ORF A, 335 codons, and ORF C, 319 codons) flanking one internal complete reading frame (ORF B, 309 codons) could be found. They all show the same orientation and were finally designated to position 746 to 3681 in Figure 2. Since the exoneuclease III deletion of nucleotides from ORF B resulted in the loss of AhaII resistance, we concluded that this orf is the desired methyltransferase gene hgiDIM. The identification of the start and stop codons of hgiDIM was performed using a series of methyleating and non-methyleating subclones with a resolution of 64 and 48 bp (data not shown). As indicated in Figure 2, ORF C turned out to contain 39% identical positions in comparison to the first 320 amino acids of the metB gene product of E. coli, an enzyme of the S-Adenosylmethionine biosynthesis pathway (39). In contrast to this, the incomplete ORF A did not show any significant DNA or protein homologies within the same data bank search, and thus was the most probable candidate for hgiDIR.

**Cloning of the hgiDIR endonuclease gene**

In order to find overlapping, prolonging DNA fragments within the H. giganteus Hpa2 chromosomal DNA a Southern-blot analysis using the cloned 3 kb HindIII fragment was performed. As expected, several hybridizing fragments were detected (data not shown). Unexpectedly, the most promising fragment found, a BglII fragment of 4.5 kb, which might have contained enough nucleotides to code for the complete, functional HgiDI restriction-modification system, could by no attempt be cloned on any plasmid system available in our laboratory. We interpreted this result as a strong hint for the existance of the endonucleolytic activity within this 4.5 kb fragment.
Figure 3. Overexpression of M. HgiDI. Panel A shows the construction of plasmids for overexpression and functional characterization of M. HgiDI. In plasmid pHK540 a 1.95 kb SmaI/HindIII fragment derived from an exonuclease III deletion variant of pMD1-2, in which 27 bp upstream of the hgiDIM start codon were left, was recombined with the lac promoter/operator region of expression vector pHK255. As a control plasmid pHK545 was built up from the 1.05 kb SmaI/HindIII fragment of a hgiDIM~ deletion variant starting with nucleotide 2628 in the C-terminus of hgiDIM. Panel B shows a Coomassie stained 12.5% SDS-PAA gel separating proteins from IPTG induced pHK540 E. coli HB101 and pHK545 E. coli HB101 cells. The order of lanes is: 1, molecular weight marker; 2, pHK540/HB101 without induction; 3 to 6, pHK540/HB101 induced with 1 mM IPTG for 30, 60, 120 and 240 min; 7, pHK545/HB101 without induction; 8 to 10, pHK545/HB101 induced with 1 mM IPTG for 60, 120 and 240 min. Arrows indicate the position of two induced proteins. Since the lower band appears in both cell types, it must be coded for in the truncated ORF C, while the disappearing band must correspond to M. HgiDI.

In order to estimate the lethality during the establishment of the RM. HgiDI system in E. coli in vivo, a quantitative transformation assay was performed. For this purpose a frameshift mutation was introduced by filling-in the HindIII site within hgiDIR by Klenow polymerase reaction. The resulting control plasmid pRMD1antiR lost all endonucleolytic activity. Equal amounts of pRMD1anti DNA as well as pRMD1antiR DNA were transformed each either into empty E. coli HB101 or into pHK560 HB101 competent cells. Table 1 shows the result of this experiment. While transformation of empty E. coli HB101 with the control plasmid pRMD1antiR resulted in a high rate of colony formation, no transformants were obtained with pHK560 HB101. In contrast to this, premethylated pHK560 HB101 recipient cells were transformed by both plasmids independently with the same efficiency. Obviously, a functional ORF A was responsible for the observed lethal effects and the gene was consequently renamed hgiDIR.

However, it is not known now, whether this observation might be limited to the heterologous host E. coli or whether it may be a hint on a specific regulation system in H. giganteus. In any case, it may well serve as another example for only seemingly unclonable endonucleolytic activities (37).

Stabilized expression of RM. HgiDI

Reversion of the insert within the plasmid pRMD1 (see Fig. 1C) led to an increased expression of the encoded endonuclease and methyltransferase genes from the lac promoter of pOM8. The
R. HgD1 activity in crude extracts of pMRD1/HBI01 was slightly higher than obtained from pRMD1anti/HBI01, accompanied by a decreased growth rate of transformed E. coli cells (data not shown). In order to improve the stability of the R-M system in E. coli, we had to delete the region upstream of the characteristic TaqI restriction site (see Fig. 1C & D; position 632 Fig. 2). Simultaneously the new vector system replaced the lac promoter by the tac promoter/lac operon region of plasmid pHK274 to increase the rate of expression. This procedure deleted a possible promoter upstream of the endonuclease gene (see E. coli +10' consensus TATAATT at position 621 in Fig. 2), but still allowed a weak expression of the methyltransferase even in the uninduced state from its proposed own promoter (ATTAAAT at position 1663 in Fig. 2) in pHKS75/HBI01 cells. IPTG induction resulted in a fairly high expression of R. HgD1, and we were able to gain enough material to partially purify the restriction enzyme. About 1800 U/g wet weight of cells were obtained. Optimizing buffer conditions for R. HgD1 disclosed a strong preference for low ionic strength (0−50 mM NaCl) and basic pH values (8−9.5) in a wide temperature range from 20 to 45°C (data not shown).

### Construction of an M. HgD1 overexpressing plasmid

For an endonuclease independent overexpression of the methyltransferase M. HgD1 the plasmid pHK540 (Fig. 3A) was constructed. It is based on our standard expression vector pHK255 and contains a Small/HindIII fragment of 1.9 kb derived from an M. HgD1 expressing exconuclease III deletion variant of pMD1-2. As a control plasmid pHK545 (see also Fig. 3A) was used, which is derived from a methyltransferase deficient variant of pMD1-2. As shown in Figure 3B two additional highly expressed proteins of about 33 and 35 kD could be observed after IPTG induction of pHK540/HBI01 cultures in denatured extracts on Coomassie stained SDS-PAA gels, whereas in the control cells pHK545/HBI01 the 33 kD product was induced, only. This identifies the 35 kD product as M. HgD1, in good agreement with the calculated size of 34,444 D. Consequently, the second
Figure 5. Schematic presentation of the localization and length of conserved and non-conserved regions within 21 m^3C-methyltransferases. Regions of homology are indicated by ten boxes (A to J) filled with different patterns, whereas unfilled areas represent generally non-homologous parts of the enzymes. Black rectangles within the variable regions of multispecific m^3C-methyltransferases show the localization of so-called target recognizing domains (22). The M.HgiDl characterized in this work is clearly the smallest m^3C-methyltransferase.

overexpressed 33 kD protein derives from ORF C, even though there is a slight difference to the calculated molecular weight of 34,678 D.

DNA sequence interpretation
Like several other R-M systems (2, 4–6, 9, 10, 13, 14, 16) the RM.HgiDI genes turned out to be high in A+T content (58% in M, 66% in R). The codon usage reveals an unusual amount of codons (30% for M, 33% for R), which are rarely (3.6%) used in strongly expressed E.coli proteins, as taken from reference 40. In contrast to the R-M genes, ORF C, has got rather normal values of 45% A+T and 20% of seldomly occurring codons. These data are in good accordance to the potentially corresponding E.coli meiB gene, revealing 48% or 23%, respectively (39).

The two reading frames of hgiDIR and hgiDIM are connected by an ATGA start/stop codon overlapping module. This was found initially for the lambda nin genes connecting tightly regulated genes and was recently used as TGATG module for an artificially connected gene expression system (56,57).

No significant similarities were found between the amino acid sequences of HgiDl methyltransferase and restriction endonuclease. The same result was observed, when R.HgiDl DNA and protein sequence were compared with 20 published sequences of type II endonucleases (2–20). This observation seems to represent a general feature within bacterial R-M systems.

Since the Narl GGCGCC recognition sequence was also protected by M.HgiDl methylation, the product had to be either nV'C or rrr'C. As expected, M.HgiDl shares some obvious homologies with one of these groups, namely the m^5C-methyltransferases. A careful comparison of M.HgiDl with 20 of these enzymes (Fig. 4) showed, that it fits into the previously described pattern of ten conserved regions (41). Among this type of enzymes M.HgiDl (309 aa) now represents the most compact.
The long variable region between block H and I, which in the case of multispecific methyltransferases contains the sequence specificity domains (42), consists of only 78 amino acids (compared to 81 in M. MspF I and 84 in M. BsuF I, but 193 in M. Ban I; for schematic representation see Fig. 5). Although special interest was focused on the comparison of this variable region between methyltransferases of different substrate recognition, no significant homology could be found, even not with M. Ban I (GGYRC) which shares the Nari recognition sequence with M. HgiD I. Additional work will be carried out to analyze the relationship of the HgiD I enzymes to the wide spectrum of R-M systems in Herpetosiphon giganteus.

ACKNOWLEDGEMENTS

We thank Prof. Dr. H. Reichenbach (GBF Braunschweig) for the provision of Herpetosiphon giganteus strains and his advice on cultivation. Special thanks have to go to Dr. W. Röwekamp (AGS GmbH Heidelberg) for his support in enzyme purification. This work is a partial fulfillment of the requirements for PhD thesis of A. Düsterhöft (Justus-Liebig-Universität Giessen, 1990) and has been supported by the Deutsche Forschungsgemeinschaft (Grant Kr 591/2-2,3).

REFERENCES