Stepwise cloning and molecular characterization of the *Hgi*DI restriction-modification system from *Herpetosiphon giganteus* *Hpa*2

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 *Hgi*DI from *Herpetosiphon giganteus* strain *Hpa*2 has been cloned in *E. coli* in a two-step procedure. Selection of the methyltransferase (*M.Hgi*DI) gene *in vitro* was performed using the heterologous restriction endonuclease *A*halI, an isoschizomer of *A*cyl and *Hgi*DI (GRCGYC). Cloning of the complete *Hgi*DI endonuclease (*R.Hgi*DI) 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 *Hgi*DI restriction-modification system was sequenced and functionally correlated with two open reading frames of 309 (M) and 359 (R) codons. In homology studies *M.Hgi*DI showed significant similarities to 20 other m⁵C-methyltransferases and turned out to be the most compact enzyme of this group described so far. Initial attempts for overexpression of *M.Hgi*DI and partial purification of *R.Hgi*DI have been successful.

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.

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.*Hgi*DI (GPuCGPyC) as another restriction-modification system, which, similar to RM.*Dde*I (24) and RM.*Bam*HI (25), could not be obtained in one cloning step. Sequence determination, gene expression data and the results of the comparison of the *Hgi*DI restriction-modification system with other known type II systems will be presented.

MATERIALS AND METHODS

Bacterial strains, plasmids and phages

Herpetosiphon giganteus *Hpa*2 was a gift of H. Reichenbach (GBF, Braunschweig). *E. coli* HB101 (*hsdR*_B[−] *hsdM*_B[−], *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 lacI^Q gene, the synthetic tac promoter and terminators derived from *rnmB* of *E. coli*: pHK255 and pHK274 are multilinker variants of pJF118 (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). [³⁵S-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 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 *Hind*III and were ligated with T4 DNA ligase to 0.5 µg *Hind*III 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 µg/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 ml of this library was used to inoculate 200 ml of Ap-LB medium for the isolation and CsCl purification of plasmid DNA.

Selection of methyltransferase clones

1 µg of the library DNA was completely digested with 30 units of the commercially available *Hgi*DI isoschizomer *Aha*II (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 purified and analyzed for *Aha*II 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 *Hind*III 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.

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.

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. Cellular debris and insoluble 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 lambda DNA in 20 µl reaction mixture containing 10 mM Tris.HCl pH 8.0, 10 mM MgCl₂, 1 mM 2-mercaptoethanol at 37°C for 1 hr.

Partial purification of endonuclease *R.Hgi*DI

Cloned *Hgi*DI restriction endonuclease was partially purified from *E. coli* HB101 carrying the expression plasmid pHK575. 10 l Ap-LB medium were inoculated with 0.5 l of a pHK575/HB101 culture and grown to an OD₆₆₀ of 0.7. For tac promoter induction IPTG was added to a final concentration of 1 mM and incubation was continued to an OD₆₆₀ 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, 5 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 KCl in 300 ml of buffer A. Individual fractions were assayed for *Hgi*DI 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 KCl. 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.Hgi*DI 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 KCl, 1 mM EDTA, 5 mM 2-mercaptoethanol, 60% glycerol) 19.5 ml of enzyme volume could be stored at 4°C. Purification effects of different steps were controlled on 12.5% SDS polyacrylamide gels (29).

Endonuclease assays of purified *R.Hgi*DI

The *R.Hgi*DI 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

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.HgiDI* 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.HgiDI*

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.HgiDI*. Plasmid DNA,

Figure 1. Construction of plasmids for functional studies of the *hgiDI* R-M system in *E. coli*. (A) The plasmids pMD1-1 and pMD1-2 contain the 3 kb *HindIII* insert obtained after *AhaII* *in vitro* selection in the original and the reversed orientation. (B) Construction of the premethylation plasmid pHK560 by insertion of the 2 kb *EcoRI/HindIII* fragment from pMD1-1 behind the strong *tac* promoter of vector pHK278 (*tet*^R). (C) In order to construct pRMD1anti for technical reasons a variant deleting the second *HindIII* site and minor parts of ORF C in pMD1-1 (not shown) was recombined with the 750 bp *HindIII/SalI* fragment from the double stranded mp18BE1.6 phage DNA to complete the *HgiDI* system. The 3.7 kb *SmaI/XbaI* fragment from pRMD1anti was fused to a *XbaI/HindIII* deleted pOM8 vector to construct pRMD1, which now bears *hgiDIRM* transcribed by the *lac* promoter. The plasmids pRMD1anti (*amp*^R) and pRMD1 (*amp*^R) could only be built up in *E. coli* HB101 host cells prematurely methylated by *M.HgiDI* expressed from plasmid pHK560 (*tet*^R). (D) For a stable maintenance in *E. coli* and a controlled endonuclease expression, plasmid pHK575 was constructed using the *HgiDI* R-M system derived from pRMD1 as a *TaqI/PstI* insert and pHK274 as vector. This both replaces a 631 bp fragment upstream of the *TaqI* restriction site in the 5'-untranslated region and changes the vector system from pOM8 with the *lac* promoter to pHK274 providing now the *tac* promoter/operator segment.

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 *Sma*I/*Hind*III 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 *tac* promoter/operator region of expression vector pHK255. As a control plasmid pHK545 was built up from the 1.05 kb *Sma*I/*Hind*III 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*.

ATG at position 676 in Figure 2, preceded by a perfect *E. coli* Shine–Dalgarno sequence.

Attempts to directly complete the putative RM.*HgiDI* system with the newly cloned fragment were again unsuccessful. We assumed a misregulation of endonuclease and methyltransferase expression in *E. coli* similar to the behavior of RM.*DdeI* (24) and RM.*Bam*HI (25) and applied a two-plasmid system. This should allow a specific premethylation of the recipient *E. coli* HB101 chromosomal DNA, before the complete R-M system was expressed from a second plasmid.

The first plasmid was constructed by fusing the tetracycline resistance mediating vector pHK278 with the 2 kb *Eco*RI/*Hind*III fragment of pMD1-1 yielding pHK560 (see Fig. 1B). Competent pHK560/HB101 cells were now used as recipients for the second plasmid pRMD1anti (for constructional details see Fig. 1C). Since this vector carries ampicillin resistance, screening could now be performed with two antibiotics to identify cells containing both pHK560 and pRMD1anti. Crude extracts from pRMD1anti/pHK560/HB101 liquid cultures showed a low, but detectable *HgiDI* restriction endonuclease activity on lambda DNA *in vitro* (data not shown). Subsequent selection on ampicillin only led to an apparent loss of the tet^R plasmid pHK560. Obviously, premethylation was only necessary for the establishment of the RM.*HgiDI* system in *E. coli* HB101, but not essential for the maintenance of this system within the heterologous host. However, minor negative effects on the growth rate and general viability of the pRMD1anti/HB101 cells were observed (data not shown).

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 *Hind*III 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 pRMD1anti DNA in 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

name	specificity	references	region A	region B	region C	region D	region E			
M.Aquia	mCYCGRG	(43)	4 KLISLFSGAGGMDIGFHAAG	27 AVAVEQDPSCCNTLRN	52 GDITS	74 DLVIGGPPCQSFSLAGKR	95 DDPGRMLVL			
M.Bani	GGYRCC	(19)	4 KFVDLFAIGIGIRIGFERAA	33 VLSSEIDKKACETYALN	55 GDINE	68 DFLLAGFPCCQPSYAGKQ	89 GDRGTLFF			
M.BepI	mCGCG	(44)	2 KVLSLFSGCGGMDLGLEGGF	56 VFANDILPFAKLAWCNF	90 VDVVN	107 DVVTGGFPCCQPSFAGKR	132 KDHNGLIYN			
M.BspRI	GmGCC	(45)	59 NVLSLFCGAGGLDLGFELAG	110 VYANDIFSEALQTYEKN	136 KDIRK	148 NLVIGGFPCCQPSFAGCR	168 DDERNFLYI			
M.BsuFI	mCCGG	(46)	102 TFDLFAIGIGIRLGFEDKY	125 VFSSEWDKYAAQTYEAN	148 GDITK	162 DVLLAGFPCCQPSFNIGKR	183 AHERRNIF			
M.BsuRI	GmGCC	(46)	60 NVLSLFSGCGGLDLGFELAG	110 IYTMDLKKEANQTYKTN	137 KDIRQ	149 NLILGGFPCCQPSFAGKR	169 DDERNFLYL			
dcm	CmCWGG	(47)	88 RFIDLFAIGIGIRRGFESIG	111 VFTSEWNKHAVRTYKAN	142 RDITL	169 DVLLAGFPCCQPSFAGVS	199 CDGTGTLFF			
M.DdeI	mCTNAG	(9)	2 NIDLFAGCGGFSHGFKMAG	24 ILAIEKDLWASQTSYFN	49 EDITT	68 DGIIIGPPCQGSFSLNGR	87 KDPNLSLV			
M.EcoRII	CmCWGG	(48)	97 RFIDLFAIGIGIRKGFTIG	120 VFTSEWNKEAVRTYKAN	148 LDIRE	178 DVLLAGFPCCQPSFAGVS	208 CEAGTLFF			
M.HaeIII	GmGCC	(49)	2 NLISLFSGAGGLDLGFKFAG	25 ICANEYDKSIUKITYESH	49 GDISK	63 DGIIIGPPCQASUSSESSL	84 DDPRLGYE			
M.HhaI	CmCGC	(50)	13 RFIDLFAIGLGGFRALALESCG	36 VYSNEWDKYAQEVYEMN	59 GDITQ	73 DILCAGFPCCQAFSISGKQ	94 EDSRGTLLF			
M.HpaII	CmCGG	(51)	33 TFDLFAIGIGIRFIAQMNLG	56 IFSSEWDEQAQKTYEAN	79 GDITL	95 DILCAGFPCCQAFSIAGRK	116 EDTRGTLFF			
M.HspI	mCCGG	(14)	106 KFIDLFSGIGIGIRRSFVNG	128 VFSSEIDPFAKFTYTYTN	150 GDITK	166 DILCAGFPCCQPSFHIGKR	188 NPTGTMFH			
M.NgoPII	GmGCC	(15)	13 KIISLFSGCGGLDLGFEEKAG	36 PAANEYDKTIWATFKAN	61 GDIRK	76 DGIIIGPPCQASUSSEAGL	97 DDARGQLFF			
M.Sau96I	GGMmCC	(20)	110 NVFETFAGAGGLDLGLESAG	123 YGAVEIDKNAETLRIN	148 NDIEF	176 DILSGGYPCQATFSYAGKR	187 ADTRGTLFY			
M.SinI	GGMmCC	(12)	76 KALSFSGAGMLDLGIEQAC	99 LLASEIDKAARDTILSN	123 GDIRD	146 DILMGPPCQAFSTAGKR	167 EDGRNVFI			
M.SssI	mCG	(52)	12 RVFEAFAGIGAQRKALEKVR	38 VGLAEVYPAIVMYQAI	119 FDIRD	133 DILTYSFPCQDLSSQGIQ	157 SGTRSGLFF			
SPR	GGMCC+mCCGG+CmCWGG	(53)	5 RVMSLFSGIGAFAEALRNIG	30 VGFSEIDKYAVKSFCAI	56 GDVSK	70 DLLVGGSPCQSFVSAGHR	91 EDTRGTLFF			
phi13T	GmCC+CmCNGC	(54)	5 RVMSLFSGIGAFAEALRNIG	30 IGFSEIDKYAISKYCAI	56 GDISK	70 DLTLSGFPCTFSVAGHR	124 KDRGTLLFF			
rho11g	GNGCC+GGCC	(55)	5 RVMSLFSGIGAFAEALRNIG	30 VGFSEIDKYAISKYCAI	56 GDVSK	70 DLLVGGSPCQSFVSAGHR	91 EDTRGTLFF			
M.HgiDI	GRCGYC (this work)	2	KTIDLFAAGCGGMSLGFMOAG	25 VAAVDNWRPAINTYQON	49 LDIAQ	67 ELIIGGPPCQDFSSAGKR	87 GLGRANLTL			
consensus			++islf+Giggt++gfe+-g	v++e+dk-a--ty-+n	gdi+-	d+l+ggfPCq+fs+ag+r	+d+rg+lft+			
			region F	region G	region H	region I	region J			
M.Aquia	110	ALPKCFVMENVKGMHWSKKG	151	EYKYA	160	VLNAADFGVPQFRERFVIGNRLGKT	71	RMLTVRELACLTQFPDLWFEFTG	104	VTMTGFGQVGNVAPVPLLAFAVAKAV
M.Bani	106	NRPKAFILLENVRGLVTHDKGR	139	GYGVS	146	VLNSSFVGVPQNRVRIYILGILGSKP	364	RRITPRECARLGGFPDDFILHS	386	NDNFAYKQLGNSVTVKVEKVIEDL
M.BepI	158	TKPKVFI AENVKGLVTLGDVCK	194	GYVVL	201	VLNAKNGVGAQNRERFVIFIGISKRYA	339	RRLTVRECALIGSFPDYEYFN	369	SASAAKYKIGNAVPPLLGFIAGRL
M.BspRI	183	VQPEIFVAENVKGMTLGGGE	218	GYRVE	225	LLNARDYGVGPQIRERVIIVGVNRDID	345	RRLSVKEIKRIQTFFPDYEFSD	377	RLDKQYKQIGNAVPPLFLTRAVAKSI
M.BsuFI	198	KQPKMFLLENVKGLLTNDNGN	233	GYSVF	240	VMDAQNGFLPQRRERVIIVGVNPDGL	349	RFLSELEKRLMGFPDVKFVP	371	SRTQMYRQFGNSVAVPMIKAVAGAM
M.BsuRI	84	AQPEIFVAENVKGMTLGKGE	219	GYRVQ	226	LLNARDYGVGPQIRERVIIEGVNRDID	346	RRLSVKEIARVQTFPDWQFSQ	378	RLDKQYKQIGNAVPPLAKAVASPI
dcm	214	RRPAMFVLENVKNLKSHDQCK	249	GYDVA	264	KTIIDGKHFLOHRRERVIIVGVNRDNL	400	RRLTPRECARLGMFEAPGEAKF	435	SDTAQYRQFGNSVVPVFAAVAKLL
M.DdeI	103	FSPKFFVMENVLGILSMKTKS	140	GYKVC	148	ILNACDYGVPQSRQRRVFIIGLSDRP	312	RNFTAREGARIGSFPDITYFGQ	341	KHLSQYQKIGNAVPPLLAQAALERI
M.EcoRII	223	KKPAIFVLENVKNLKSHDKGK	258	GYEVA	273	KVIDGKHFLPOHRRERVIIVGVNRDNL	409	RRLTPRECARLGMFEKVDGRPF	435	SDTQSYRQFGNSVVPVFEAVAKLL
M.HaeIII	98	KKPIFFLAENVKGMMAQRHMK	133	GYDVH	140	LLNANDYGVGAQNRERVIIGFRKELN	272	RRLTVRECARVGGFPDDFIHY	295	SLNDGYKQIGNAVPVLAYEIAKTI
M.HhaI	109	KKPKVFMENVKNFASHDNGN	144	DYSFH	151	VLNADYGVGPQNRERVIIGFRDNL	272	RKLHPRECARVMGYPSDYKYK	294	STSAQYKQFGNSVAVPVLAYIAYNI
M.HpaII	131	HQPKAFFLENVKGLKNHDKGR	167	GYFVP	175	IVNAKNGVGPQNRERVIIVGVNKGSTG	303	RKMTPREWARLGGFPDSYVIVP	325	SDASAYKQFGNSVAVPAIQTGKKI
M.HspI	203	KKTPVFLFLENVPLINHDDGN	238	GYKVN	245	VLDASHFGIPQKRRFYLVAFNLQNI	351	RLTTNECKAIMGFPDVIPIV	373	SRTQMYRQFGNSVAVPVTVKIAEQI
M.NgoPII	112	KQPKFFLAENVKGMMLANRNG	147	GYDVT	154	MANAKDYGVAQNRERVIIGFRKOLE	289	RRMTVREVARIGGFPDWFKFIY	312	NVNDAYKQIGNAVPVLAYEIAAAI
M.Sau96I	204	LKPKAFIAENVKGLVNHDDGK	237	GYEVY	244	ILNSWYDVAQKRRVYIIGIREDLV	377	RPFISIREYARIQSFPDEWFEFG	398	GVGAQYRQIGNAVPVLAKYIGKSL
M.SinI	182	IRPKYIVLENVRGLLSAPMKH	238	GYSVS	245	LYNSANFGVPQIRERVIICSRDGR	382	RPLSVQYKVIQGFPEEWIKG	404	KLDKYRQLGNAPVIGLGLAVGKNI
M.SssI	176	DLPKYLLMENVGALLNKKNEE	191	GYONS	218	VLNAADFGSSQARRVFMISTLNEFV	332	RKMNSDETFLYIGFDSQDGKRV	360	TENQKIFVCGNSISVEVLEATIDKI
SPR	106	KQPKFFVLENVKGLINHDKGN	141	GYRID	148	LLNSKFFNVQNRERVIYIIGIREDL	378	RRLTPLECFRLQAFDDEFEKA	405	SNSQLYKQAGNSITVTVLESIFKEL
phi13T	139	KQPKFVILLENVKGLINSGNGQ	174	GYRID	181	LLNSKFFNVQNRERVIYIIGIREDLV	382	RKLSPLECFRLQAFDDEFEKA	409	SNSQLYKQAGNSITVTVLESIFKEL
rho11g	106	KQPRYFVLENVKGLINHDKGN	141	GYRID	148	LLNSKFFNVQNRERVIYIIGIREDLI	443	RKLTPLECFRLQAFDDEFEKA	469	SNSQLYKQAGNSITVTVLESIFKEL
M.HgiDI	102	IQPAWVIMENVERARLSKIHQ	134	GYSLA	141	VLDASLCGVPQLKRTFVIGHRHGSI	244	RPLTKERSLIQTFFDLDFKFGV	266	TKSEQEQMIGNAVPVLAFPLATSL
consensus		++pk+fk++ENVKgl++++g-	gY-v-			+lne++gvpQ-ReR++iig+r++-	R+l++E-ar+q+Fp+++++		s-++-ykq+GNsv++v-++-a+g++	

Figure 4. Detailed comparison of ten conserved regions in m⁵C-methyltransferases including *M.HgiDI*. The figure aligns the names of the enzymes, their recognition specificities, number of references for protein sequences, the number of each first position of the conserved block and a possible consensus sequence at the bottom line. Since the borders of the conserved blocks are set not exactly like in reference 41, they are referred to as A to J in this work, rather than I to X. Following rules were obeyed for the evaluation of this consensus sequence: Highest conserved amino acids among all 21 m⁵C-methyltransferases are written in capital letters using the standard single letter abbreviation; conserved residues in more than 10 enzymes are written in small letters; '+' indicates more than 10 residues belonging to a common functional group; '-' shows less, but significant preference for certain amino acids or functional groups of residues.

R.HgiDI activity in crude extracts of pRMD1/HB101 was slightly higher than obtained from pRMD1anti/HB101, 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 most of 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 operator 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 (ATTAATG at position 1663 in Fig. 2) in pHK575/HB101 cells. IPTG induction resulted in a fairly high expression of *R.HgiDI*, 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.HgiDI* 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.HgiDI* overexpressing plasmid

For an endonuclease independent overexpression of the methyltransferase *M.HgiDI* the plasmid pHK540 (Fig. 3A) was constructed. It is based on our standard expression vector pHK255 and contains a *SmaI/HindIII* fragment of 1.9 kb derived from an *M.HgiDI* expressing exonuclease 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/HB101 cultures in denatured extracts on Coomassie stained SDS-PAGE gels, whereas in the control cells pHK545/HB101 the 33 kD product was induced, only. This identifies the 35 kD product as *M.HgiDI*, 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^5C -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^5C -methyltransferases show the localization of so-called target recognizing domains (22). The *M.HgiDI* characterized in this work is clearly the smallest m^5C -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 metB* 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 *HgiDI* methyltransferase and restriction endonuclease. The same result was observed, when R.*HgiDI* 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 *NarI* GCGGCC recognition sequence was also protected by *M.HgiDI* methylation, the product had to be either m^4C or m^5C . As expected, *M.HgiDI* shares some obvious homologies with one of these groups, namely the m^5C -methyltransferases. A careful comparison of *M.HgiDI* 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.HgiDI* (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.MspI* and 84 in *M.BsuFI*, but 193 in *M.BanI*; 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.BanI* (GGYRCC), which shares the *NarI* recognition sequence with *M.HgiDI*. Additional work will be carried out to analyze the relationship of the *HgiDI* 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

- Wilson, G.G. (1988) *Gene* **74**, 281–289.
- Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J. (1981) *J. Biol. Chem.* **256**, 2143–2153.
- Schoner, B., Kelly, S. and Smith, H.O. (1983) *Gene* **24**, 227–236.
- Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) *Nucleic Acids Res.* **12**, 3659–3676.
- Walder, R.Y., Hartley, J.L., Donelson, J.E. and Walder, J.A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1503–1507.
- Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) *Nucleic Acids Res.* **13**, 6403–6421.
- Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) *Nucleic Acids Res.* **13**, 8441–8461.
- Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, B. (1986) *Cell* **46**, 993–1000.
- Sznyter, L.A., Slatko, B., Moran, L., O'Donnell, K.H. and Brooks, J.E. (1987) *Nucleic Acids Res.* **15**, 8249–8266.
- Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M. and Wilson, G.G. (1987) *Nucleic Acids Res.* **15**, 9781–9796.
- Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) *Gene* **70**, 387–392.
- Karreman, C. and De Waard, A. (1988) *J. Bacteriol.* **170**, 2527–2532.
- Kita, K., Kotani, H., Sugisaki, H. and Takanami, M. (1989) *J. Biol. Chem.* **264**, 5751–5756.
- Lin, P.M., Lee, C.H. and Roberts, R.J. (1989) *Nucleic Acids Res.* **17**, 3001–3011.
- Sullivan, K.M. and Saunders, J.R. (1989) *Mol. Gen. Genet.* **216**, 380–387.
- Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) *Gene* **85**, 1–13.
- Heidmann, S., Seifert, W., Kessler, C. and Domdey, H. (1989) *Nucleic Acids Res.* **17**, 9783–9796.
- Ito, H., Sadaoka, A., Kotani, H. and Nakamura, T. (1990) *Nucleic Acids Res.* **18**, 3903–3911.
- Mackawa, Y., Yasukawa, H. and Kawakami, B. (1990) *J. Biochem. (Tokyo)* **107**, 645–649.
- Szilak, L., Venetianer, P. and Kiss, A. (1990) *Nucleic Acids Res.* **18**, 4659–4664.
- Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
- Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305–312.
- Kröger, M., Hobom, G., Schütte, H. and Mayer, H. (1984) *Nucleic Acids Res.* **12**, 3127–3141.
- Howard, K.A., Card, C., Benner, J.S., Callahan, H.L., Maunus, R., Silber, K., Wilson, G. and Brooks, J.E. (1986) *Nucleic Acids Res.* **14**, 7939–7951.
- Brooks, J.E., Benner, J.S., Heiter, D.F., Silber, K.R., Sznyter, L.A., Jager-Quinton, T., Moran, L.S., Slatko, B.E., Wilson, G.G. and Nwankwo, D.O. (1989) *Nucleic Acids Res.* **17**, 979–997.
- Oberbäumer, I. (1986) *Gene* **49**, 81–91.
- Fürste, J.P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M. and Lanka, E. (1986) *Gene* **48**, 119–131.
- Mayer, H. and Reichenbach, H. (1978) *J. Bacteriol.* **136**, 708–713.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
- Henikoff, S. (1984) *Gene* **28**, 351–359.
- 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.
- Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
- Benton, W.D. and Davis, R.W. (1977) *Science* **196**, 180–182.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Kröger, M. and Kröger-Block, A. (1984) *Nucleic Acids Res.* **12**, 113–120.
- Szomolanyi, E., Kiss, A. and Venetianer, P. (1980) *Gene* **10**, 219–225.
- Wilson, G.G. (1988) *Trends in Genetics* **4**, 314–318.
- Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., Decruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z. and Graham, M.W. (1989) *Nucleic Acids Res.* **17**, 3469–3478.
- Duchange, N., Zakim, M.M., Ferrara, P., Saint-Girons, I., Park, I., Tran, S.V., Py, M.C. and Cohen, G.N. (1983) *J. Biol. Chem.* **258**, 14868–14871.
- Gouy, M. and Gautier, C. (1982) *Nucleic Acids Res.* **10**, 7055–7074.
- Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
- Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. and Trautner, T.A. (1988) *EMBO J.* **7**, 2601–2609.
- Karreman, C. and De Waard, A. (1990) *J. Bacteriol.* **172**, 266–272.
- Kupper, D., Zhou, J.G., Venetianer, P., Kiss, A. (1989) *Nucleic Acids Res.* **17**, 1077–1088.
- Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) *J. Mol. Biol.* **170**, 597–610.
- Walter, J., Noyer-Weidner, M. and Trautner, T.A. (1990) *EMBO J.* **9**, 1007–1013.
- Hanck, T., Gerwin, N. and Fritz, H.-J. (1989) *Nucleic Acids Res.* **17**, 5844.
- Som, S., Bhagwat, A.S. and Friedman, S. (1987) *Nucleic Acids Res.* **15**, 313–332.
- Slatko, B.E., Croft, R., Moran, L.S., Wilson, G.G. (1988) *Gene* **74**, 45–50.
- Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) *J. Biol. Chem.* **262**, 4770–4777.
- Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A. and Roberts, R.J. (1990) *Nucleic Acids Res.* **18**, 1377–1383.
- Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S. and Razin, A. (1990) *Nucleic Acids Res.* **18**, 1145–1152.
- Buhk, H.-J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Günthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) *Gene* **29**, 51–61.
- Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) *Gene* **42**, 89–96.
- Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesch, T.S. and Trautner, T.A. (1987) *EMBO J.* **6**, 1137–1142.
- Kröger, M. and Hobom, G. (1982) *Gene* **20**, 25–38.
- Mashko, S.V., Veiko, V.P., Lapidus, A.L., Lepedeva, M.I., Mochulsky, A.V., Schechter, I.I., Trukhan, M.E., Ratmanova, K.I., Rebentisch, B.A., Kaluzhsky, V.E. and Debabov, V.G. (1990) *Gene* **88**, 121–126.