Eight new restriction endonucleases from *Herpetosiphon giganteus* – divergent evolution in a family of enzymes

Manfred Kröger and Gerd Hobom
Institut für Biologie III der Universität, Schänzlestrasse 1, D-7800 Freiburg im Breisgau, FRG, and

Horst Schütte and Hubert Mayer
Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-3300 Braunschweig, FRG

Received 5 January 1984; Revised and Accepted 8 March 1984

ABSTRACT

Characterization of eight restriction endonucleases isolated from five strains of *Herpetosiphon giganteus* is described. *Hgi*CI from strain Hpg9 recognizes and cleaves the degenerate sequence: ↓GGPyPuCC, producing 5'-hexanucleotide protruding ends. Endonucleases *Hgi*BI, *Hgi*CII and *Hgi*EI are isoschizomers of *Ava*II; *Hgi*CIII and *Hgi*DII are isoschizomers of *Sal*I; and *Hgi*DI and *Hgi*GI are isoschizomers of *Acy*I. Based upon their closely related and in part overlapping recognition specificities a close evolutionary relationship is proposed for all known *Hgi* restriction endonucleases.

INTRODUCTION

Type II restriction endonucleases have been identified in a wide variety of prokaryotes over the last years, in almost one hundred types of recognition specificities (see [1]). New restriction enzymes may serve for more flexibility in generating and cloning DNA fragments, but they may also serve as models in studying protein - nucleic acid recognition or the endonuclease reaction mechanism [2-4]. Along these lines it is important to characterize a number of restriction enzymes which interact with the same DNA sequence palindrome, but come from widely different sources. These isoschizomer restriction endonucleases will allow to compare their protein structures for elucidation of a common mechanism of DNA sequence recognition in an almost ideal case of convergent evolution. On the other hand, divergent steps in evolution might be studied through comparing a group of endonucleases from closely related strains of a single bacterial species, if they have very similar but slightly different specificities of DNA sequence recognition.

The first step into this direction is an extensive search for

homologous restriction enzymes. We chose *Herpetosiphon giganteus* [5] for such an exploration and isolated a number of enzymes from several strains of this bacterium. Characterization of eight of these *Hgi* enzymes is presented in this report. A correlation together with six other known types of *Hgi* recognition specificities is in agreement with a case of divergent evolution from a common ancestor for the genes coding for this family of restriction endonucleases.

MATERIALS AND METHODS

Herpetosiphon giganteus strains Hpg5 ('*Hgi*B'), Hpg9 ('*Hgi*C'), Hpa2 ('*Hgi*D'), Hpg24 ('*Hgi*E'), and Hpa1 ('*Hgi*G') were grown and kindly given to us by Dr. H. Reichenbach, GBF Braunschweig. DNA sequences and sequence boundaries on λ DNA for λ- or hybrid λ-derived plasmids λdvh93 (λ:37,768-40.091), λdv021 (λ:37,183-40,335), λdv1 (λ:35,465-42,646), and λdv*imm*21AB5 (21:~36,400-λ:~43,260) have been described previously [6-8]. The numbering system of the λ DNA sequence is according to Sanger et al. [9], numbering of the 21 DNA sequence (M. Kröger and G. Hobom, unpublished) is derived from the λ DNA sequence with reference to the common *Eco*RI site within the origin of λ and 21, at position 39,169. Numbering in the φ80 DNA sequence refers to the *Eco*RI site at φ80-ori=40,000 (R. Grosschedl and G. Hobom, unpublished).

All restriction enzyme tests were performed in a buffer containing 10 mM $MgCl_2$ and 10 mM Tris·HCl, pH 7.5 at 37°C. Ligase reactions were done as described previously for the characterization of *Eca*I [10]. Following digestion the cleavage products were separated on agarose gels (0.8-1.5%) or on polyacrylamide gels (5-15%). Fragment sizes as described in RESULTS were calculated using partially cleaved λdv021 octamer plasmid (multimers of 3153bp [7,8]), and *Hae*III digested λdv1 DNA as standard size markers (see Figure 1), together with fragments obtained in specific, parallel cleavages such as isoschizomer digestions.

For determination of the 5'-terminal nucleotide the fragments obtained after enzymatic hydrolysis were dephosphorylated and 5'-labelled using [γ-^{32}P]ATP and T4 polynucleotide kinase. The mixtures of ^{32}P-labelled fragments were cleaved in additional restriction endonucleolysis prior to gel electrophoretic separa-

tion. Fragments with only one 5'-labelled end were isolated and digested to completion with 1 µg DNAseI and 1 µg snake venom phosphodiesterase in 30 µl of 10 mM Tris pH 8.5 and 10 mM $MgCl_2$ for 1 h at 37°C. 40 µl of a mixture of unlabelled mononucleotides (about 0.2 mg/ml each) were added as carrier to the ^{32}P-labelled mononucleotide preparations in order to achieve optical visibility. The reaction mixtures were separated by electrophoresis on Whatman 3MM paper at pH 3.5. The four mononucleotide spots observed were cut out under UV-light and used directly for measuring their ^{32}P-activities.

For determination of the exact length of a *Hgi*CI generated fragment a conventional Maxam-Gilbert sequencing reaction [11] was used for preparing a sequencing ladder against which sizing of the corresponding *Hgi*CI fragment could be done.

RESULTS

General isolation procedure of restriction endonucleases

The following purification has been standardized for 10 g of cells. Strains were grown as described previously [5] at 30°C, collected by centrifugation and kept at -20°C. Cells were thawed and suspended in 10 mM potassium phosphate buffer (pH 7.5), 0.1% (v/v) 2-mercaptoethanol, 1 mM EDTA and 10% glycerol (buffer A). The concentration of the cell suspension was adjusted to 40%. For preparation of the crude extract the cells were broken by a 10 min treatment with a Branson Sonifier at 60 W in an ice bath. Cell debris was removed by centrifugation at 45,000 x g for 30 min. The supernatant was directly applied to a column of DEAE-cellulose DE52 (2.6 x 15 cm) which had been equilibrated with buffer A until the absorption at 280 nm decreased to the baseline. The restriction endonucleases were then eluted gradually by increasing the sodium chloride concentration in buffer A from 0 to 0.5 M.

Most of the restriction endonucleases eluted in the range of 0.1-0.25 M NaCl; only *Hgi*DII eluted at 0.45 M NaCl. Restriction endonucleolytic activity was assayed for every fraction by incubation with λ DNA as the general substrate. Fractions with identical activities were pooled and dialyzed against buffer A. The dialyzed enzyme solution was rechromatographed on a phosphocel-

lulose P11 column (2.6 x 15 cm) equilibrated with the same buffer. Elution was carried out by increasing the NaCl concentration in buffer A either in a stepwise manner to 200 mM, 500 mM, and 1 M, or by using a linear gradient from 0 to 0.4 M. The restriction endonucleases were again located in the effluent through their ability to cleave λ DNA. In *Herpetosiphon* strains containing more than a single restriction enzyme activity notations such as *Hgi*CI, *Hgi*CII, and *Hgi*CIII were given according to their order of elution from the P11 column.

As a final step all fractions with identical activities were pooled and concentrated by dialysis against buffer A containing 20% PEG 2000. Storage was possible for short periods at 4°C. For longer periods of storage at -20°C glycerol was added to yield a final concentration of 50% (v/v).

Recognition sequence determination

General strategies. The recognition sequence of the individual enzymes was analyzed through collecting mapping data with increasing precision. An initial set of data was obtained through cleavage of pBR322 and a series of closely related, but distinctly different λdv plasmid DNAs: λdv1, λdv021, λdvh93, and λdv*imm*21AB5, which constitute overlapping segments of the λ (or λ-hybrid) replication region. The group of four λdv plasmids when digested with one of the restriction endonucleases gave a small number of identical fragments from the common part of λ DNA, and a single or few nonidentical fragment(s). The latter fragments could directly be identified as containing either heterologous sections and/or the plasmid fusion point. Thus, comparison of the four cleavage patterns always led to a preliminary localization for a group of *Hgi* cleavage sites. Appropriately selected endonucleases with nearby single restriction sites served to confirm and to better define that original map in double digestions. Finally, detailed measurements for at least two individual cleavage positions were obtained through mixed digestions in a series with enzymes which generate small fragments such as *Hae*III or *Sau*3A. This was either done directly in mixed digestions, or with preisolated fragments containing only a single cleavage site of the respective enzyme. If a computer aided search [12] detected similarities with fragmentation pat-

terns of an already described enzyme this identification was followed up in parallel digestions with the potential isoschizomeric enzyme.

During this multistep procedure some of the *Hgi* enzymes were detected to show identical cleavage patterns, and further could be assigned to known enzymatic activities as follows: *Hgi*BI, *Hgi*CII and *Hgi*EI as isoschizomers with *Ava*II [13,14], *Hgi*CIII and *Hgi*DII as isoschizomers with *Sal*I [15], and *Hgi*DI and *Hgi*GI as isoschizomers with *Acy*I [16]. Fragmentation patterns distinctly different from any of the known restriction enzymes were observed for *Hgi*CI and for a second endonucleolytic activity present in *H. giganteus* strain Hpg 24 (*Hgi*EII; G. Hobom and H. Mayer, unpublished).

*Hgi*BI, *Hgi*CII, *Hgi*EI. For all three enzymes three recognition sites were found near λ-ori in the identical segment of the λdv plasmids, at λ:39,005; λ:39,438; and λ:39,480 with positions measured relative to λ-*Eco*RI:39,169, λ-*Bgl*II:38,815 and λ-*Hpa*I: 39,609. In accordance with the known *Ava*II fragmentation no other *Hgi*BI, *Hgi*CII or *Hgi*EI restriction site was observed within the λdv DNA boundaries. A distance of 148 bp from the pBR322-*Sal*I site was also measured for the *Hgi*BI, *Hgi*CII, and *Hgi*EI site at pBR322:799, again as calculated for the *Ava*II site: $GG^{A}_{T}CC$.

*Hgi*CIII, *Hgi*DII. Mixed digestions with *Eco*RI or *Bam*HI of pBR322 and λdv*imm*21 DNA suggested a similarity with *Sal*I. This was confirmed through cleaving total λ DNA. As for *Sal*I, the resulting 499 bp fragment (λ:32,746-33,245) could be recleaved by *Cla*I (λ:32,965) in a very specific, partially inhibited manner because of overlapping GATC-specific methylation [17]. Another *Sal*I specific pattern has been observed in φ80 DNA digestion, including formation of a 175 bp fragment from φ80-*c*I DNA (φ80: 38,258-38,433; R. Grosschedl and G. Hobom, unpublished), identical to fragmentation by *Sal*I: GTCGAC.

*Hgi*DI, *Hgi*GI. A series of six recognition sites was observed for these enzymes in pBR322 DNA in mixed digestions with *Eco*RI, *Bam*HI, *Sal*I, and *Pvu*I which appeared to be similar to the pattern described for endonuclease *Acy*I [16]. This was confirmed by the highly specific production of a 30 bp fragment from the *cro* gene in λdv*imm*21 DNA (21:38,167-38,197) and its precise locali-

zation through subfragmentation of an isolated λdv*imm*21-*Hae*III fragment (21:~36,950-38,265). Recognition sites at λ:40,808, λ:41,115, and λ:42,249 have been localized relative to λ-*Cla*: 41,365 and λ-*Cla*:42,022. Cleavage of GGCGCC, GGCGTC, and GACGTC in pBR322, and GACGCC in λdv*imm*21 DNA proved that all four possible combinations of the degenerate *Acy*I sequence GPuCGPyC are recognized.

*Hgi*CI. Digestion of pBR322 DNA with endonuclease activity *Hgi*CI resulted in a pattern of nine products out of which two small fragments, but no others appeared to be identical with two of the six *Hgi*DI fragments. In mixed digestions with *Hgi*CI + *Hgi*DI eight of the nine *Hgi*CI fragments stayed unchanged, and only a single one was cleaved into three subfragments by *Hgi*DI. This indicated an overlap or very close proximity for four cut of six *Hgi*DI restriction sites with four out of nine *Hgi*CI recognition sequences, see Figure 1. Precise measurement after mixed digestions with *Bam*HI and *Sal*I allowed to locate two of the coincident *Hgi*CI and *Hgi*DI cleavage sites at identical positions (pBR322:435 and pBR322:549), and the same result was obtained for the GGCGCC sequence present in the *cro* gene of λdv*imm*21 (21:38,197) through subfragmentation of the *Hae*III fragment already mentioned above for *Hgi*DI; the GACGCC sequence at 21-*Hgi*DI:38,167 is not cleaved by *Hgi*CI, however.

The four GGCGCC sequences present in pBR322 and in λdv*imm*21 DNA are not only cleaved by *Hgi*CI and *Hgi*DI, but also by endonuclease *Hae*II (PuGCGCPy [18]), which again has seven additional sites in pBR322 DNA not recognized by either of the two other enzymes (Fig. 1). This situation suggested recognition of another degenerate group of restriction sites for *Hgi*CI, with a degeneracy overlapping, but different from the two degeneracies known for *Hae*II and *Hgi*DI: PuGCGCPy and GPuCGPyC, respectively. A recognition sequence: GGPyPuCC for *Hgi*CI concluded from these comparisons matched perfectly the experimentally obtained restriction patterns both for the GGCGCC sites commonly recognized by all three enzymes, and also for the remaining fragmentation sites which are recognized only by *Hgi*CI. This prediction finally was confirmed through distance measurements for a GGTGCC site in λdv1 DNA at λ:39,908, next to λ-*Hpa*I:39,837, and for a GGCACC

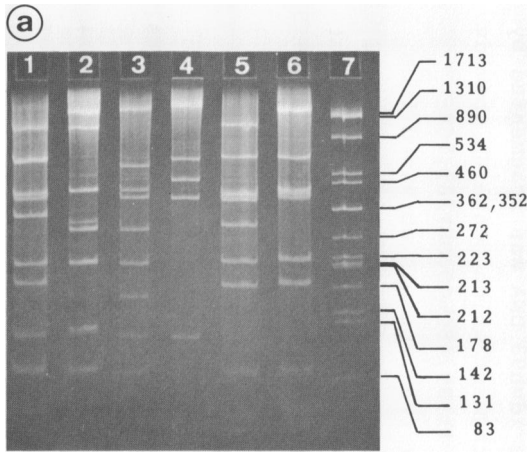
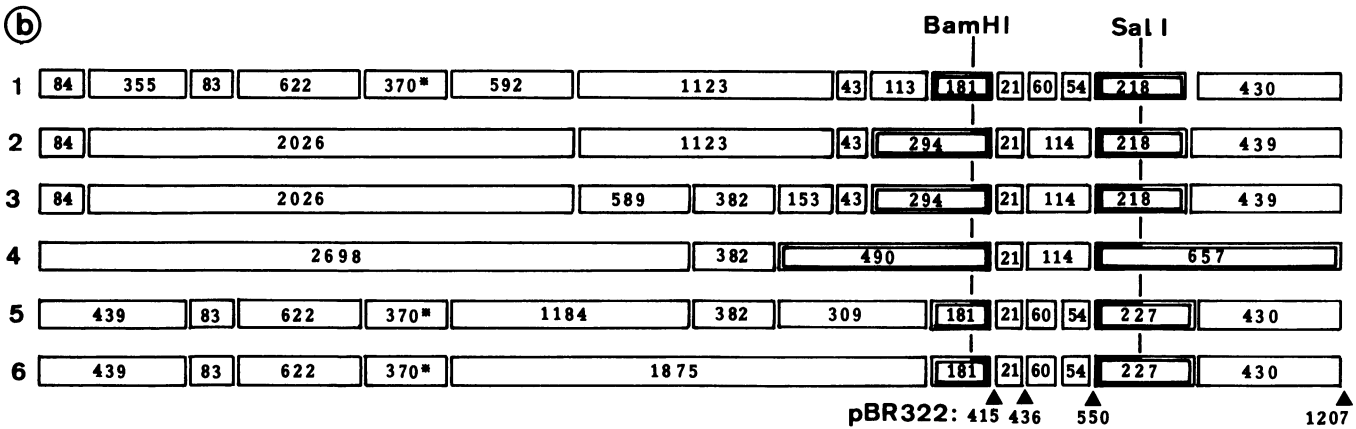


Figure 1. Determination of *Hgi*CI and *Hgi*DI cleavage sites on pBR322 DNA. a) 5% polyacrylamide gel electrophoresis and b) restriction map for the following digestions 1: *Hgi*CI + *Hae*II, 2: *Hgi*CI, 3: *Hgi*DI + *Hgi*CI, 4: *Hgi*DI, 5: *Hgi*DI + *Hae*II, 6: *Hae*II. Lane 7 contains the marker pattern of *Hae*III cleaved λdv1 DNA; marker fragment sizes are indicated on the right. Larger fragments are measured in parallel on agarose gels, very small fragments on 15% polyacrylamide gels (not shown). The pBR322 fragments given in the map in bp are not drawn to scale. Cleavage points in common for all three enzymes are marked with arrows below the map. The appropriate map coordinates for pBR322 refer to *Eco*RI = 1. *Bam*HI and *Sal*I double or triple digestions have been used for a precise measurement of the resulting subfragments indicated in heavy lining (on 10% polyacrylamide; not shown). * denotes origin containing fragment with unusual electrophoretic mobility.



site in pBR322 DNA at pBR322:77, next to pBR322-*Eco*RI:4,361 (not shown). Since only three out of four pyrimidine/purine combinations possible for a GGPyPuCC consensus sequence are present in pBR322 and in the λdv DNAs used we had to test for *Hgi*CI sensitivity of the *Kpn*I sequence GGTACC in another plasmid DNA. For this purpose we used pHL188, which contains a 370 bp fragment of φ80 DNA (36,665-37,035) including a *Kpn*I sequence, but no other GGPyPuCC site in the inserted segment. The expected 353 bp fragment was found in both double digestions done in parallel, with *Kpn*I and *Eco*RI, and with *Hgi*CI and *Eco*RI. Thus we conclude that the recognition site for *Hgi*CI is GGPyPuCC.

Determination of cleavage sites

Three strategies have been used for the various enzymes in determining the position of endonucleolytic cleavage within their recognition sequences: i) cross-ligation of fragments generated by the enzyme under study with fragments of different size generated by another restriction endonuclease, in particular in cases suspected to be isoschizomers [10]; ii) chemical analysis of nucleotides exposed at 5'-terminal ends of fragments after cleavage; iii) sizing of (small) restriction fragments against corresponding sequencing ladders.

*Hgi*BI, *Hgi*CII, *Hgi*EI. Within this group determination of the cleavage position relative to the known *Ava*II reaction was done by cross-ligation. For this purpose λdvh93 DNA was fragmented by *Ava*II, and the resulting 1849 bp fragment which contains the λdv circular fusion point was isolated (λ:39,480-40,091/37,768-39,004). A very similar, but larger fragment was obtained from λdv021 DNA (2663 bp; λ:39,480-40,335/37,183-39,004) after *Hgi*BI, *Hgi*CII, or *Hgi*EI digestion. It is known from the λ sequence that both recognition sites located at the boundaries of these fragments consist of identical GGACC sequences. The *Ava*II-1849 bp fragment and the *Hgi*BI-2663 bp fragment (or *Hgi*CII-2663 bp, *Hgi*EI-2663 bp fragment) were mixed and co-ligated. The reaction products were analyzed on agarose gels next to ligation products obtained from individual reactions, see Figure 2. In all three co-ligation reactions the specific heterologous dimers of 1849 + 2663 = 4512 bp and traces of higher heterologous oligomers were observed in addition to homologous dimers and oligomers. Finally,

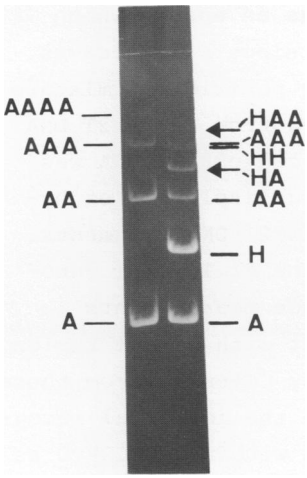


Figure 2. Interspecific ligation reaction between DNA fragments produced by isoschizomeric enzymes *Hgi*BI and *Ava*II. An 1849 bp *Ava*II fragment (A) obtained from λdvh93, and a 2663 bp *Hgi*BI fragment (H) obtained from λdv021 DNA have been ligated both individually to produce a series of homologous oligomers, and after mixing the two substrates for an analysis of the production of heterospecific dimers and oligomers. A key on the left and right denotes the composition of bands present among ligation products in the *Ava*II homoligation (lane 1), and in the *Hgi*BI-*Ava*II heteroligation (lane 2) reactions. The hetero-dimer band of 1849 + 2663 = 4512 bp and the hetero-trimer band of 1849 + 1849 + 2663 = 6361 bp are indicated by black arrows. Similar results have been obtained for the *Hgi*CII-*Ava*II and *Hgi*EI-*Ava*II interspecific ligation reactions.

the reaction mixtures obtained were redigested by *Ava*II and by the respective *Hgi* enzyme used before, and in both reactions the series of oligomers were converted back completely into the two heterologous monomer fragments. From these results we conclude that all three enzymes *Hgi*BI, *Hgi*CII, and *Hgi*EI are true isoschizomers of endonuclease *Ava*II: G↓G$^{A}_{T}$CC.

*Hgi*CIII and *Hgi*DII. Determination of the cleavage position for these two enzymes was also done by cross-ligation using the *Sal*I linearized pBR322 DNA and the *Hgi*CIII or *Hgi*DII digested λdv*imm*21 DNA (4361 bp, and close to 6850 bp, respectively) as monomer substrates for interspecific ligation. Formation of heterologous dimers and oligomers which could be redigested by both enzymes (data not shown) proved both endonucleases to be true isoschizomers of *Sal*I: G↓TCGAC.

*Hgi*DI, *Hgi*GI. Analysis of the 5'-terminal nucleotide revealed in both cases a pC residue (>95% and >90%). However, this cytosine residue could have been derived from either of the two invariable C residues within the GPuCGPyC sequence. C residues from the variable Py position have been excluded, since all four possible cleavage sites for both enzymes have been analyzed without any deviation. In order to distinguish between the two possibilities we again applied a variation of the crossligation

analysis. For this purpose *Cla*I was chosen as an enzyme which also produces 5'-pCpG single stranded ends, since *Cla*I cleaves pBR322 DNA in a unique and favorable cloning position. A mixture of *Hgi*DI and of *Hgi*GI fragments was generated from λdv*imm*21 DNA and interspecific ligation with *Cla*I linearized pBR322 DNA was observed on agarose gels and confirmed through isolation of hybrid plasmid clones for the four larger λdv*imm*21 DNA fragments. In these hybrid plasmid DNAs recleavage of the *Cla*I-*Hgi*DI junctions was impossible as expected, but distance measurements across these junction sites were in agreement with direct fusion of the two components via their complementary CG-ends. From these results we conclude that *Hgi*DI and *Hgi*GI use the identical recognition sequence and cut at the same position within GPu↓CGPyC as does *Acy*I [16].

*Hgi*CI. The 5'-nucleotide was determined for all four possible cleavage sites to be pG (93% - 98%). This result was again ambiguous since we could not decide, which of the two invariant G positions within GGPyPuCC would be the actual cleavage site. Because of the centrally degenerated sequence recognized by *Hgi*CI the method of cross-ligation could not be applied in this case. Therefore, we decided to characterize the cleavage position through direct size comparison of a small *Hgi*CI fragment with the corresponding sequencing ladder across that recognition site, as has been successfully applied to characterize the cleavage site for *Hgi*AI [19]. The result obtained in this approach is illustrated in Figure 3. The *Bam*HI-*Sal*I pBR322 DNA fragment (pBR322: 376-652) containing three of the *Hgi*CI sites (pBR322:414, 435, and 549) was ^{32}P-labelled and recleaved by *Sau*3A. The resulting 91 bp subfragment was isolated and subsequently both cleaved with *Hgi*CI and chemically degraded according to the Maxam-Gilbert protocol [11]. After separation in a sequencing gel a single band was observed in the lane carrying the endonuclease reaction product(s) which in the sequencing ladder corresponded with the product of chemical cleavage before the first G residue within the recognition sequence at pBR322:414. This allows the conclusion that *Hgi*CI specific hydrolysis occurs at the phosphodiester bond 5'-proximal to the recognition sequence. Very low yields in ^{32}P-ATP phosphorylation with polynucleotide kinase if the *Hgi*CI

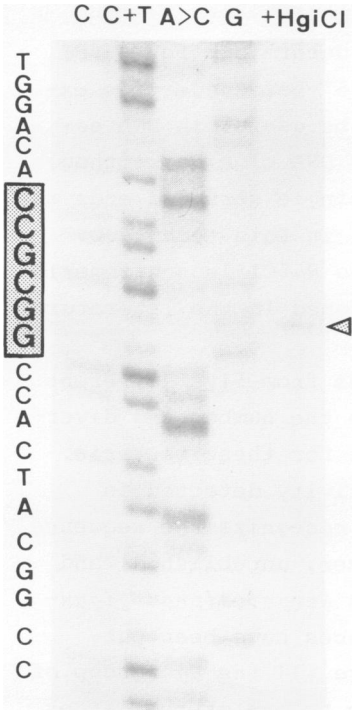


Figure 3. Determination of the cleavage site for endonuclease *Hgi*CI. Autoradiograph of a DNA sequencing ladder (lanes 1-4) across the recognition site for *Hgi*CI at pBR322:414 is shown as a reference for the position of the fragment band produced in the *Hgi*CI cleavage reaction (lane 5). The 91 bp DNA fragment ^{32}P-labelled at the *Bam*HI site (pBR322:376) and recleaved by *Sau*3A at pBR322:467 has been used for both reactions. The *Hgi*CI generated fragment (arrow) occupies a position slightly below the second G-specific band of the chemical degradation reaction, for the boxed *Hgi*CI sequence GGCGCC. Since additional phosphate groups at the 3' ends of the chemical degradation products result in increased electrophoretic mobilities of otherwise identical fragments, the observed *Hgi*CI band corresponds to a cleavage reaction before the two invariant G residues in the degenerate ↓GGPyPuCC *Hgi*CI recognition sequence.

fragment is prepared without treatment by phosphatase, and high yields following pretreatment by phosphatase proved the production of 5'-phosphate ends in *Hgi*CI cleavage reactions. In conclusion, the *Hgi*CI recognition sequence has been determined to be ↓GGPyPuCC, a first example of a 5'-hexanucleotide staggered cut.

DISCUSSION

A majority of endonucleases described in this report was characterized as isoschizomers of restriction enzymes known before, such as *Acy*I, *Ava*II, and *Sal*I. However, availability of these enzymatic activities from two non-related sources, *Herpetosiphon* and *Anabaena*, may add to their accessibility and also offers a basis for phylogenetic comparison. *Hgi*DI or *Hgi*GI (*Acy*I) are attractive enzymes for DNA cloning, because they produce medium size fragments with 5'-CG single stranded ends, which can be cloned into *Cla*I cleaved vector DNAs such as pBR322 or derivatives, similar to fragments generated by *Taq*I, *Hpa*II,

*Hin*P1I, *Nar*I, or *Asu*II [17].

A recognition specificity of the type determined for *Hgi*CI has not yet been described. Due to their 5'-hexanucleotide extended ends *Hgi*CI specific fragments may be useful in DNA sequencing and also for special purposes of DNA cloning, although the degenerate positions present in the single stranded ends may interfere with a more general application in this technique. During preparation of this publication two *Hgi*CI isoschizomeric recognition specificities have been mentioned in the literature: *Hgi*HI [1], and *Ban*I [1,20].

Our eight new restriction endonucleases from five different strains of *Herpetosiphon giganteus* add to the number and diversity of restriction enzymes already known for these bacteria. Together with an additional enzymatic activity detected in strain Hpg24, which has been observed to recognize the sequence ACC------GGT (*Hgi*EII; G. Hobom and H. Mayer, unpublished) and five other restriction endonucleases from *Herpetosiphon giganteus* strains for which recognition sequences have been published [19] or mentioned in the literature [1] the *Hgi* group of 14 endonucleases (seven types) appears to be one of the larger series of enzymes investigated to date from very closely related strains of a single bacterial species.

A prominent feature in the characterization of different enzymatic activities from *Herpetosiphon giganteus* strains has been the repeated observation of degenerate recognition sequences. These also in part overlap with each other, as has been described above in particular for the *Hgi*CI and *Hgi*DI pair of enzymes. A similar overlap can be pointed out for the pair of *Hgi*AI [19] and *Hgi*JII [1] enzymes. Although some of the enzymatic activities, such as *Hgi*CIII (*Sal*I) or *Hgi*EII, may appear to be unrelated in this regard, it is possible to arrange the complete series of *Hgi* enzymes into a pattern of close correlation among their recognition sequences.

All of the presently known *Hgi* enzymatic activities fit into the correlation scheme presented in Figure 4. Two of the six boxes in this tabulation which represent degenerate recognition in positions 2 and 5 are not yet filled, or only occupied by a specific, not a degenerate hexanucleotide recognizing activity

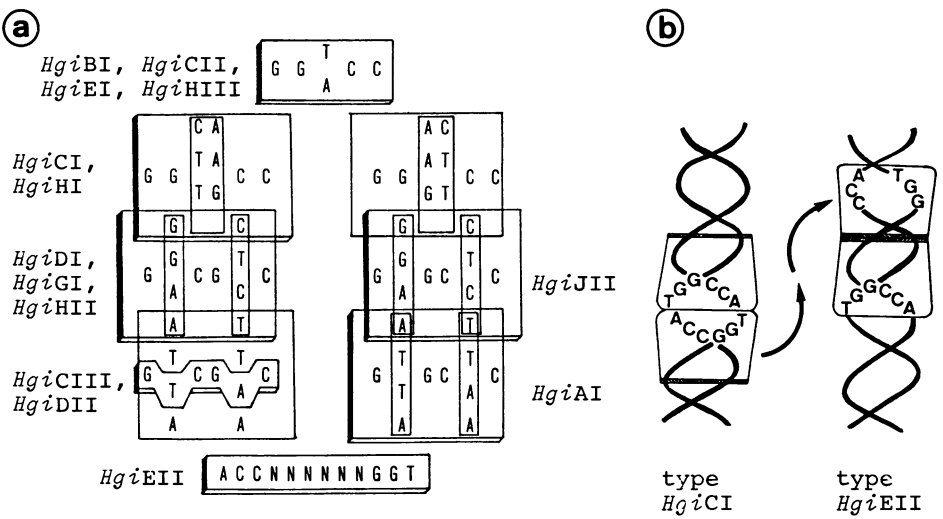


Figure 4. a) Tabulation of recognition sequences for the known *Herpetosiphon giganteus* restriction enzymes. A majority of the known *Hgi* endonucleases recognizes a specific group of degenerate sequences (boxed) out of which individual recognition sequences may be shared by two different enzymes, as reflected in partial overlaps between adjacent specificity groups. Boxes of recognition sequences shown in heavy lining have been determined for *H. giganteus* enzymes to date. b) Proposed correlation between recognition sequences for endonucleases *Hgi*CI and *Hgi*EII. A rearrangement of the dimerization mode for two identical subunits of *Hgi*CI is shown to provide the structural basis for an interaction with a similarly arranged, 'gapped' recognition sequence of endonuclease *Hgi*EII.

(*Hgi*CIII/*Hgi*DII), see Figure 4a. A number of additional enzymatic activities have, however, been observed in our laboratories in extracts from several other strains of *H. giganteus*, such as *Hgi*FI and *Hgi*KI, which remain to be characterized in detail in recognition and cleavage properties.

The left part of the scheme presented in Figure 4a contains central CG(PyPu) sequences and has cleavage sites with protruding 5'-ends, while the right branch in an otherwise symmetrical pattern has central GC(PuPy) sequences and single stranded 3'-ends after cleavage. The close correlation of recognition sequences for the various *H. giganteus* endonucleases suggests the existence of an evolutionary tree behind that scheme, with the genes for all restriction (and possibly modification) systems

in *H. giganteus* strains related to each other.

This inference may be extended to include also the endonuclease *Hgi*EII system. The *Hgi*EII recognition sequence although different at first sight appears to be related to one of the *Hgi*CI specific sequences (GGTACC) by a separation of the two constituting elements and reorganization into a new type of 'gapped' palindrome: ACC------GGT. In terms of the two enzyme subunits recognizing such a separated palindrome sequence we propose that in the *Hgi*EII enzyme the subunits may also have been rearranged in a correlated way in their subunit association relative to 'regular' enzymes such as *Hgi*CI. Consequently a rearranged *Hgi*EII enzyme will recognize the two halves of its rearranged palindrome sequence one helical turn above each other (on the same side of the DNA substrate; see Figure 4b). In this model the 'gap' basepairs in the *Hgi*EII recognition sequence are assigned to a spacer function that derives from the recognition complex which extends from one helical turn over to the next. A similar interpretation may apply to other restriction endonucleases which interact with two halves of a recognition sequence spaced over a distance of approximately 10 bp, such as *Bgl* I: GGC-----GCC [21], and in particular has been proposed to account for the gapped recognition complexes of *E.coli* type I restriction-modification enzymes [22].

ACKNOWLEDGEMENTS

The expert technical assistance of Mrs. G. Schlingmann, Miss M. Melzer, and Mrs. P. Schliephake-Rueß is gratefully acknowledged. This work has been supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Roberts,R.J. (1983) Nucleic Acids Res. 11, r135-r167.
2. Modrich,P. and Roberts,R.J. (1982) in Nucleases, Linn,S.M. and Roberts,R.J., Eds., pp. 109-154, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
3. Lu,A.-L., Jack,W.E. and Modrich,P. (1981) J. Biol. Chem. 256, 13200-13206.
4. Alves,J., Pingoud,A., Langowski,J., Urbanke,C. and Maas,G. (1982) Eur. J. Biochem. 124, 139-142.
5. Mayer,H. and Reichenbach,H. (1978) J. of Bacteriol. 136, 708-713.

6. Schwarz,E., Scherer,G., Hobom,G. and Kössel,H. (1980) Biochem. Int. 1, 386-394.
7. Landsmann,J., Kröger,M. and Hobom,G. (1982) Gene 20, 11-24.
8. Kröger,M. and Hobom,G. (1982) Gene 20, 25-38.
9. Sanger,F., Coulson,A.R., Hong,G.F., Hill,B.F. and Peterson, G.B. (1982) J. Mol. Biol. 162, 729-773.
10. Hobom,G., Schwarz,E., Melzer,M. and Mayer,H. (1981) Nucleic Acids Res. 9, 4823-4832.
11. Maxam,A.M. and Gilbert,W. (1980) in Methods in Enzymology, Grossman,L. and Moldave,K., Eds., Vol. 65, pp. 499-560, Academic Press, New York.
12. Kröger,M. and Kröger-Block,A. (1984) Nucleic Acids Res. 12, 113-120; 193-201.
13. Hughes,S.G. and Murray,K. (1980) Biochem. J. 185, 65-75.
14. Fuchs,C., Rosenvold,E.C., Honigman,A. and Szybalski,W. (1978) Gene 4, 1-23.
15. Arrand,J.R., Myers,P.A. and Roberts,R.J. (1978) J. Mol. Biol. 118, 127-135.
16. DeWaard,A., Korsuize,J., van Beveren,C.P. and Maat,J. (1978) FEBS Letters 96, 106-110.
17. Mayer,H., Grosschedl,R., Schütte,H. and Hobom,G. (1981) Nucleic Acids Res. 9, 4833-4845.
18. Tu,C.-P.D., Roychoudhury,R. and Wu,R. (1976) Biochem. Biophys. Res. Comm. 72, 355-362.
19. Brown,N.L., McClelland,M. and Whitehead,P.R. (1980) Gene 9, 49-68.
20. Sugisaki,H., Maekawa,Y., Kanazawa,S. and Takanami,M. (1982) Nucleic Acids Res. 10, 5747-5752.
21. Bickle,T.A. and Ineichen,K. (1980) Gene 9, 205-212.
22. Kröger,M. and Hobom,G. (1984) Nucleic Acids Res. 12, 887-899.