Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains

Qing Xu*, Richard D. Morgan†, Richard J. Roberts†, and Martin J. Blaser*‡§

*Department of Microbiology and Immunology and ‡Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, TN 37232; and †New England Biolabs, Inc., Beverly, MA 01915

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved June 16, 2000 (received for review March 15, 2000)

A total of 22 type II restriction endonucleases with 18 distinct specificities have been identified in six *Helicobacter pylori* strains. Among these 18 specificities are three completely new endonucleases, *Hpy*178III, *Hpy*99I, and *Hpy*188I, that specifically cleave DNA at TCNNGA, CGWCG, and TCNGA sites, respectively. The set of endonucleases identified in each strain varies, but all have four- or five-base recognition sequences. Among 16 *H. pylori* strains, examination of the DNA modification status at the recognition sites of 15 restriction endonucleases reveals that each strain has a substantially different complement of type II modification systems. We conclude that the type II restriction-modification systems in *H. pylori* are highly diverse between strains, a unique characteristic of *H. pylori*. The diverse methylation status of *H. pylori* chromosomal DNA may serve as a new typing system to discriminate *H. pylori* isolates for epidemiological and clinical purposes. This study also demonstrates that *H. pylori* is a rich source of type II restriction endonucleases.

Type II restriction-modification (R-M) systems are pairs of enzymes: one, a restriction endonuclease (RE); the other, a methyltransferase, with opposing intracellular activities (1, 2). The RE recognizes specific sequences in DNA and cleaves the DNA at a particular site whereas the cognate methyltransferase modifies DNA within the same recognition sequence, thereby preventing cleavage by the RE. By possessing these two opposing enzymes, bacteria may protect their own DNA and degrade foreign DNA, thus limiting the spread of invading DNA molecules within the bacterial population (3–6). In addition, DNA methylation may be involved in regulation of chromosomal DNA replication (7, 8) and gene expression (9, 10), transposon movement (11), or DNA mismatch repair (12).

Helicobacter pylori are bacteria that colonize the human stomach and increase the risk of development of peptic ulcer disease and gastric adenocarcinoma (13–15). Analysis of the entire genomic sequence of *H. pylori* strains 26695 and J99 predicted that they possess an unusually high number (14 or 15) of type II R-M systems (16, 17). Comparison of these two strains showed that the two genomes are quite similar, with only 6–7% strain-specific genes (17); R-M systems are a major source of the strain differences. Furthermore, comparison of strains J166 and 26695 using a PCR-based subtractive hybridization method (18) showed that 7 of 18 J166-specific DNA clones had homology to the genes of R-M systems. In total, these studies demonstrated that, at the genomic level, R-M systems are quite diverse among the three strains examined. Whether this diversity is a general phenomenon among various *H. pylori* strains and which functional R-M systems are present in these strains remain as unanswered questions. Thus far, only the *iceA-hpyIM* locus in *H. pylori*, an ulcer-related *Nla*III-like type II R-M system (19, 20), has been studied among various strains, which indicates that the M gene is conserved whereas the R gene (*iceA1*) is not. To investigate whether *H. pylori* strains indeed have diversified R-M systems, we sought to study the *H. pylori* R-M systems by identifying type II REs at the biochemical level in six strains. In complementary studies, we also have examined the DNA modification status at multiple RE recognition sites in a variety of *H. pylori* strains.

Materials and Methods

Bacterial Strains, Growth Conditions, and Reagents. The bacterial strains used in this study are listed in Table 1. *H. pylori* strains were grown under conditions described (20). Digestion buffers were from New England Biolabs. All columns used for protein purification were obtained from Amersham Pharmacia, unless specifically indicated.

Restriction Endonuclease Assay. Chromatography fractions were assayed for RE activity by incubation at 37°C in NEB buffer 4 (50 mM KOAc/20 mM Tris·OAc/10 mM $Mg(OAc)_2$/1 mM DTT, pH 7.9), using phage λ, or ϕX174 DNAs as substrates, and were further examined by agarose gel electrophoresis. One unit of an RE is defined as the amount of the enzyme required to cleave completely 1 μg of ϕX174 DNA in 50 μl of NEB buffer 4 within 1 h at 37°C.

Purification of Restriction Endonucleases from Various *H. pylori* Strains. *Preparation of crude cell extracts of H. pylori. H. pylori* cells, grown in Brucella broth for 48 h, were harvested by centrifugation and were resuspended in sonication buffer (20 mM Tris·HCl/0.5 mM EDTA/1 mM DTT pH 7.5), then were sonicated until ≈50 mg of protein per gram of cells was released. The lysate was centrifuged, and the supernant was collected for column chromatography. Based on the characteristics of REs present in each strain, a different set of columns was used for the purpose of chromatography, as determined experimentally.

Chromatography of 60190 crude cell extract. The crude extract made from 15 g of 60190 cells was first applied to a 20-ml heparin HyperD column (Biosepra, Marlborough, MA), which was eluted with a linear gradient of 0.05–1.0 M NaCl in a total volume of 200 ml. After assay, the fractions containing RE activity were pooled and applied to either 3-ml Heparin-TSK (Tosohaas, Philadelphia) or 6 ml Source Q-15 columns. The TSK column was then eluted with a 50-ml linear gradient of either 0.05–0.60 M or 0.20–0.85 M NaCl whereas the Source Q-15 column was eluted with a 100-ml linear gradient of 0.05–1.0 M NaCl. At this stage, all REs in 60190 were substantially purified and ready for characterization.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: R, restriction; M, modification; RE, restriction endonuclease; HAP, hydroxylapatite.

§To whom reprint requests should be addressed at present address: New York University School of Medicine, Department of Medicine, New York, NY 10016. E-mail: martin.blaser@med.nyu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. *H. pylori* strains used in this study

H. pylori strain	Relevant genotype	Origin	Age/Sex*	Ethnicity*	Diagnosis*
60190	*iceA1*, *cagA*$^+$, *vacA* s1a m1	U.K.	NA	NA	NA
J166	*iceA1*, *cagA*$^+$, *vacA* s1b m1	U.S.	56M	Caucasian	Duodenitis
CH4	*iceA1*, *cagA*$^+$, *vacA* s1a m2	China	NA	Oriental	NA
9366	*iceA1*, *cagA*$^+$, *vacA* s2 m2	U.S.	57M	Caucasian	DU
A101	*iceA1*, *cagA*$^+$, *vacA* s1a m2	U.S.	75M	Caucasian	NUD
26695	*iceA1*, *cagA*$^+$, *vacA* s1a m1	U.K.	NA	NA	NA
97690	*iceA1*, *cagA*$^+$, *vacA* s1a m1	India	45M	Caucasian	NUD
HPK5	*iceA1*, *cagA*$^+$, *vacA* s1a m1	Japan	NA	Oriental	NA
J195	*iceA2*, *cagA*$^-$, *vacA* s2 m2	U.S.	44M	Caucasian	NUD
J178	*iceA2*, *cagA*$^+$, *vacA* s1a m1	U.S.	61M	African	CA
J188	*iceA2*, *cagA*$^-$, *vacA* s2 m2	U.S.	54M	Caucasian	NUD
J262	*iceA2*, *cagA*$^-$, *vacA* s2 m2	U.S.	63M	Caucasian	NUD
8822	*iceA2*, *cagA*$^-$, *vacA* s2 m2	U.S.	NA	NA	NA
97762	*iceA2*, *cagA*$^+$, *vacA* s1a m1	India	62F	Oriental	NUD
98518	*iceA2*, *cagA*$^+$, *vacA* s1a m1	India	30F	Oriental	NUD
J99	*iceA2*, *cagA*$^+$, *vacA* s1b m1	U.S.	48M	Caucasian	DU

*Patient information: DU, duodenal ulcer; NUD, non-ulcer dyspepsia; CA, gastric cancer; NA, data not available, reference strains.

Chromatography of J178 crude cell extract. The crude extract of ≈4 g of J178 cells was applied to a 5-ml Heparin Hitrap column, which was eluted with a 50-ml linear gradient of 0.12–1.0 M NaCl. After endonuclease assay, the eluted fractions were combined into two pools, 0.50–0.60 M and 0.65–0.85 M NaCl, which were applied to 3-ml TSK columns separately. The TSK columns were eluted with a 60-ml linear gradient of either 0.10–1.0 M or 0.15–1.0 M NaCl. For further purification, the TSK fractions between 0.35 and 0.48 M NaCl were pooled and applied to a 2-ml Mono Q column, which was eluted with a 40-ml linear gradient of 0.05–0.60 M NaCl.

Chromatography of J188 crude cell extract. The crude extract of 8 g of J188 cells was applied to a 20-ml Heparin HyperD column, which was eluted under the same conditions as described above. Endonuclease assay on eluted fractions revealed the presence of *Hpy*188I and one other specificity (21). Purification of *Hpy*188I has been described (21). To purify the second RE, the RE-containing fractions were applied to a 7-ml Mono S column, which was eluted with a linear gradient of 0.10–0.80 M NaCl.

Chromatography of J99 crude cell extract. The crude extract of 3.2 g of J99 cells was applied to a 20-ml Heparin HyperD column, which was eluted under the same conditions as described above. The bound REs were eluted from the column between 0.35 and 0.44 M, 0.48 and 58 M, and 0.9 and 1.2 M NaCl, respectively. The fractions in each of these ranges were pooled and applied to a 3-ml Heparin TSK column separately. The TSK columns each were eluted with a 45-ml linear gradient of 0.05–0.60 M, 0.10–0.70 M, or 0.50–1.5 M NaCl. A 7-ml hydroxylapatite (HAP) column was used to purify REs in Hyper D flow-through fractions and was eluted with a 70-ml linear gradient of 0.05–0.40 M KPO_4 (pH 7.4) in the presence of 0.20 M NaCl.

Chromatography of CH4 crude cell extract. The crude extract of 8 g of CH4 cells was applied to a 20-ml Heparin HyperD column and was eluted under the same conditions as described above. Four peaks of RE activities were detected among fractions eluted from the column. After dilution, each pool was applied to either 3-ml Heparin-TSK or 2-ml Mono S columns. The TSK column was eluted by using 55 ml of a linear gradient of 0.05–0.60 M, 0.12–0.60 M, or 0.15–0.70 M NaCl whereas the Mono S column was eluted with a 0.05–0.60 M NaCl linear gradient. After assay, RE-containing fractions were further purified over either 7-ml SP-Sepharose or 2-ml Mono Q columns, which were eluted by linear gradients of NaCl (0.05–0.80 M or 0.05–0.60 M).

Chromatography of J166 crude cell extract. The crude extract of 5.5 g of J166 cells was applied to a 20-ml Heparin HyperD column and was eluted under the same conditions as described above. RE fractions eluted from the column were pooled. Further purification used a 3-ml Heparin TSK column, which was eluted with a 45-ml linear gradient of 0.10–0.70 M NaCl. The flow-through fraction was applied to a 7-ml HAP column, eluted with a linear gradient of 0.05–0.40 KPO_4. The RE fractions from the HAP column were applied to a 2-ml Mono Q column, then to a 1-ml Heparin Hitrap column, which was eluted with a 30-ml gradient of 0.05–0.70 M NaCl.

Determination of Recognition Sequences of Purified Restriction Endonucleases. The recognition sequences of purified REs were determined as described (21).

Determination of the DNA Modification Status of Various *H. pylori* Strains. Chromosomal DNAs from *H. pylori* were prepared as described (20) and were digested with purified REs, under the conditions used for assay of these purified *H. pylori* REs. The digested DNAs were subjected to electrophoresis through 1% agarose gels.

Results

Purification and Identification of Restriction Endonucleases in *H. pylori*. *Endonucleases from 60190.* We first sought to study the type II R-systems in *H. pylori* strains, using column chromatography. To purify REs from strain 60190, a crude extract was applied to Heparin Hyper D, Heparin TSK, and Source Q-15 columns. A total of four different type II REs, termed *Hpy*II, *Hpy*IV, *Hpy*V, and *Hpy*VIII, were purified from strain 60190. *Hpy*V, *Hpy*II, and *Hpy*VII were eluted from the Heparin TSK columns in the 0.07–0.15 M, 0.34–0.37 M, and 0.40–0.48 M NaCl fractions, respectively, whereas *Hpy*IV was identified in the 0.29–0.32 M NaCl fractions eluted from the Source Q-15 column. Mapping experiments indicated that the recognition sequence of *Hpy*II is GAAGA (isoschizomer of *Mbo*II); of *Hpy*IV is GANTC (*Hin*fI); of *Hpy*V is TCGA (*Taq*I); and of *Hpy*VIII is CCGG (*Hpa*II) (Table 2). *Hpy*V had the highest total activity (≈2.7×10^4 units), followed by *Hpy*IV, *Hpy*VIII, and *Hpy*II (≈7,800, ≈3,400, and ≈2,500 units, respectively).

Endonucleases from J178. A crude extract of J178 was applied to Heparin Hitrap, Heparin TSK, and Mono Q columns, before four REs, designated *Hpy*178II, *Hpy*178III, *Hpy*178VI, and *Hpy*178VII, were completely separated. *Hpy*178III and *Hpy*178VI were eluted between 0.35 and 0.38 M or 0.27 and 0.30 M NaCl from the TSK columns, respectively, whereas *Hpy*178II was eluted around 0.16 M NaCl from the Mono Q column, and *Hpy*178VII was in the flow-through fractions of the column. Mapping experiments

Table 2. Restriction endonucleases isolated from six *H. pylori* strains

H. pylori strain	Designation of endonuclease	Recognition site	Isoschizomer	Units of activity/g cells
60190	*Hpy*II	GAAGA	*Mbo*II	1.7×10^2
	*Hpy*IV	GANTC	*Hinf*I	5.2×10^2
	*Hpy*V	TCGA	*Taq*I	1.8×10^3
	*Hpy*VIII	CCGG	*Hpa*II	2.3×10^2
J178	*Hpy*178II	GAAGA	*Mbo*II	2.3×10
	*Hpy*178III	TCNNGA	None	1.5×10^2
	*Hpy*178VI	GGATG	*Fok*I	2.3×10^2
	*Hpy*178VII	GGCC	*Hae*III	2.8×10^3
J188	*Hpy*188I*	TCNGA	None	1.3×10^3
	*Hpy*188III	TCNNGA	*Hpy*178III	6.2×10
J99	*Hpy*99I	CGWCG	None	1.6×10
	*Hpy*99II	GTSAC	*Tsp*45 I	8.0×10^2
	*Hpy*99III	GCGC	*Hha*I	2.9×10^3
CH4	*Hpy*CH4I	CATG	*Nla*III	1.4×10^2
	*Hpy*CH4II	CTNAG	*Dde*I	1.6×10^2
	*Hpy*CH4III	ACNGT	*Tsp*4CI	1.4×10^3
	*Hpy*CH4IV	ACGT	*Mae*II	3.1×10
	*Hpy*CH4V	TGCA	*Cvi*RI	2.1×10^2
	*Hpy*CH4VI	TCNNGA	*Hpy*178III	4.5×10
J166	*Hpy*166I	TCNGA	*Hpy*188I	2.4×10^3
	*Hpy*166II	GTNNAC	*Mja*IV	7.3
	*Hpy*166III	CCTC	*Mnl*I	5.4

Enzymes with new specificities are written in bold.
*Purification and cloning of this restriction system are described in ref. 21.

showed that *Hpy*178III was an RE with a specificity TCNNGA; *Hpy*178II, also present in *H. pylori* 60190, is an isoschizomer of *Mbo*II; *Hpy*178VI recognizes GGATG (*Fok*I); and *Hpy*178VII recognizes GGCC (*Hae*III) (Table 2). The complement of REs in J178 was different from that in 60190, and, as in strain 60190, the enzymatic levels of the REs isolated from J178 varied substantially. In J178, *Hpy*178VII was the most abundant RE detected, with $\approx 1.14 \times 10^4$ units, followed by *Hpy*178VI, *Hpy*178III, and *Hpy*178II ($\approx$900, $\approx$600, and $\approx$90 units, respectively).

Endonucleases from J188. A crude extract of J188 was applied to Heparin Hyper D and Mono S columns. In total, only two REs, *Hpy*188I and *Hpy*188III, were identified from strain J188. *Hpy*188I is an RE with the specificity TCNGA (21). *Hpy*188III was eluted around 0.22 M NaCl from the Mono S column. Mapping experiments indicated that *Hpy*188III recognizes TCNNGA, the phenotype also found in J178 (Table 2). *Hpy*188I was the most abundant RE ($\approx 10^4$ units), followed by *Hpy*188III ($\approx$500 units).

Endonucleases from J99. To identify type II REs from the sequenced strain J99, a crude cell extract was applied to Heparin Hyper D, Heparin TSK, and HAP columns. In total, three REs, designated *Hpy*99I, *Hpy*99II, and *Hpy*99III, were isolated from strain J99. *Hpy*99I was eluted from the TSK column between 0.35 and 0.40 M NaCl whereas *Hpy*99II and *Hpy*99III eluted at $\approx$0.54 M or $\approx$0.90 M NaCl, respectively. Mapping of their recognition sites indicated that *Hpy*99I is an RE with the specificity CGWCG; *Hpy*99II recognizes GTSAC (an isoschizomer of *Tsp*45I); and *Hpy*99III recognizes GCGC (*Hha*I) (Table 2). Based on the total activities isolated, *Hpy*99III was the most abundant RE ($\approx$9,400 units), followed by *Hpy*99II ($\approx$2,550 units) and *Hpy*99I ($\approx$50 units).

Endonucleases from CH4. In addition to strains from the U.S. and Europe, we also screened an Asian strain, CH4, for its RE activities. Heparin Hyper D, Heparin TSK, Mono S, SP-Sepharose, and Mono Q columns were used for purification. In total, six REs, designated *Hpy*CH4I, *Hpy*CH4II, *Hpy*CH4III, *Hpy*CH4IV, *Hpy*CH4V, and *Hpy*CH4VI, were isolated. *Hpy*CH4I, *Hpy*CH4IV, and *Hpy*CH4V were eluted between 0.13 and 0.23 M, 0.42 and 0.44 M, and 0.52 and 0.59 M NaCl, from the TSK columns, whereas *Hpy*CH4II and *Hpy*CH4VI were eluted around 0.15 M and 0.33 M NaCl from the Mono Q column. Furthermore, *Hpy*CH4III was eluted from the Mono S column between 0.35 and 0.42 M NaCl. Mapping of their recognition sites showed that *Hpy*CH4I recognizes CATG (an isoschizomer of *Nla*III); *Hpy*CH4II recognizes CTNAG (*Dde*I); *Hpy*CH4III recognizes ACNGT (*Tsp*4CI); *Hpy*CH4IV recognizes ACGT (*Mae*II); *Hpy*CH4V recognizes TGCA (*Cvi*RI); and *Hpy*CH4VI recognizes TCNNGA, the specificity also identified in strains J178 and J188 (Table 2). *Hpy*CH4III had the highest activity ($\approx 1.1 \times 10^4$ units), followed by *Hpy*CH4V, *Hpy*CH4II, *Hpy*CH4I, *Hpy*CH4VI, and *Hpy*CH4IV ($\approx$1,700, $\approx$1,250, $\approx$1,100, $\approx$360, $\approx$250 units, respectively). Of these REs, five were not found in any other strain screened.

Endonucleases from J166. A crude extract of J166 was applied to Heparin Hyper D, Heparin TSK, HAP, Mono Q, and Heparin Hitrap columns, which permitted separation of three REs, designated *Hpy*166I, *Hpy*166II, and *Hpy*166III. *Hpy*166I was detected in fractions between 0.36 and 0.50 M NaCl from the Heparin Hyper D column, while *Hpy*166II was eluted from Mono Q column at $\approx$0.22 M NaCl, whereas *Hpy*166II was eluted from the Hitrap column around 0.20 M NaCl. Mapping experiments showed that *Hpy*166I recognizes TCNGA, the specificity identified in strain J188 (21); *Hpy*166II recognizes GTNNAC (an isoschizomer of *Mja*IV); and *Hpy*166III recognizes CCTC (*Mnl*I) (Table 2). Among the REs isolated, *Hpy*166I has the highest activity ($\approx 1.3 \times 10^4$ units), followed by *Hpy*166III ($\approx$40 units); and *Hpy*166II ($\approx$30 units). *Hpy*166II and *Hpy*166III were not identified in any of the other five strains.

Examination of the DNA Modification Status in *H. pylori* Strains. In total, we isolated 22 type II REs with 18 distinct specificities from six *H. pylori* strains. Each strain had a varying complement of REs, suggesting considerable variation in R-M systems among *H. pylori* strains. However, an alternative explanation is that variable expression of a common set of R-M genes may be occurring, as seems to be the case in *Neisseria gonorrhoeae* (22). To assess this possibility, we examined the modification status of 16 clinical isolates of *H. pylori*, including the six strains used above for RE screening. These strains differ in their genotypes at the *iceA* locus; *iceA1* is a homolog of the *Nla*III gene whereas *iceA2* is unrelated (19). Eight strains each are *iceA1* or *iceA2* genotype. Chromosomal DNA was prepared from these strains and was digested with 15 of the 18 distinct

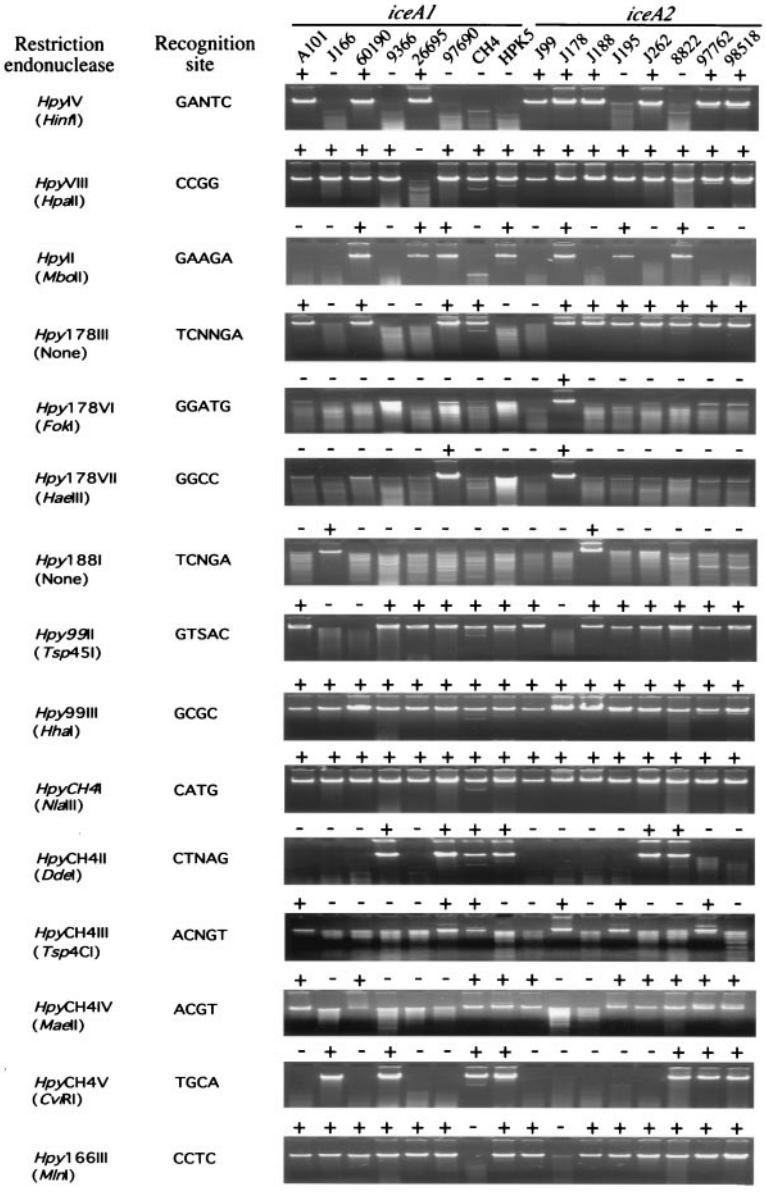

Fig. 1. Digestion of chromosomal DNA from 16 *H. pylori* strains by 15 different type II REs. The digests, resolved in a 1% agarose gel, are listed in the right panel. "+" indicates digestible by the RE whereas "−" indicates resistant to digestion. The name, recognition site, and prototype of each RE are listed in the left panel. The sources of the chromosomal DNA are listed at the top of the figure. The *iceA* genotypes of the DNA are also indicated.

type II RE specificities we purified (Fig. 1). The other three specificities (*Hpy*V, *Hpy*99I, and *Hpy*166II) could not be used, because most activity was lost during storage. Substantially different patterns of chromosomal DNA resistance to digestion by the 15 REs were observed among these *H. pylori* strains. Only two of the REs tested, *Hpy*99II and *Hpy*CH4I, failed to digest any of the DNAs, indicating that all chromosomal DNAs were modified at GCGC and CATG sites, and confirming our previous observation that *hpyIM* is conserved (20). Another 13 recognition sites examined are modified in some, but not all, of the strains (Fig. 1; Table 3). Modification at the recognition site of *Hpy*VIII (CCGG) was present in 15 of the 16 DNAs tested. Only DNA from strain 26695 was digestible, indicating that all strains, except 26695, carried a methyltransferase with the specificity of M.*Hpy*VIII. Modification at the *Hpy*166III site (CCTC) also was common, because 14 of the 16 DNAs were resistant to digestion by *Hpy*166III. In contrast, modification at GGATG, the *Hpy*178VI site, was the least common, present only in strain J178, from which *Hpy*178VI was isolated. Modification at GGCC and TCNGA, the sites of *Hpy*178VII and *Hpy*188I, also was uncommon. Two strains, 97690 and J178, were protected against *Hpy*178VII; and two other strains, J166 and J188, against *Hpy*188I. Modification at the remaining nine recognition sites occurs in an intermediate number of strains (6–11 of the 16 strains) (Table 3).

The complement of the M systems in each strain appears different from that of the 15 others (Table 3). Among the 16 chromosomal DNAs tested, those from strains 97690, 8822, and 97762 were modified to the highest degree, with 10 of the 15 RE sites modified. In contrast, the DNA from two strains, J166 and 26695, was modified at only six of the 15 sites tested. We further compared the modification status by plotting the number of specific recognition sites shared between individual strains (Table 4); the shared sites range from 5 to 14. Three pairs of strains, A101/97762, 97762/98518 and HPK5/8822, possessed

Table 3. Modification status of chromosomal DNA from 16 *H. pylori* strains

Modification site	*H. pylori* strains examined†															
	A101	J166	60190	9366	26695	97690	CH4	HPK5	J99	J178	J188	J195	J262	8822	97762	98518
GANTC	+	−	+*	−	+	−	−	−	+	+	+	−	+	−	+	+
CCGG	+	+	+*	+	−	+	+	+	+	+	+	+	+	+	+	+
GAAGA	−	−	+*	−	+	+	−	+	−	+*	−	+	−	+	−	−
TCNNGA	+	−	+	−	−	+	+*	−	−	+*	+*	+	+	+	+	+
GGATG	−	−	−	−	−	−	−	−	−	+*	−	−	−	−	−	−
GGCC	−	−	−	−	−	+	−	−	−	+*	−	−	−	−	−	−
TCNGA	−	+*	−	−	−	−	−	−	−	−	+*	−	−	−	−	−
GTSAC	+	−	−	+	+	+	+	+	+*	−	+	+	+	+	+	+
GCGC	+	+	+	+	+	+	+	+	+*	+	+	+	+	+	+	+
CATG	+	+	+	+	+	+	+*	+	+	+	+	+	+	+	+	+
CTNAG	−	−	−	+	−	+	+*	+	−	−	−	−	+	+	−	−
ACNGT	+	−	−	−	−	+	+*	−	−	+	−	+	−	−	+	−
ACGT	+	−	+	−	−	−	+*	+	+	−	−	+	+	+	+	+
TGCA	−	+	−	+	−	−	+*	+	−	−	−	−	−	+	+	+
CCTC	+	+*	+	+	+	+	−	+	+	−	+	+	+	+	+	+

''+'' indicates ''modified,'' and ''−'' indicates ''not modified.''

*Restriction enzymes isolated from each of the six strains (J166, 60190, CH4, J99, J178, and J188) correspond to the methylation in their DNA.

†Strains from A101 to HPK5 have the *iceA1* genotype; strains from J99 to 98518 have the *iceA2* genotype.

the most similar M systems; their modification status at the 15 sites tested are the same for each pair except for one site TGCA or GAAGA, respectively. In contrast, strain J178 shares with strains 9366 or HPK5 modification at only 5 of the 15 sites tested. The geographical relationship of M systems was assessed among these strains from the U.S., Europe, and Asia. Strains originating from the same geographic area generally do not share more methylation patterns than geographically unrelated strains (Tables 1 and 4). Overall, our studies on the 16 *H. pylori* strains indicates that *H. pylori* has many different active M systems and confirms the diversity of R-M systems among strains.

Discussion

Our screening of six *H. pylori* strains reveals a total of 22 different type II REs (Table 2), which indicates that *H. pylori* is rich in type II restriction systems. The enzymatic activities of these REs are varied, but they are in a normal range, when compared with those in other bacteria. Each *H. pylori* strain tested had a different complement of REs, and, in each of these six strains, modification of additional sequences was observed. Furthermore, in a survey of 16 *H. pylori* strains for the presence of the 15 type II M systems, the modification profiles were found to be substantially different (Fig. 1; Table 3). Overall, these data demonstrate that R-M systems are highly diversified in *H. pylori* strains. Comparison of the complete genome sequences of strains 26695 and J99 shows that the two strains share 11 of the 15 type II R-M genes (17), and comparable differences between them were observed in our study. We found six functional M systems in 26695, and seven in J99. The two strains share five systems, including the M systems of *Hpy*99III and *Hpy*CH4I, that are conserved among all strains tested. Our demonstration of highly diversified R-M systems at a functional level is consistent with studies at the genomic level in strains 26695, J99, and J166 (16, 17). Our previous studies of two *H. pylori* R-M systems (*Hpy*188I and *Hpy*I) also demonstrate substantial interstrain variation (19–21). First, the *Hpy*188I R-M genes are not present in 9 of 11 strains tested (21). Second, although *hpyIM* is functionally conserved in all strains tested (20), its cognate RE gene, *iceA1*, is present in less than half of the strains, whereas the entirely unrelated *iceA2* occupies the same locus in other strains (19, 23).

The total numbers of REs identified in each of the six strains vary from two to six. Thus far, CH4 represents the *H. pylori* strain with the largest number of REs identified at the biochemical level. Although it was predicted that strain J99 has 15 type II R-M systems (17), we found activity of only three REs from this strain. One explanation may be that some RE activity was actually present in J99, but could not be detected; such low

Table 4. Sharing of chromosomal DNA modification status among 16 *H. pylori* strains at 15 different recognition sequences

H. pylori strain	J166	60190	9366	26695	97690	CH4	HPK5	J99	J178	J188	J195	J262	8822	97762	98518
A101	8	12	9	10	10	11	9	13	9	12	13	13	10	14	13
J166		9	12	9	7	8	10	10	6	11	8	8	9	9	10
60190			8	11	9	8	10	12	10	11	12	12	11	11	12
9366				10	10	11	13	11	5	10	9	11	12	10	11
26695					9	6	10	12	8	11	10	10	9	9	10
97690						10	10	8	10	9	12	10	11	9	8
CH4							11	9	7	8	11	11	12	12	11
HPK5								11	5	8	11	11	14	10	11
J99									7	12	11	13	10	12	13
J178										8	9	7	6	8	7
J188											10	12	9	11	12
J195												11	12	12	11
J262													12	12	13
8822														11	12
97762															14

This table indicates the number of sites at which each pair of strains had the same DNA modification status.

activity could be attributable to a low level of expression, or poor enzyme stability. In any event, the detected RE activities in each of the six strains varied considerably (Table 2). Another explanation is that some of these R-M systems are degenerate and no longer functional, which clearly is the case for *iceA1* (23), an *Nla*III homolog in *H. pylori*. Most *iceA1* genes are prematurely truncated and are not predicted to encode a functional RE. Screening three *iceA1* strains in this study revealed that only CH4 encoded a functional *Nla*III-like RE, consistent with the fact that the *iceA1* gene in CH4 is intact whereas those in 60190 and J166 are degenerate with frame-shift or nonsense mutations (23).

H. pylori plasmids have been reported previously and are also evident in strains CH4, 97762, and 98518 in this study (Fig. 1). R-M systems may be encoded by plasmids in bacteria (24–26). However, cloning of the R-M genes from *H. pylori*, including the sequenced strains J99 and 26695 and CH4, which possesses an unusually high number of REs, shows that all of the genes are chromosomal (ref. 21; R.D.M., Q.X., and H. M. Kong, unpublished work). These data suggest that the *H. pylori* R-M systems are largely, if not exclusively, encoded chromosomally.

N. gonorrhoeae differs from other bacteria by possessing a high number of type II M systems and at least seven linked REs (22, 27–30). The genes of Type II M systems in this organism are generally conserved among strains and always are expressed to keep the host genome modified. In contrast, the cognate RE genes sometimes are silent, possibly because of specific regulation of their gene expression. Thus, although R-M systems in *N. gonorrhoeae* are similar at the gene level, expression of their REs varies among different isolates (22). *H. pylori* resembles *N. gonorrhoeae* in possessing an unusually high number of R-M systems (16, 17), but differs in that different sets of R-M systems are present in each strain at the gene level. Thus, *H. pylori* is unique because of the unusually high numbers and diversity of its R-M systems.

It is not clear why *H. pylori* possesses so many different R-M systems. One major function of R-M systems in bacteria is to protect the host from bacteriophage infections. An advantage of having diversified R-M systems among a population of strains is that when one strain is successfully infected by a phage, other strains, possessing different R-M systems, can be protected. Thus, the entire bacterial population will not be destroyed. We speculate that R-M systems must play other important roles in this organism. Because *H. pylori* is naturally competent (31), possessing a large number of R-M systems may prevent exogenous DNA from integrating into the host genomic DNA. All of the REs identified in *H. pylori* have 4- or 5-bp recognition sequences, suggesting that efficient digestion of DNA is required for this organism. There are two highly conserved methyltransferases, M.*Hpy*99II and M.*Hpy*I, among the 16 strains examined, modifying *H.pylori* DNA at GCGC and CATG sites, respectively. Strong conservation of these two methyltransferases in contrast to the diversified R-M systems in *H. pylori* suggests that GCGC and CATG-methylation may be involved in important cellular processes. For example, *dam* methylation (at GATC) is essential for regulation of chromosomal DNA replication and mismatch repair in other organisms (7–12). Modification at GATC has been found in most but not all of the 16 *H. pylori* strains examined (Q.X. and M. J. Blaser, unpublished work). It is conceivable that other methyltransferases, perhaps M.*Hpy*99II or M.*Hpy*I, play a similar role in *H. pylori*.

Whereas *H. pylori* is cosmopolitan and highly prevalent in most populations (13–15), fewer than 10% of people carrying the bacteria develop *H. pylori*-associated diseases. One explanation may be that *H. pylori* strains of particular genotypes are more virulent than other strains (32, 33). Typing systems based on genomic variability have been used to discriminate *H. pylori* isolates for epidemiological and clinical purposes (34–37). Because diversified R-M systems in *H. pylori* are major strain-specific genes (17), the modification status of chromosomal DNA may be used to develop a new system for typing *H. pylori* isolates. In this study, all 16 strains examined had unique DNA modification profiles, demonstrating the potential discriminatory power of this typing system. A typing scheme based on DNA modification status may become a simple, reproducible, and discriminatory means of documenting diversity within *H. pylori* populations.

Finally, although numerous Type II REs with differing specificities are currently available, there is a continuing need for additional type II REs, because new enzymes recognizing novel sequences improve flexibility for genetic manipulation, and enable specialized applications (38). Finding REs with novel specificities is becoming increasingly difficult. However, because *H. pylori* has such an unusually high number of R-M systems that are diversified among strains, this organism could be an excellent source for the discovery of new type II REs. Thus far, nearly 30 type II REs with a total of 22 distinct specificities (39, 40), including those from this study, have been identified from *H. pylori*. Importantly, from a screen of six *H. pylori* strains, we found three novel type II REs, an unusually high number. It seems possible that more REs with novel specificities will be identified from *H. pylori*, as more screening efforts are made.

This work was supported in part by the Dissertation Enhancement Grant from the Vanderbilt Graduate School and National Institutes of Health Grants GM56534 and DK53707.

1. Roberts, R. J. & Halford, S. E. (1993) in *Nucleases*, eds, Linn, S. M., Lloyd, S. & Roberts, R. J. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 35–88.
2. Wilson, G. G. & Murray, N. E. (1991) *Annu. Rev. Genet.* **25,** 585–627.
3. Arber, W. & Dussoix, D. (1962) *J. Mol. Biol.* **5,** 18–36.
4. Dussoix, D. & Arber, W. (1962) *J. Mol. Biol.* **5,** 37–49.
5. Kuhnlein, U. & Arber, W. (1972) *J. Mol. Biol.* **63,** 9–19.
6. Smith, J. D., Arber, W. & Kuhnlein, U. (1972) *J. Mol. Biol.* **63,** 1–8.
7. Messer, W. & Noyer-Weidner, M. (1998) *Cell* **54,** 735–737.
8. Smith, D. W., Garland, A. M., Herman, G., Baker, T. A. & Zyskind, J. W. (1985) *EMBO J.* **4,** 1319–1326.
9. Braaten, B. A., Nou, X, Kaltenbach, L. S. & Low, D. A. (1994) *Cell* **76,** 577–588.
10. Heithoff, D. M., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. (1999) *Science* **284,** 967–970.
11. Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. (1985) *Cell* **43,** 117–130.
12. Modrich, P. (1989) *J. Biol. Chem.* **264,** 6597–6600.
13. Berg, D. E. & Logan, R. P. H. (1997) *BioEssays* **19,** 86–90.
14. Blaser, M. J. & Parsonnet, J. (1994) *J. Clin. Invest.* **94,** 4–8.
15. Goodwin, C. S., Armstrong, J. A., Chilvers, T., Peters, M., Collins, M. D., Sly, L., McConnell, W. & Harper, W. (1989) *Int. J. Syst. Bacteriol.* **39,** 397–405.
16. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., *et al.* (1997) *Nature (London)* **388,** 539–547.
17. Alm, R. A., Ling, L. L., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Guild, B. C., deJonge, B. L., Carmel, G., *et. al.* (1999) *Nature (London)* **397,** 176–180.
18. Akopyants, N. S., Fardkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S., Sverdlov, E. D. & Berg, D. E. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 13108–13113.
19. Peek, R. M., Jr., Thompson, S. A., Atherton, J. C., Blaser, M. J. & Miller, G. G. (1998) *Proc. Assoc. Am. Phys.* **110,** 531–544.
20. Xu, Q., Peek, R. M., Miller, G. G. & Blaser, M. J. (1997) *J. Bacteriol.* **179,** 6807–6815.
21. Xu, Q., Stickel, S., Roberts, R. J., Blaser, M. J. & Morgan, R. D. (2000) *J. Biol. Chem.* **275,** 17086–17093.
22. Stein, D. C., Gunn, J. S., Radlinska, M. & Piekarowicz, A. (1995) *Gene* **157,** 19–22.
23. Figueiredo, C., Quint, W. G. V., Sanna, R., Sablon, E., Donahue, J. P., Xu, Q., Miller, G. G., Peek, R. M., Blaser, M. J. & van Doorn, L. J. (2000) *Gene* **246,** 59–68.
24. Zakharova, M. V., Beletskaya, I. V., Kravetz, A. N., Pertzev, A. V., Mayorov, S. G. & Solonin, A. S. (1998) *Gene* **208,** 177–182.
25. Lubys, A. & Janulaitis, A. (1995) *Gene* **157,** 25–29.
26. Wayne, J., Holden, M. & Xu, S. Y. (1997) *Gene* **202,** 83–88.
27. Korch, C., Hagblom, P. & Normark, S. (1983) *J. Bacteriol.* **142,** 1325–1332.
28. Norlander, L., Davies, J. K., Hagblom, P. & Normark, S. (1981) *J. Bacteriol.* **140,** 788- 795.
29. Piekarowicz, A. & Stein, D. C. (1995) *Gene* **157,** 101–102.
30. Ritchot, N. & Roy, P. H. (1990) *Gene* **76,** 103–106.
31. Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H. & Kunstmann, E. (1998) *Proc. Natl. Acad. Sci.* USA **95,** 12619–12624.
32. Cover, T. L., Glupczynski, Y., Lage, A. P., Burette, A., Tummuru, M. K., Perez-Perez, G. I. & Blaser, M. J. (1995) *J. Clin. Microbiol.* **33,** 1496–1500.
33. van Doorn, L. J., Quint, W. G. V., Schneeberger, P. M., Tytgat, G. N. J. & de Boer, W. A. (1997) *Lancet* **350,** 71–72.
34. Burucoa, C., Lhomme, V. & Fauchere, J. L. (1999) *J. Clin. Microbiol.* **37,** 4071–4080.
35. Kansau, I., Raymond, J., Bingen, E., Courcoux, P., Kalach, N., Bergeret, M., Braimi, N., Dupont, C. & Labigne, A. (1996) *Res. Microbiol.* **147,** 661–669.
36. Owen, R. J., Fraser, J., Costas, M., Morgan, D. & Morgan, D. R. (1990) *J. Clin. Pathol.* **43,** 646–649.
37. Taylor, N. S., Fox, J. G., Akopyants, N. S., Berg, D. E., Thompson, N., Shames, B., Yan, L., Fontham, E., Janney, F., Hunter, F. M., *et al.* (1995) *J. Clin. Microbiol.* **33,** 918–923.
38. Schiestl, R. H., Nicklen, S. & Petes, T. D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7585–7589.
39. Ivic, A., Jakeman, K. J., Penn, C. W. & Brown, N. L. (1999) *FEMS Microbiol. Lett.* **179,** 175–180.
40. Repin, V. E., Puchkova, L. I., Ananko, G. G., Reshetnikov, O. V. & Kurilovich, S. A. (1999) *Mol. Gen. Microbiol. Virusol.* **3,** 37–38.