Plasticity of Repetitive DNA Sequences within a Bacterial (Type IV) Secretion System Component

Rahul A. Aras, Wolfgang Fischer, Guillermo I. Perez-Perez, MariaLuisa Crosatti, Takafumi Ando, Rainer Haas, and Martin J. Blaser

1Department of Medicine and Department of Microbiology, New York University School of Medicine, and VA Medical Center, New York, NY 10016
2Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Ludwig-Maximilians-Universitat, 80336 Munich, Germany
3First Department of Internal Medicine, Nagoya University School of Medicine, 466-8550 Nagoya, Japan

Abstract
DNA rearrangement permits bacteria to regulate gene content and expression. In Helicobacter pylori, cagY, which contains an extraordinary number of direct DNA repeats, encodes a surface-exposed subunit of a (type IV) bacterial secretory system. Examining potential DNA rearrangements involving the cagY repeats indicated that recombination events invariably yield in-frame open reading frames, producing alternatively expressed genes. In individual hosts, H. pylori cell populations include strains that produce CagY proteins that differ in size, due to the predicted in-frame deletions or duplications, and elicit minimal or no host antibody recognition. Using repetitive DNA, H. pylori rearrangements in a host-exposed subunit of a conserved bacterial secretion system may permit a novel form of antigenic evasion.

Key words: Helicobacter pylori • antigenic variation • genome plasticity • pathogenicity island • type IV secretion system

Introduction
The ability of microbial parasites to persist in vertebrate hosts protected by adaptive immunity is a central biological problem. Antigenic variation of surface components is one strategy by which microbial parasites facilitate evasion of host immune responses during long-term colonization (1). Although microbial populations are constantly changing in all colonized niches (2), generation of antigenic diversity at rates faster that the immune system can respond is a challenge that many successful long-term colonizers overcome (3). DNA recombination between paralogous loci containing variable regions is an efficient mechanism for generating diversity, and is used by a wide range of microbes. Such recombination typically involves separate loci and multiple genes (3–6), or tandem repeats within an individual open reading frame (ORF: 7, 8).

Helicobacter pylori, Gram-negative, curved bacteria that colonize the human stomach for decades, if not for life, increase risk for peptic ulcer disease and gastric adenocarcinoma (9). Although H. pylori elicits vigorous host immune responses, ongoing over the full course of colonization (10, 11), the mechanisms for avoiding immune-based clearance remain unknown. H. pylori strains may contain a 40-kb pathogenicity island (termed the cag-island; 9, 12, 13). Strains that are cag+ induce increased IL-8 production by gastric epithelial cells in vitro (13, 14) and are associated with increased gastric mucosal cytokine production in vivo (15–17) and increased risk of developing peptic ulcer disease and gastric adenocarcinoma (18). The cag-island includes genes encoding subunits of bacterial type IV secretory systems (12–14, 19) that inject their substrate, the CagA protein, into host epithelial cells (20–22) altering specific signal transduction pathways (23, 24).

Type III and IV bacterial secretion systems typically involve the direct injection into host cells of macromolecules that are crucial to the lifestyle of the bacteria (25, 26). This interface represents a frontier important to understanding the evolution of virulence. One component of the H. pylori secretion system, cagY (HP0527), whose product has homol-
ology to Agrobacterium tumefaciens VirB10 (27), is required for gastric epithelial cell IL-8 induction (13) and CagA translocation into epithelial (28), indicating its important role in the H. pylori–host interaction.

Why is the host unable to mount an effective response against a bacterial constituent that so profoundly affects its physiology? An organism may reside in an immunologically privileged locale, preventing significant host immune recognition, but such is not the case with H. pylori (10). Alternatively, H. pylori may evade immune clearance through altering its superficial structures. The deduced cagY product from H. pylori strains 26695, J99, and NCTC11638 has striking NH₂-terminal and middle region amino acid repeat patterns, and a high multiplet count dominated by lysine- and glutamate-containing doublets (29). Therefore, we hypothesized that the DNA sequences of these repeat structures (29) represent opportunities for persistence of a potentially vulnerable and critical element of the system pilus (CagY), which would facilitate immune evasion of a potentially vulnerable and critical element of the persistent bacterial interaction with its host.

Materials and Methods

Bacterial Strains. In this study, 90 H. pylori isolates from persons born in different regions of the world (United States, 31; Netherlands, 26; Mexico, 5; Argentina, 3; Peru, 4; Columbia, 4; China, 8; Thailand, 3; Korea, 5; United Kingdom, 1) were obtained from a stock collection of H. pylori isolates from the NYU Helicobacter/Campylobacter strain reference center stored at −70°C in our laboratory. From H. pylori strain 84–183, 249 (107 low in vitro passage and 142 high in vitro passage) single colony isolates were examined. 14 strains belonging to a single clonal group had been obtained from 7 members of an extended family from The Netherlands (31). 12 other strains from The Netherlands were obtained from 6 patients from whom H. pylori was isolated from antral biopsy specimens at both an initial and follow up endoscopy 7–10 yr apart (32). We also studied 15 isolates of USA strain J166 that had been recovered from the stomach of a male rhesus monkey 10 mo after experimental challenge (33). Chromosomal DNA from each of the strains studied was prepared using a phenol extraction method (34). In addition, isolates of strain B128 (35) used in experimental mouse challenge were studied. Strains included the prechallenge isolate as well as gastric isolates from two FVB/N mice 8 mo after experimental challenge (36).

PCR Amplification and Restriction Fragment Length Polymorphism (RFLP) Analysis. Oligonucleotide primers and PCR conditions were used to amplify the 1.5-flaA and 1.7-kb spa-glmM gene segments, as previously described (37). The FRRF (5′-ATGAATGAAGAAAACGATAAATTG), FRRR (5′-CACCTGAACTTTTTGTGTTCTCAG), MRRF (5′-GCTTACTGAAACCAACAAAGTTCA), and MRRR (5′-CGCTCAAAACATCCTCAACCTTTC) primers were designed based on the cagY sequence of strain 26695 (27). PCR amplification was performed with a reaction mixture containing 100 ng template DNA, 200 ng of each primer, and 0.5 U Taq polymerase (QIAGEN) in a 100-µl volume. PCR included 45 cycles at 94°C for 1 min, 54°C for 1 min, and 72°C for 3 min. RFLP analysis was performed as previously described (37).

Analysis of Genomic Repeat Sequences. The Genetics Computer Group program, Repeat (Wisconsin Package Version 9.1), was used to assess the presence of DNA direct repeats ≥10 bp in five arbitrarily selected 5,000 bp regions in H. pylori strain 26695. Repeats of 10 (mean = 44.8 ± 4.7), 11 (mean = 14.2 ± 4.0), 12 (mean = 4.6 ± 1.1), 13 (mean = 1.2 ± 0.8), 14 (mean = 0.2 ± 0.4), and 15 bp (mean = 0.2 ± 0.4) were found. However, no repeat sequences ≥16 bp were found within the 25,000 bp examined, indicating that use of a threshold ≥16 bp minimizes or eliminates background repeats. Sequences of cagY in three independently isolated strains, 26695, J99, and NCTC11638 obtained from GenBank were examined using Repeat to identify identical direct repeats ≥16 bp flanking or within the gene. Locations of the repeat sequences as well as areas subject to deletion were mapped using Microsoft Excel.

cagY Sequence. For analysis of specific cagY sequences, relevant PCR products generated with primers FRRF and FRRR or MRRF and MRRR were purified using the Qiagen gel extraction kit (QIAGEN). The purified PCR products were then sequenced on both strands using an automated Applied Biosystems sequencer in the New York University Cancer Center Core Laboratory, and analyzed using Sequencer 3.1.1 (Gene Code Corp., Inc.).

Immunoblot Assay. H. pylori cells were harvested from 48-h plate cultures, resuspended in 500 µl loading buffer (5% SDS, 0.003% bromophenol blue, 20% glycerol, 0.5% DTT, 1.5% Tris base), and 10 µl protein samples were assayed by electrophoresis on a 7% sodium-dodecyl sulfate-polyacrylamide gel. CagY proteins were detected by chemiluminescence with polyclonal rabbit antisera (1:3,000 dilution) against the H. pylori strain J99 CagY subunit (38). The secondary antibody (1:2,000 dilution) was goat α-rabbit immunoglobulin G–alkaline phosphatase (Boehringer).

ELISA. All assays were performed as previously described (32). In brief, microtiter plates were coated with 500 ng of a recombinant 1,000 amino acid CagY fragment isolated from Esche- nrichia coli strain 2136 (2136/CagY-J99; reference 38). α-CagY hyperimmune rabbit serum or sera from H. pylori or H. pylori patients was used as the primary antibody. The secondary antibody was goat α-human IgG coupled with horseradish peroxidase (Sigma-Aldrich). In addition to controls without primary or secondary antibody, cells of E. coli strain 2136 were included in each assay. To determine specific binding, the optical density values from wells coated with E. coli 2136 were subtracted from wells containing E. coli 2136/CagY-J99 at the same protein concentration (500 ng per well), and all specimens were examined at least in duplicate in at least three separate assays.

Results

H. pylori cagY Possesses an Extraordinary Number of DNA Repeats. Addressing whether the extensively repeated CagY amino acid sequence (29) was reflected at the nucleotide level in three strains studied, a mean of 264 ± 34
The DNA repeats within cagY are clustered within two highly conserved areas, that we have termed the 5′ repeat region (FRR) and the middle repeat region (MRR; Fig. 1). Within each, the repeats frequently overlap so that a single cagY nucleotide in strain 26695 might be part of up to 10 different repeat sequences (Fig. 1 A). However, the downstream copies of the repeat sequences have entirely different overlap patterns. Based on our analyses, cagY can be divided into five regions conserved within the three strains: FRR, 5′ conserved region (FCR), MRR, 3′ conserved region (TCR), and VirB10 homology region (VHR; Fig. 1 C). Although VirB10 from A. tumefaciens and nine other organisms have high homology (BLAST E values <10−20) to the region encoded by the cagY VHR, multiple alignments indicate no significant homologies in GenBank to the regions upstream of the VHR. Similarly, separate BLAST searches of the four upstream regions revealed no significant homologies. G+C content of cagY upstream of the MRR is lower (32–35%) than the remainder of the gene (38–40%; Fig. 1 C) and the entire H. pylori chromosome (39%). The cagY ORFs in the three strains range from 5,172 to 5,784 bp. The major differences are a 329-bp deletion in the FRR of strain J99 and NCTC11638, as well as smaller MRR deletions that vary from strain to strain (Fig. 1 C). The FCR, TCR, and VHR each are highly conserved (DNA sequence identity of 96.3–97.3%) in the three strains, similar to the level observed amongst H. pylori housekeeping genes (43). Thus, although annotated as a VirB10 homologue (19, 27), further sequence analysis suggests a complex evolutionary history and the presence of other functional domains.

**cagY Intragenic Recombination Events Always Yield In-Frame Gene Products.** Accordingly, we next examined the regions with multiple DNA repeats. The presence of direct DNA repeats allows for deletion (or duplication) of the intervening region plus one copy of the repeat through DNA recombination, involving either slipped strand misalignment or intrachromosomal exchange (44). To understand the potential of the repeats to permit cagY gene structure changes, sequences subject to change in copy number by their location between repeats were determined. Two regions mapping to the FRR and MRR were observed (Fig. 1 B). Importantly, each is independent because no sequences ≥16 bp are present in both the FRR and MRR. Each strain has >300 different potential recombination events involving the MRR (Table I), predominantly mapping to the middle of the MRR (Fig. 1). Because bacteria often use recombination between repeat sequences for phase-variable gene regulation (1), random distances between paired repeats would yield in-frame deletions one

<table>
<thead>
<tr>
<th>H. pylori strain</th>
<th>Total number (independent)</th>
<th>Mean ± SD copies</th>
<th>Mean ± SD size (bp)</th>
<th>Number of regions</th>
<th>Range in fragment size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26695</td>
<td>302 (99)</td>
<td>3.1 ± 1.6</td>
<td>36.1 ± 22.8</td>
<td>443</td>
<td>111–2,187</td>
</tr>
<tr>
<td>J99</td>
<td>235 (78)</td>
<td>3.0 ± 1.7</td>
<td>48.2 ± 42.7</td>
<td>347</td>
<td>114–2,205</td>
</tr>
<tr>
<td>NCTC11638</td>
<td>255 (82)</td>
<td>3.1 ± 2.0</td>
<td>34.9 ± 23.5</td>
<td>425</td>
<td>114–2,544</td>
</tr>
</tbody>
</table>

*Table I. Characteristics of H. pylori cagY MRR Direct DNA Repeat Sequences and Their Recombination Potential*

1. The MRR of cagY in the three strains are shown in Fig. 1.
2. ≥16 nucleotides (refer to Materials and Methods).
3. Recombination potential defined as number of regions flanked by direct DNA repeats, and fragment size as intervening segment plus one repeat.
4. All potential recombination events involving the 1,215 regions identified (leading to either DNA deletion or duplication) result in in-frame ORF (P < 0.001).
third of the time. However, in the three strains studied, ev-ery one of the 1,215 potential recombination events pro-
duced an in-frame cagY, a highly (P = 4.0 \times 10^{-243}) non-
random phenomenon. Therefore, with two independent
regions, the number of potential recombination events
within each gene, in each case producing a full-length pro-
tein product, is enormous.

Variation in cagY Repeat Region Sizes. To determine the
extent of intergenic cagY repeat region variation, PCRs
were performed using primers that flank either the FRR or
MRR using 62 strains from locations around the world.
FRR products ranging from 1.1 to 2.0 kb were amplified
(Fig. 2). For most strains, MRR products were
2.8 kb, but all varied between 2 and 3 kb. The geographic origins
of individual strains and FRR or MRR size showed no cor-
relation (unpublished data). Thus, whereas cagY size varies
amongst different strains, there appear to be length bound-
aries on both FRR and MRR. To examine cagY stability in
vitro, MRRs of laboratory-passaged (since 1984) strain 84–
183 (84–183HP) and the original 84–183 (84–183LP; <10
passed since 1984) were examined by PCR. PCR of the
cagY MRR from all 107 single colony 84–183LP isolates
yielded a 2.5-kb product, whereas 142 single colony 84–
183HP isolates yielded 74 (52%) products of 2.5 kb, 64
(45%) products of 2.0 kb, and 4 (3%) colonies yielded 2.0,
2.2, and 2.5 kb products. Immunoblots, using α-CagY se-
rum, performed on cell lysates from the three different 84–
183HP colony types yielded CagY products whose sizes
were consistent with the corresponding cagY PCR products
from that colony (Fig. 3 A). These results indicate that dur-
ing in vitro passage, an H. pylori population develops that is
heterogeneous in its expressed CagY products, potentially
through deletions and/or duplications in cagY. To examine
in vivo size variability, the MRR from 14 highly related
isolates obtained from 7 members of an extended family in
The Netherlands reflecting a total of >150 person-years

---

Figure 1. Direct nucleotide repeats in the structure of H. pylori cagY. (A) The abscissa represents the nucle-
otide position in each ORF and the ordinate for each of three strains (26695, J99, NCTC1168) represents
the number of times each nucleotide is part of a direct repeat sequence ≥16 nucleotides. For each strain, the
repeats are clustered in two regions. (B) The abscissa represents the nucle-
otide position in each ORF and the ordinate represents the number of
times that identical repeats ≥16 bp flank each nucleotide. The presence
of a nucleotide between identical repeats indicates that it could be de-
leted or duplicated if recombination occurred between the repeats. (C) For
the three strains analyzed, cagY might be separated into five defined
regions present in each sequence, which we termed the FRR, the
FCR, the MRR, the TCR, and the VHR. None of the repeats are
shared in both the FRR and MRR. Downstream of both the FRR and
MRR are highly conserved regions of ~550 bp (FCR and TCR), fol-
lowed by the VHR portion of cagY, encoding a polypeptide highly homol-
ogous to the A. tumefaciens virB10 product. The FCR, TCR, and
VHR have high levels of DNA identity between strains, with FCR
size differences due to small variations in the boundaries of the adjacent
repeat regions. FRR size variation is largely due to a 329-bp deletion
(hatched region). The G+C content is lowest and deviates substantially
from the remainder of the genome in the portion of cagY upstream of the
MRR.
isolation from one another (33) were analyzed. Using 3 different restriction digestions, RFLP patterns for segments of \( \text{flaA} \) (unpublished data) and \( \text{glmM} \) (Fig. 3 B) were nearly identical for the 14 strains (Fig. 3 B), reconfirming the clonal origin of the familial isolates (31, 45). Although PCR of the \( \text{cagY} \) MRR yielded 2.8 kb amplicons for all 14 isolates, RFLP analysis of the products indicated at least 4 different patterns (Fig. 3 B). In two patients (6 and 7), isolates show differing MRR-RFLP patterns. These data provide evidence that the \( \text{cagY} \) MRR undergoes diversification in vivo, during the decades-long human colonization.

**Intrahost MRR Diversity Resulting from Deletion between Direct DNA Repeats.** Next, we sought to determine whether the direct DNA repeats detected by in silico analysis were in fact involved in size variation. Parallel studies were performed on pairs of isolates from antral biopsies from six patients over 7–10-yr intervals. Sequence, random amplification of polymorphic DNA (RAPD)-PCR, and AFLP analyses indicate that each pair of strains are clonal variants (32). From patient 13, PCR of the \( \text{cagY} \) MRR yielded a 2.8-kb product from the initial isolate (13aqs) and a 2.5-kb product from the isolate (13bqs) obtained 7.4 yr later (Fig. 4 A). Sequence analysis showed that in 13bqs, a 168-bp segment plus one of the flanking 156-bp direct repeats was deleted (Fig. 4 B). As predicted from analyses of \( \text{cagY} \) in the sequenced strains (Table I), deletion of this 324-bp segment allows for expression of an in-frame protein (Fig. 4 C). Analysis using PCR and immunoblots of \( H. pylori \)
pylori B128 isolates harvested 8 mo after experimental challenge of FVB/N mice confirmed that in vivo intragenomic recombination alters cagY sequence. RAPD and RFLP indicated the same clonal origin for preinoculation strain B128 and postchallenge gastric isolates obtained from mice 6.11 and 6.16 (unpublished data). By PCR using primers that flank the MRR, cells obtained from population sweeps and from five colonies from each strain yielded a shorter

Figure 4. Size and sequence diversity in the cagY MRR in paired H. pylori isolates obtained 7.4 yr apart from an individual host. (A) The cagY MRR was amplified from isolates 13aqs (original strain) and 13bqs (obtained 7.4 yr later). (B) Sequence analysis of the PCR products showed that relative to strain 13aqs, strain 13bqs has a 324-bp deletion including at its 5’ end a 156-bp sequence (underlined) identical to the 156-bp sequence immediately 3’ of the deletion. Both copies of the repeat are present in 13aqs. (C) Translation of 13aqs and 13bqs indicate that cagY is in-frame in both strains.

Figure 5. In vivo cagY MRR deletion occurs during experimental colonization of FVB/N mice. (A) CagY MRR PCR performed on single colonies from H. pylori strains before and after 8 mo of challenge of FVB/N mice. Lane 1, PCR control; lane 2, B128 prechallenge strain (sweep of colonies); lanes 3–7, single colonies from B128 prechallenge strain; lane 8, cells recovered from mouse 6.11 (sweep of colonies); lanes 9–13, single colonies from mouse 6.11; lane 14, cells recovered from mouse 6.16 (sweep of colonies); lanes 15–18, single colonies from mouse 6.16. (B) Sequence analysis of the PCR products showed that relative to strain B128, strain 6.16 has a 441-bp (in-frame) deletion, including at its 3’ end a 33-bp sequence (underlined) identical to the 33-bp sequence immediately 5’ of the deletion. (C) Immunoblotting with α-CagY rabbit serum of cell lysates from H. pylori strains before and during the experimental challenge. Lane 1, B128 strain before challenge; lane 2, B128 isolated from mouse 6.11; lane 3, B128 isolated from mouse 6.16.
product from mouse 6.16 compared with the other two strains (Fig. 5 A), indicating that deletion of a fragment of the cagY MRR had occurred during colonization. Sequence analysis of strain 6.16 identified a 441-bp deletion (including a 33-bp direct repeat of the immediately upstream sequence) in relation to strain 6.11 (Fig. 5 B). Immunoblots using α-CagY serum indicated that strain 6.16 expressed smaller CagY protein products than those produced by strain B128 or 6.11 (Fig. 5 C), consistent with the sequence data. That for each of the three strains two major CagY protein products were identified, indicates that within a population of cells, the potential exists to produce CagY protein products that differ in size. These comparisons confirm that in vivo recombination between cagY direct repeats that result in deletions alters the composition of the in-frame–expressed protein.

Intrahost FRR Diversity Resulting from DNA Duplication. The presence of repetitive sequences allows for duplication as well as deletion (44). Examination of a rhesus monkey 10 mo after experimental challenge with seven H. pylori strains including strain J166 (33) identified an isolate, 98–147, from which a 1.8-kb FRR PCR product was amplified that was ~0.4 kb larger than the FRR product amplified from the prechallenge J166 strain (unpublished data). RAPD identified the two strains as being of the same clonal origin (33). Sequence analysis of the FRR from these two isolates showed that J166 contained three copies of a 393-bp tandem repeat that begins 23 bp downstream of the start codon, whereas strain 98–147 contained four copies of the repeat. The sequences were otherwise identical, indicating that the size variation resulted from duplication rather than acquisition of another allele through horizontal gene transfer. Immunoblots using the α-CagY serum indicated that strain 98–147 expressed CagY protein products larger in size than those expressed by strain J166 (not depicted). These in vivo results confirm that recombination between repetitive DNA allows H. pylori to undergo chromosomal deletion or duplication, which alters CagY amino acid composition.

H. pylori Strains Produce Varied CagY Proteins. To determine whether the observed size differences at the DNA level affect CagY protein expression and to assess interstrain and intrastain CagY size diversity, wild-type H. pylori strains were examined by immunoblot using α-CagY serum. Consistent with the deduced CagY size (19, 27), the α-CagY serum recognized proteins of ~180–220-kD in all cagY+ strains. Each strain had one to two major bands as well as multiple minor bands, varying in both size and band intensity (Fig. 6). That the antiserum did not recognize any proteins in cagY− control strain 88–22 (Fig. 6) or in a cagY deletion mutant (not depicted) indicates that the minor bands in the cagY+ strains are CagY specific and is consistent with H. pylori cultures containing heterogeneous cagY populations (Fig. 3, A and B). Differences in the intensity of CagY products identified in immunoblots may result from differential antibody recognition of particular products, variation in cagY transcription levels, or stability of the cagY RNA and protein products. To determine whether CagY protein size variation resulted from RecA-dependent cagY recombination, immunoblots were performed on a recA mutant (26695ΔrecA; Fig. 6). The presence of multiple CagY products in strain 26695ΔrecA indicates that CagY size differences do not require RecA-dependent events.

CagA Translocation and IL-8 Induction by Paired cagY Variant Strains. To examine the effect of cagY diversity on IV secretion system injection of the CagA substrate into host cells (20–22) and cytokine induction (10, 14), CagA translocation and IL-8 induction were examined in AGS cells cocultured with H. pylori strains 13abqs, 13bqs, J166, 98–147, and 88–22. For the AGS cells cocultured with each of the four H. pylori cag+ strains, immunoblots using α-CagA antibodies identified ~130 kD cell–associated products (not depicted) indicating CagA injection into the cells and thus, as expected, function of the type IV secretion system in each, and no products were identified in immunoblots of cells cocultured with cag− strain 88–22 or without coculture. These findings indicate that in-frame cagY deletions and duplications do not affect CagA translocation. For cagY MRR variant paired strains 13aqs and 13bqs (Fig. 4), coculture with AGS cells yielded significantly different IL-8 induction by 13aqs (4.7 ± 2.9 ng/AGS cell) and 13bqs (0.2 ± 0.2 ng/AGS cell). In contrast, there was no significant difference in IL-8 induction between cagY FRR variant paired strains J166 (3.7 ± 2.5 ng/AGS cell) and 98–147 (3.8 ± 3.1 ng/AGS cell). These results suggest that intragenomic deletions and duplications in the CagY MRR, but not the FRR, could affect IL-8 production by host cells, or that other, unidentified, genetic differences between paired strains 13aqs and 13bqs may influence IL-8 induction.

CagA Evades Humoral Immunity. To assess whether surface-exposed CagY domains (38) are recognized by the immune system, IgG responses to a recombinant CagY MRR fragment were examined in sera from persons carrying
cagY+ strains and from controls. First, because CagY proteins show strain-specific variation (Fig. 6), serum from the person from whom genome strain J99 was isolated (46) was examined. This index patient showed no IgG recognition of either the recombinant CagY fragment (Fig. 7 A), nor of the native CagY in lysates of J99 cells (Fig. 7 B), despite IgG recognition of multiple other H. pylori antigens (Fig. 6), serum from the person from whom strain J99 was isolated (46), and his robust responses to other strains (Fig. 7 B), the CagY protein evaded immune (IgG) recognition.

Next, we sought to determine whether this phenomenon was generalized among persons carrying cag-island+ H. pylori strains. First, to determine immune responses to H. pylori group antigens, we examined serum IgG responses in 195 patients to H. pylori whole cell lysates. As expected, the 114 H. pylori+ patients showed significantly greater recognition of H. pylori group antigens (mean ODU ± SD = 4.16 ± 2.36) than the 81 H. pylori- patients (0.37 ± 0.27). To ensure a test population capable of mounting strong immune responses to CagY, from among the 114 H. pylori+ patients, we selected 38 persons whose sera showed significantly (P < 0.0001) greater IgG recognition (6.03 ± 1.68) of the H. pylori antigens. Next, to confirm the antigenicity of the CagY preparation in the ELISA format, α-CagY rabbit immune serum was tested. As expected, this serum recognized the recombinant CagY MRR fragment to a substantially greater extent than the control rabbit serum (Fig. 7 C). These ELISA results are consistent with previous immunofluorescence experiments in which the α-CagY immune rabbit serum recognized the surface-exposed CagY region on intact H. pylori cells (38), indicating that the native CagY structure is intrinsically antigenic. However, CagY-specific serum IgG responses in humans were low, and not significantly different in persons colonized with cag+ or cag− strains (Fig. 7 C). As a control, and as expected, the 27 persons carrying cag+ strains had significantly (P < 0.001) higher IgG responses to a recombinant CagA antigen (18) than sera from 11 patients carrying cag− strains (Fig. 7 C), indicating that the CagA protein is antigenic and recognized during colonization. Of 27 sera examined from persons with cag+ strains, only five (18.5%) had α-CagY serum IgG responses >mean ± 3 SD for persons with cag− strains (Fig. 7 A). These five patients also had significantly higher IgG responses (7.1 ± 1.3 vs. 5.5 ± 1.5; P = 0.049) to H. pylori whole cell antigens than the 22 other cag+ patients, but showed no significant differences in CagA-specific IgG responses. In total, the results of these experiments show that among naturally colonized human hosts who mount strong immune responses to both H. pylori and CagA anti-
gens, antibody recognition of the surface-exposed CagY protein and its MRR is minimal or absent despite years or decades of CagY expression by *H. pylori*.

Discussion

The ability of *H. pylori* to persist within hosts that recognize its presence (10, 47) indicates important adaptations, undoubtedly involving programmed cross-talk between microbe and host (48). However, that common mechanisms for adapting to changing environments, including stringent response and two-component regulatory systems, are underrepresented in *H. pylori* (49), suggests the existence of alternate mechanisms for gene regulation. Recombination involving direct DNA repeat sequences, for example, flanking the *hpyl*II restriction modification system (45), is one mechanism that affects gene content and thus phenotype. That a large number of direct repeat sequences are present in *H. pylori* cagY in a highly skewed distribution and that rearrangements involving these repeats invariably yield in-frame products is a paradigm for use of repetitive DNA to alter amino acid content of an expressed protein (50). The RecA independence of cagY rearrangement is consistent with the recent observation that intragenomic recombination in *E. coli* between direct repeats <400 bp does not require RecA function (30). Although the data presented in this study indicate the role of intragenomic (DNA) recombination in influencing CagY sequence diversity, they do not exclude roles at the transcriptional or translational levels.

*H. pylori* populations within individual hosts are diverse (31, 32, 46) with recombination between repeats leading to varied genotypes (45, 51). The presence of multiple cagY alleles in a single host indicates generation of phenotypic variation (Figs. 4 and 6), which provides substrate for host selection based on differential fitness (52). Analysis of CagY protein expression indicates that *H. pylori* cell populations that contain dominant cagY alleles also include multiple variants. Comparison of 84–183LP and in vitro–passaged 84–183HP CagY products suggests that development of cagY allele variants is a spontaneous event that leads to heterogeneous populations. Within a host, the existence of heterogeneous CagY populations provides substrate for selection of the allele with greatest fitness for that in vivo context. The natural competence of *H. pylori* (53, 54), which provides opportunities for both interstrain (55, 56) and intrastrain (45) recombination between cagY alleles, yields an even larger repertoire of phenotypes that are substrate for host selection. However, although there is potential for enormous size variation of the MRR due to deletion or duplication, finding that PCR products from all 107 examined strains range between 2 and 3 kb indicates strong selection for that size range. These findings are consistent with constraints involved in formation of pilus that function in type IV systems (25, 26), in which the CagY surface–exposed regions must extend beyond the bacterial envelope and yet not interfere with translocation of the substrate macromolecule(s). The presence of deletions and duplications, and intergenomic recombination as well (37, 43, 53), may generate an extremely large number of cagY alleles. *H. pylori* cells with alleles that are too long or short would be selected against.

Antigenic variation is well recognized in bacterial pili, structures highly subject to host immune surveillance (57, 58). By homology with other genes encoding VirB10-like sequences, cagY is a chimera, in which the 3¢ VHR has classical type IV secretion system function (12, 19–22), and the unique upstream sequences have alternative function. Variations in the 5¢ G+C content suggests that the current form of cagY developed after *H. pylori* had become a distinct species. Confocal laser scanning microscopy and electron microscopy indicate that CagY sheaths the needle-like structure of the type IV secretion system (38) involved in CagA injection into host cells (21), thereby making it susceptible to host immune attack. Despite the presence of the needle-like type IV secretion system at one bacterial pole, CagY was identified on the surface of the entire bacterial cell, suggesting that its role is not limited to the type IV secretion system (38). The extraordinary number of DNA repeats within the cagY FRR and MRR facilitates sequence variation through intragenomic recombination at a frequency markedly higher than the expected rate of DNA mutation or alteration in *H. pylori* housekeeping genes (38, 55), providing a mechanism for rapid changes in amino acid composition, and thus epitopes, of CagY surface–exposed regions. Our findings indicating both potential and actual extensive strain–specific protein variation, as well as the observed absent or minimal host responses to the cagY product, provide evidence that FRR– and MRR–based antigenic variation reduces CagY vulnerability to host immune responses. Alternatively, the repetitive DNA sequences within CagY may facilitate binding of the type IV secretion apparatus to cellular receptors, provide mechanical support, or protect the apparatus from the surrounding acidic environment.

Bacterial type III or IV secretion systems that inject their substrates into host cells typically include components exposed to host immunity (25, 26). The decades-long, naturally occurring *H. pylori* colonization of its host (59), with ongoing immune recognition (16), provides particularly strong selection for antigenic variation of such a critical constituent. The experimental data from mice, rhesus monkeys, and humans indicate that over relatively short intervals (weeks to months), *H. pylori* cells with cagY variants created by intragenomic recombination replaced previous populations. The high mutation rates seen in *H. pylori* (60) may generate heterogeneous CagY populations to evade host immune responses. This hypothesis is consistent with immunoblots of CagY antigenic expression (Fig. 6), which indicate that *H. pylori* cell populations vary in several CagY isoforms and in their host recognition. The ability of the rabbit to mount an immune response to the recombinant CagY protein, whereas the natural host mounts minimal or no antibody responses to native CagY, might be because the rabbit is hyperimmunized with a high concentration of...
the CagY recombinant protein as a single protein, delivered parenterally, as opposed to crossing the mucosa.

The general lack of human antibody response to CagY, despite colonization that has been ongoing for years to decades, suggests that cagY variation could be occurring so rapidly that the host immune system is not capable of mounting an effective response. Alternatively, lack of detectable host antibodies may reflect the power of even low levels of antibodies to act as selecting agents. H. pylori adaptation to this phenomenon would also involve generation of variants, but in this model, host responses to particular epitopes are lost over time and most would not be detectable on an ongoing basis. During the long course of colonization, hosts develop robust immune responses to numerous H. pylori proteins. However, because these antigens are usually internal to either the bacterial (e.g., urease) or host (e.g., CagA) cells, their immune recognition should not facilitate H. pylori clearance. Other H. pylori surface constituents, such as BabA and BabB, also have diversity regions (61), and the outermost lipopolysaccharide constituents, the Lewis antigens present on host cells (62) may also be involved in immune evasion. That several other putative H. pylori surface proteins also contain repetitive DNA sequences (unpublished data), suggests that CagY immune evasion might be a model for a more general phenomenon in H. pylori. The selection inferred from observing a highly evolved system of immune evasion in H. pylori implies that functional host immunity exists against organisms living in the lumen of the human stomach. That Plasmodium falciparum, an important cause of human malaria, contains Csp, a surface-exposed protein containing direct repeat sequences that undergo high frequency (10^-3–10^-2) intragenomic recombination (63), suggests that the observed CagY deletions and duplications might be representative of immune evasion mechanisms that span a broad range of microbial species.

In conclusion, our findings demonstrate that H. pylori uses extensive DNA rearrangement to alter amino acid composition of the surface-exposed region of a required constituent of its type IV secretion system, which effectually evades the host immune response. That several other bacterial species with relatively small (<2 Mb) genomes also possess extensive and nonrandomly distributed DNA repeats (64), suggests that recombination between repetitive DNA sequences (50, 65) might be a common mechanism in such organisms to regulate gene content. The localization of the cagY repeats indicates the evolution of programmed rearrangements, whereas the remarkably extensive possibilities suggest a stochastic process. This combination is a powerful mechanism to effect variation of critical host-exposed structures.

This investigation was supported by National Service Awards 5 T32 AI07180-20 and 5 T32 CA09385, and R01 GM63270 from the National Institutes of Health, the Medical Research Service of the Department of Veterans Affairs, and by SFB576, project B1.

References


