Published in final edited form as:

Helicobacter. 2000 March ; 5(1): 1–12.

Analysis of *iceA1* transcription in *Helicobacter pylori*

John P. Donahue^{*}, Richard M. Peek Jr.^{†,‡}, Leen-Jan van Doorn[§], Stuart A. Thompson^{*}, Qing Xu[¶], Martin J. Blaser^{*,‡,¶}, and Geraldine G. Miller^{*,¶}

^{*}Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, TN [†]Division of Gastroenterology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN [‡]Department of Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN [§]Medical Service, Department of Veterans Affairs Medical Center, Nashville, TN [¶]Delft Diagnostic Laboratory, Delft, The Netherlands

Abstract

Background—Transcription of the *Helicobacter pylori* *iceA1* gene is induced following adherence of the bacterium to gastric epithelial cells *in vitro*, suggesting that this gene might be involved in *H. pylori* pathogenesis. Consequently, the current studies were undertaken to characterize *iceA1* transcription and to define the structure of *iceA1*-containing transcripts to evaluate the potential of this gene to encode functional proteins.

Materials and Methods—Northern blots and primer extension of RNA isolated from broth-grown cultures of various *H. pylori* strains was done to analyze *iceA1*-specific gene transcription. Reverse transcriptase (RT)-PCR was used to determine the levels of *iceA1* transcripts derived from readthrough transcription that was initiated upstream of *iceA1* within the 5'-flanking *cysE* gene.

Results—Three major transcripts were detected and each was initiated from a common promoter, designated P_I. Two of these transcripts were comprised of *iceA1* sequence, while a third transcript was dicistronic and included the downstream gene, *hpyIM*. In addition, 10-fold lower levels of *iceA1* transcripts were initiated upstream of P_I, either within or immediately downstream of *cysE*.

Conclusions—The present analysis suggests that *iceA1* does not encode a functional protein in the majority of *H. pylori* strains. However, transcription of *hpyIM*, which encodes a highly conserved DNA adenine methyltransferase, is linked to *iceA1* transcription. Therefore, *iceA1* may affect *H. pylori* virulence *in vivo* through transcriptional regulation of *hpyIM* expression levels, which may result in specific variations in DNA methylation patterns leading to alteration in the expression of genes involved in virulence or pathogenesis.

Helicobacter pylori is a Gram-negative, microaerophilic, curved bacterium that persistently colonizes the mucous layer overlying the gastric epithelium [1]. Colonization with *H. pylori* results in chronic superficial gastritis [2], which increases the risk for the development of duodenal and gastric ulcers, gastric adenocarcinoma, or non-Hodgkin's gastric lymphoma [3–5]. However, the majority of persons infected with *H. pylori* remain asymptomatic [2] and, although several strain-specific factors have been identified that are potentially markers for the differential clinical outcome of *H. pylori* colonization [6–9], we presently have an incomplete understanding of the bacterial factors involved in progression to ulceration or more severe disease.

The *iceA* genetic locus was identified using an experimental strategy based on the hypothesis that adherence of *H. pylori* to gastric epithelial cells may induce the expression of genes related to virulence or pathogenesis [10]. DNA sequencing of *iceA* from *H. pylori* clinical isolates demonstrates that this gene exists as two distinct allelic types, designated *iceA1* and *iceA2* [10,11]. Epidemiological evidence suggested that strains containing the *iceA1* allele are associated with duodenal ulcer disease [10,12], although this finding was not supported in another study [13].

The *iceA1* allele shows significant homology (60% nucleotide identity) to *nlaIII*R, which encodes the *Neisseria lactamica* restriction endonuclease *NlaIII* that recognizes the sequence CATG [14]. Located immediately downstream of *iceA* is *hpyIM*, which encodes a DNA adenine methyltransferase that also specifically recognizes the sequence CATG [15]. Consequently, the genetic arrangement of *iceA1-hpyIM* resembles a typical type II restriction-modification system found in many bacterial species [16]. However, the majority of *iceA1* sequences studied thus far contain different, strain-specific frameshift and nonsense mutations within its potential ORF that would prevent translation of a full length protein with homology to *NlaIII* [10,11]. The *iceA2* allele shows approximately 40% homology to *iceA1* but has an entirely different genetic structure [10,11]. Thus, these observations suggest that it is unlikely that *iceA-hpyIM* constitutes a functional type II restriction-modification system in the majority of *H. pylori* strains. However, evidence for potential *iceA1* function is suggested by the demonstration that *iceA1* transcription is induced following *H. pylori* contact with epithelial cells *in vitro*. Consequently, we sought to characterize *iceA1* transcription and determine the structure of *iceA1*-containing transcripts to evaluate the potential of this gene to encode functional proteins.

Materials and Methods

Bacterial Strains, Plasmids and Oligonucleotides

Strains used in this study were *H. pylori* 60190 [*vacA*⁺ (s1a/m1) *cagA*⁺ *iceA1*⁺], which is a well characterized laboratory strain [17]; 60190I⁻ (*iceA1::aphA*); and J178 [*vacA*⁺ (s1a/m1) *cagA*⁺ *iceA2*⁺], a clinical isolate from a USA patient with gastric inflammation [10]. To construct *H. pylori* 60190I⁻ an *aphA* cassette, encoding kanamycin resistance, derived from pILL600 [18] was inserted into the unique *EagI* site (encompassing nucleotide residues 564-569 within ORF-2 in Fig. 1) in *iceA1* to create plasmid pVU1005. The *iceA1* (*orf2*):*aphA* fusion was introduced into the chromosome of strain 60190 by natural transformation using 5 µg of pVU1005, as previously described [15]. Transformants were selected on blood agar plates containing 30 µg/ml kanamycin. Genomic DNA was isolated from a kanamycin resistant transformant and the presence of the *aphA* cassette in *iceA1* was confirmed by polymerase chain reaction (PCR) analysis using oligonucleotide primers F1 and R10 (Table 1; data not shown). Oligonucleotides used in this study (Table 1) were synthesized by the DNA core facility of the Vanderbilt University Cancer Center.

Preparation of Cellular RNA

Brucella broth (Difco Laboratories, Detroit MI) containing 10% fetal calf serum was inoculated with *H. pylori* strains from 24 hours blood agar plate cultures of the organisms. Broth cultures were incubated at 37°C in 5% CO₂ with shaking at 100 rpm. Cellular RNA was isolated from cultures in the early exponential ($A_{600} = 0.4-0.5$) phase of growth by extraction with TRI REAGENT (Molecular Research Center, Inc., Cincinnati OH) as described by the manufacturer. Contaminating DNA was removed from RNA preparations by treatment with RNase-free DNase I (1 unit/µg of RNA) (Life Technologies, Gaithersburg, MD).

Northern Blotting of Cellular RNA

RNA (10 µg) was separated by electrophoresis in a 1.2% formaldehyde agarose gel and then transferred to a Hybond N (Amersham Life Science, Arlington Heights IL) membrane, as described [19]. RNA was crosslinked to the membrane by exposure to ultraviolet light (1200 microjoules for 30 sec) in a Stratalinker 2400 (Stratagene, La Jolla, CA). Filters were prehybridized in 2 × standard sodium chloride-sodium phosphate-EDTA (SSPE), 5 × Denhardt's solution, 0.1% sodium dodecyl sulfate (SDS) and 150 µg/ml salmon sperm DNA (Sigma, St. Louis, MO) at 65°C for 4 hours. After prehybridization, filters were hybridized to ³²P-labeled DNA fragments, as indicated in individual figure legends, in 2 × SSPE, 2 × Denhardt's solution, 50% dextran sulfate, 0.1% SDS and 150 µg/ml salmon sperm DNA at 65°C for 18–20 hours. After hybridization, filters were washed twice in 0.1 × SSPE, 0.1% SDS at 65°C for 30 min and then autoradiographed using Kodak Bio-Max X-ray film. The 0.24–9.5 kb RNA ladder used to estimate transcript sizes was from Life Technologies.

Preparation of DNA Probes for Northern Blots

DNA fragments P-1 through P-7 (Fig. 2A and Fig. 3A) were produced by PCR using *iceA1*- and *hpyIM*-specific oligonucleotide primers and *H. pylori* genomic DNA. A 522-bp DNA fragment encoding a segment of 16S rRNA was produced by PCR using primers HP1 and HP2 and *H. pylori* genomic DNA [20]. After PCR, DNA fragments were isolated by agarose gel electrophoresis and extracted and purified from gel slices using the QiaexII kit (Qiagen, Inc., Chatsworth CA), as described by the manufacturer. DNA fragments were labeled with [α -³²] dCTP using the Prime-It RmT Random Primer labeling kit (Stratagene), as recommended by the manufacturer. Radiolabeled DNA fragments were separated from unincorporated [α -³²] dCTP by centrifugation through small Sephadex G-50 columns.

Primer Extension Mapping of *iceA1* Transcripts

RNA (10 µg) from *H. pylori* strains 60190 or J178 was mixed with 1 µl 5'-³²P-labeled oligonucleotide primers R3 or R7 (2 pmol/µl) and 4 µl 5 × First Strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂] in a total volume of 16 µl. Mixtures were heated at 100°C for 3 min, incubated at 60°C for 1 hour, placed on ice and then 2 µl 0.1 M DTT and 1 µl 10 mM dNTP mixture (10 mM each dATP, dGTP, dTTP, dCTP) added to each. Reactions were prewarmed at 42°C for 2 min and then incubated at 42°C for 1 hour with 1 µl Superscript II reverse transcriptase (Life Technologies), and then terminated by incubation at 70°C for 15 min. Each primer extension reaction (6 µl) was combined with 4 µl of DNA sequencing gel dye (95% formamide, 20 mM Na₂ ethylenediaminetetraacetate (EDTA), 0.05% bromophenol blue and xylene cyanol FF). Samples were heated at 80°C for 3 min, chilled on ice, and then 4 µl was loaded alongside an appropriate DNA sequencing ladder on an 8% polyacrylamide sequencing gel containing 8 M urea.

Reverse Transcriptase (RT)-PCR of *iceA1* Transcripts

RNA (5 µg) from *H. pylori* strain 60190 was mixed with 2 pmol of the oligonucleotide primer R10 (Table 1) in a total volume of 12 µl, heated at 70°C for 10 min and quick chilled on ice. Then, 4 µl 5 × First Strand buffer, 2 µl 0.1 M DTT, and 1 µl 10 mM dNTP mixture was added. The reaction was prewarmed at 42°C for 2 min and then incubated at 42°C for 1 hour with 1 µl of Superscript II reverse transcriptase. Reactions were terminated by incubation at 70°C for 15 min. PCR was performed as described [21] using 1 µl of the first strand cDNA reaction as template and oligonucleotide primers F1 and R3 (Table 1). Reaction conditions were 94°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute for 30 cycles followed by a final extension at 72°C for 10 minutes. For quantitative comparison of transcript levels by RT-PCR, RNA from *H. pylori* strain 60190 was reverse transcribed as described above using the oligonucleotide primers, R5 or R9 (Table 1). for PCR, 10-fold serial dilutions of the

oligonucleotide R5-primed RT reaction were amplified using oligonucleotide primers F2 and R4, or 10-fold serial dilutions of the oligonucleotide R9-primed RT reaction were amplified using oligonucleotide primers F6 and R8, using the reaction conditions described above.

Results

Demonstration of Three *iceA1* Transcripts

The *iceA1* gene in *H. pylori* strain 60190 has 60% nucleotide identity to *nlaIII*, which encodes the restriction endonuclease *NlaIII* in *Neisseria lactamica*. The *iceA1* sequences homologous to *nlaIII* begin with the ATG at nucleotide residues 206–208 and end at nucleotide residue 888 (Fig. 1) and define a potential ORF. However, a single nucleotide deletion between residues 331 and 332 results in a frameshift and premature translation termination, so that the longest potential protein that could be synthesized is 64 amino acids in length (ORF-1 in Fig. 1). In the absence of this deletion and frameshift, a 228 amino acid protein with 66% homology (identity + similarity) to *NlaIII* could be translated. A second potential ORF (ORF-2 in Fig. 1) begins with the ATG at nucleotide residues 508–510 and could encode a 127 amino acid protein with 66% homology (identity + similarity) to *NlaIII*, but lacking substantial sequence at the amino terminus. If frameshift suppression exists in *H. pylori*, transcripts initiating upstream of nucleotide 206 might be translated into an active *NlaIII* homolog. Also, we have found that *iceA1* sequences in four of 19 strains lack frame-shift mutations and contain an intact ORF comprised of both ORF-1 and ORF-2 in Fig. 1 (L. J. van Doorn, manuscript submitted). Therefore, we sought to characterize *iceA1* transcription to determine whether transcripts were synthesized that could encode the 228 amino acid protein. Northern blot analysis was performed on RNA isolated from the wild-type strain 60190 and an isogenic mutant, 60190I⁻, which was produced by insertion of an *aphA* cassette into the *EagI* site of *iceA1* (*orf2*) (Fig. 1 and 2A). This insertion mutation is predicted to interrupt transcripts initiating upstream of ORF-2.

Using the *iceA1*-specific probe P-1 (Fig. 2A), three transcripts were demonstrated in the wild type strain 60190; two major transcripts, approximately 530 and 330 nucleotides in length, and a third less prominent transcript, approximately 1460 nucleotides in length (Fig. 2B). These transcripts were not detected in the mutant strain 60190I⁻. To determine whether the 1460 nucleotide transcript included sequences downstream of the 684 bp *iceA1* region, we used probe P-2 which hybridizes to *hpyIM* (Fig. 2A). Probe P-2 detected the 1460 but not the 530 and 330 nucleotide transcripts in strain 60190 and no transcripts were detected in strain 60190I⁻. Control hybridization with a probe for *H. pylori* 16S rRNA indicated that the levels of RNA loaded in each lane and transferred were similar. These results suggested that the 530 and 330 nucleotide transcripts lie within *iceA1*, because they are not detected by probe P-2, and that *hpyIM* is cotranscribed with *iceA1* to yield a 1460 nucleotide dicistronic transcript which is of sufficient length to contain the entire protein coding sequence of *hpyIM*. The absence of the 1460 nucleotide *hpyIM*-specific transcript in strain 60190I⁻ is consistent with the dicistronic nature of this transcript, because insertion of the *aphA* cassette in *iceA1* is expected to exert a polar effect on transcription of downstream genes in a polycistronic mRNA. In addition, transcription of *hpyIIM*, which is located immediately downstream of *hpyIM*, is unaffected in strain 60190I⁻ demonstrating that transcription of this gene is independent of *iceA1-hpyIM* transcription (data not shown).

To evaluate the possibility of strain-dependent differences in *iceA1* transcription, we also did Northern blot analysis of RNA isolated from *iceA1* strains J166 [10], 26695 [22] and CH4. This analysis demonstrated the presence, in these strains, of the *iceA1* and *iceA1-hpyIM* transcripts observed in strain 60190, although strain-dependent differences in the levels of each of these transcripts were observed (data not shown).

Localization of the 5'-termini of the 530, 330 and 1460 Nucleotide Transcripts Within *iceA*

We next sought to determine the approximate origin of the 530, 330 and 1460 nucleotide transcripts within *iceA1* by Northern blotting of RNA isolated from strain 60190. The probes used for this analysis spanned defined segments of *iceA1* and are shown schematically in Fig. 3A. Neither probe P-3 nor P-4, which encompass the region from nucleotide residues 149–419 (Fig. 1), detected any of the transcripts, indicating that the 5'-termini of all three transcripts are further downstream. Probe P-5, which spans the insertion site of the *aphA* cassette at nucleotide residue 564 in 60190I⁻ (Fig. 1), and probe P-6, which lies downstream of this insertion site, detected all three transcripts. These results suggested that all three transcripts originate between nucleotide residues 419 and 564. Only probe P-7 revealed differential hybridization to the three transcripts. It detected the 1460 and 530 nucleotide transcripts, but did not detect the 330 nucleotide transcript. These findings suggested that the 3'-end of the 330 nucleotide transcript is located upstream of fragment P-7 and within fragment P-6. Control hybridization with a probe for 16S rRNA confirmed similar loading and transfer of RNA in these blots. Together with the results from the *aphA* insertion (Fig. 2B), these data indicated that all three transcripts have a common 5' terminus that originates within fragment P-5 and that transcriptional termination varies resulting in the 530 and 330 nucleotide *iceA1* transcripts and a discistronic transcript which includes *iceA1* and *hpyIM*.

Determination of the Site of *iceA1* Transcriptional Initiation

The site of transcriptional initiation within the *iceA1* region defined by fragment P-5 in Northern blots was identified by primer extension mapping of the 5'-termini of *iceA1* transcripts (Fig. 4A). The results of this analysis showed that transcriptional initiation occurs primarily at nucleotide residues 444 and 446 in strain 60190. As expected, primer extension products were not detected in the negative control *iceA2* strain, J178. The identified *iceA1* transcript 5'-termini in strain 60190 are preceded by the sequence TATAAT (433–438 in Fig. 1) which is identical to the consensus – 10 sequence of *Escherichia coli* σ^{70} promoters [23]. The sequence TTGATA (410–415 in Fig. 1) located 17 nucleotides upstream of the – 10 sequence differs by only a single nucleotide from the *E. coli* σ^{70} –35 consensus sequence [23]. These results define the location of the *iceA1* promoter, designated P_I, within fragment P-5 (Fig. 3A) and establish a common site of transcriptional initiation for the three major *iceA1* transcripts identified by Northern blot analysis. This transcriptional initiation site is located 64 nucleotide residues upstream of potential ORF-2 (Fig. 1). Based on the transcript sizes determined in Northern blots, the 530 and 1460 nucleotide transcripts, but not the 330 nucleotide transcript, would be of sufficient length to encode the complete polypeptide predicted by this ORF. In strain 60190, this polypeptide would be 127 amino acid residues in length with 66% amino acid homology (identity + similarity) to the C-terminal end of *NlaIII*.

Demonstration of an *iceA1* transcript 5'-Terminus in the *cysE-iceA1* Intergenic Region

Because *iceA1* shows 60% nucleotide identity to *N. lactamica nlaIIIR* beginning at nucleotide residue 206 and ending at nucleotide residue 888 (Fig. 1), we initially predicted that *iceA1* transcriptional initiation might occur in the region upstream of the ATG triplet that is homologous to the *NlaIII* translational initiation codon (206–208 in Fig. 1). Transcripts initiated upstream of this ATG could encode a full-length *NlaIII* homolog, if frameshift correcting mechanisms exist in *H. pylori*, or in a minority of strains that do not contain frameshift mutations in the *iceA1* ORF. However, the experiments described above showed that the major *iceA1* transcripts detected by Northern blot analysis initiated downstream of the ATG at nucleotide residues 206–208 and beyond the predicted ORF-1 (Fig. 1). Nevertheless, using primer extension analysis, we sought to identify transcripts originating in the region upstream of this ATG. The results of this analysis demonstrated a unique 5'-terminus located at nucleotide residue 150 (Fig. 4B). Unlike the *iceA1* promoter P_I identified above, this

transcriptional initiation site is not preceded by -10 and -35 sequences that resemble those found in *E. coli* σ^{70} promoters [23]. The complete genome sequences of *H. pylori* strains 26695 and J99 [22,24] revealed that this organism contains two alternative sigma factor homologs, σ^{54} and σ^{28} , in addition to σ^{70} . However, the sequences located immediately up stream of the mapped transcript 5'-terminus do not match the consensus sequences shown to be recognized by these σ factors in other bacterial species. Consequently, it is not clear whether the 5'-terminus of this transcript results from transcriptional initiation, or from the processing of transcripts initiated upstream of *iceA1*, perhaps at the *cysE* promoter (see below). In either case, the mapped 5'-terminus of this transcript is located 56 nucleotide residues upstream of potential ORF-1 that could encode a 64 amino acid polypeptide (Fig. 1).

Demonstration of Readthrough Transcription from *cysE* into *iceA1*

The intergenic region between the *cysE* translation termination codon (119–121 in Fig. 1) and the *iceA1* ATG triplet which is homologous to the *NlaIII* initiation codon (206–208 in Fig. 1) is only 84 nucleotide residues and does not contain obvious signals for the termination of transcripts initiated at the *cysE* promoter. Therefore, transcripts initiated at the *cysE* promoter could continue into *iceA1*, and processing of this long transcript could account for the *iceA1* 5'-terminus at nucleotide residue 150 that was detected in this region by primer extension (Fig. 4B). To address this issue, we used RT-PCR to determine whether readthrough transcription from *cysE* into *iceA1* occurs. RNA from strain 60190 was reverse transcribed using the R10 oligonucleotide primer (Fig. 5A and Table 1) which anneals to sequences near the 5'-terminus of *hpyIM*. When PCR amplification of the first strand cDNA reaction was done using oligonucleotide primers F1 and R3 (Table 1), a product of the predicted size (287 bp) was detected, indicating that readthrough transcription into *iceA1* from *cysE* occurs (Fig. 5B, lane 2). PCR amplification of RNA that had not been reverse transcribed showed no product (Fig. 5B, lane 3), confirming that the 287 bp product was the result of reverse transcription of RNA into cDNA, and not from DNA contamination. Because the primer used for reverse transcription of the RNA into cDNA anneals to sequences downstream of *iceA1* within *hpyIM*, the results of this analysis also demonstrated that these readthrough transcripts can contain the entire *iceA1* sequence and extend into the 5'-end of *hpyIM*. Thus, these data indicate that the 5'-terminus of the *iceA1* transcript mapped by primer extension at nucleotide residue T₁₅₀ (Fig. 4B) could result from processing of transcripts initiated upstream of *iceA1* in *cysE*, and therefore, might not indicate that a second *iceA1* promoter is located in this region.

Quantitation of *iceA1* Promoter P_I-initiated Transcripts and Transcripts Initiated Upstream of P_I

Northern blot analysis of RNA isolated from strain 60190 did not show the presence of *iceA1* transcripts corresponding to the region upstream of nucleotide residue 419 as defined by probe P-4 (Fig. 3A and 3B). However, both primer extension and RT-PCR analysis showed that transcripts encoding sequences from this region were present in strain 60190 (Figs 4B and 5B). We sought to reconcile this discrepancy using a quantitative RT-PCR method, which allowed us to compare the levels of *iceA1* promoter P_I-initiated transcripts to the levels of transcripts initiated upstream of nucleotide residue 419. RNA isolated from strain 60190 was reverse transcribed using the R5 oligonucleotide primer, which anneals to nucleotide residues 419–396 located upstream of the *iceA1* P_I promoter, or the R9 oligonucleotide primer, which anneals to nucleotide residues 803–781 located within the 530 and 1460 nucleotide transcripts. These primers were chosen for reverse transcription of RNA so that cDNA generated from transcripts initiated either at P_I or upstream of P_I were approximately the same length, thereby minimizing bias in the estimation of transcript levels that might be caused by premature termination of transcript copying by reverse transcriptase. The relative locations of these primers in *iceA1*, and the DNA fragments that were amplified by PCR, are shown schematically in Fig. 5A. Ten-fold serial dilutions of the oligonucleotide R5-primed RT reaction were

amplified using oligonucleotide primers F2 and R4, which allows detection of cDNA that was extended from within transcripts initiated upstream of the P_I promoter. Similarly, 10-fold serial dilutions of the oligonucleotide R9-primed RT reaction were amplified using primers F6 and R8, which allows detection of cDNA that was extended from within the transcripts initiated at P_I. As shown in Fig. 5C, for the F2/R4 primer set, a 212-bp product was detected to the 10⁻³ dilution of the RT reaction. For the F6/R8 primer set, a 193-bp product was detected to the 10⁻⁴ dilution of the RT reaction. Comparison of band intensities by densitometry indicated that transcript levels resulting from transcriptional initiation from a putative upstream *iceA1* promoter or readthrough transcription from *cysE* were 10-fold lower than transcription initiated from the P_I promoter. We determined that this result was not due to the differential ability of the F2/R4 and F6/R8 primer sets to amplify their target sequences, because both primer sets exhibited identical amplification efficiencies, with resolution to 10 pg of 60190 genomic DNA (data not shown). Thus, these results confirm that the major transcripts from the *iceA1* locus are initiated from a promoter located downstream of the ATG that is homologous to the *NlaIII* translation initiation codon (nucleotide residues 206–208 in Fig. 1) and lack sequences needed to encode either ORF-1 or a full-length homolog of *NlaIII*. However, 10-fold lower levels of long transcripts that include these sequences are expressed.

Discussion

The *H. pylori iceA* genetic locus was identified because transcription of the *iceA1* allelic variant of this gene was induced following adherence of the organism to gastric epithelial cells in vitro [10]. This observation suggested to us that *iceA1* might be a virulence-associated gene, because epithelial cell contact induces the expression of genes associated with virulence in other gastrointestinal pathogens [25]. This hypothesis was subsequently supported by epidemiological evidence showing an association of *iceA1* strains with enhanced gastric inflammation and duodenal ulcer disease [10,12]. However, recently published epidemiological data do not support the association of *iceA1* with clinical outcome, but suggest that *iceA* (*iceA1* vs. *iceA2*) status reflects geographic differences in *H. pylori* strains rather than a virulence-associated trait [13]. Further understanding of the function of the *iceA* locus and expression of its products may help resolve these discordant results.

The function of *iceA1*, whether it encodes a protein product and whether it plays a role in *H. pylori* pathogenesis, is unknown. Therefore, the goal of these studies was to define the transcriptional products of the *iceA1* locus to establish its potential for encoding functional proteins. The strong nucleotide homology between *iceA1* and *nlaIII*R, which encodes the type II restriction endonuclease *NlaIII*, together with the tandem arrangement of *iceA1* and *hpyIM*, suggested that *iceA1* might encode a functional endonuclease with *NlaIII*-like recognition specificity for CATG sequences which are modified by *M.HpyI*. However, the frameshift mutation in the *iceA1* sequence of strain 60190 is but one example of multiple types of mutations, including longer deletions, insertions, and point mutations, found in approximately 80% of *iceA1* sequences examined thus far [10,11]. Therefore, multiple compensatory mutation or suppression mechanisms must be invoked to postulate translation of a full-length protein homologous to *NlaIII* in the majority of *iceA1* strains. In addition to this, our results show that the three major transcripts originating from the P_I promoter lack either the N-terminal sequence (530 and 1460 nucleotide transcripts) or both N- and C-terminal sequence (330 nucleotide transcript) needed to encode a full-length homolog of *NlaIII*. However, the 530 and 1460 nucleotide transcripts each could encode a 127 amino acid polypeptide with 66% homology (identity + similarity) to the C-terminus of *NlaIII* (ORF-2 in Fig. 1). It is presently unknown whether this protein, if synthesized, would have restriction endonuclease activity or some alternative function.

In addition to the three major transcripts originating from P_I , transcripts that include sequences homologous to the N-terminus of *NlaIII* and potentially capable of encoding both short and long homologous proteins (ORF-1 and ORF-2, respectively, in Fig. 1) are expressed in *H. pylori* strain 60190, but at approximately 10-fold lower levels than the major P_I -initiated transcripts. It is unknown whether mechanisms exist in *H. Pylori* to correct the frameshift mutation in strain 60190 *iceA1* and allow translation of a full-length homolog of *NlaIII*. Efforts to identify restriction enzyme activity with *NlaIII*-like specificity in partially purified extracts of *H. pylori* strains 60190 and J166, which contain frameshift mutations, have been unsuccessful, while preliminary studies on extracts from strain CH4 which lacks such mutations demonstrated *NlaIII*-like activity (Q. Xu, unpublished data). Considering the high levels of synthesis of transcripts initiated from P_I , these results suggest that ORF-2 (Fig. 1) does not encode a functional *NlaIII*-like endonuclease, although other functions for this putative protein can not be ruled out. If frameshift suppression occurs with the low levels of *cysE*-initiated transcripts, it does not lead to expression of sufficient *NlaIII*-like restriction activity for detection. Thus we suspect that only those strains lacking frameshift mutations express significant levels of this endonuclease.

Finally, the structures of the three major transcripts initiated at the P_I promoter share a common 5' origin and differ at their 3' termini. This finding suggests that the shorter 530 and 330 nucleotide transcripts terminating before *hpyIM* must result either from premature transcriptional termination or from processing or degradation of the 1460 nucleotide transcript that includes both *iceA1* and *hpyIM* sequences. The presence of RNA secondary structure likely to be involved in transcriptional termination and transcript processing mechanisms are not evident in the defined the 3'-ends of the 530 and 330 nucleotide transcripts. However, regardless of the mechanism by which the smaller transcripts are generated, the net result is a reduction in the levels of the dicistronic message initiated at P_I , which have regulatory significance for *hpyIM* expression. The ability of *H. pylori* to regulate *hpyIM* transcription and thereby modulate DNA adenine methylase levels in the cell may be significant parallel to the recent demonstration of Dam methylase involvement in *Salmonella* pathogenesis [26].

Acknowledgments

We thank Mark Forsyth and Dawn Israel for valuable discussions and suggestions, and also D. Israel for technical advice on the preparation of figures. This work was supported by National Institutes of Health grant RO1-AI37659, by National Cancer Institute Center grant CA68485 awarded to the Vanderbilt Cancer Center, and by the Medical Research Service of the Department of Veterans Affairs.

References

1. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1984;1:1311–5. [PubMed: 6145023]
2. Dooley CP, Fitzgibbons PL, Cohen H, Appleman MD, Pérez-Pérez GI, Blaser MJ. Prevalence of *Helicobacter pylori* infection and histologic gastritis in asymptomatic persons. *N Engl J Med* 1989;321:1562–6. [PubMed: 2586553]
3. Peterson WL. *Helicobacter pylori* and peptic ulcer disease. *N Engl J Med* 1991;324:1043–8. [PubMed: 2005942]
4. Nomura A, Stemmermann GN, Chyou P-H, Kato I, Pérez-Pérez GI, Blaser MJ. *Helicobacter pylori* infection and gastric carcinoma in a population of Japanese-Americans in Hawaii. *N Engl J Med* 1991;325:1132–6. [PubMed: 1891021]
5. Parsonnet J, Hansen S, Rodriguez L, Gelb AB, Warnke RA, Jellum E, Orentreich N, Vogelstein JH, Friedman GD. *Helicobacter pylori* infection and gastric lymphoma. *N Engl J Med* 1994;330:1267–71. [PubMed: 8145781]

6. Atherton JC, Peek RM, Tham KT, Cover TL, Blaser MJ. The clinical and pathological importance of heterogeneity in *vacA*, encoding the vacuolating cytotoxin of *Helicobacter pylori*. *Gastroenterology* 1997;112:92–9. [PubMed: 8978347]
7. Goossens H, Glupczynski Y, Burette A, Lambert J-P, Vlaes L, Butzler J-P. Role of the vacuolating toxin from *Helicobacter pylori* in the pathogenesis of duodenal and gastric ulcer. *Med Microbiol Lett* 1992;1:153–9.
8. Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, Massone A, Papini E, Xiang Z, Figura N, Rappuoli R. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci USA* 1993;90:5791–5. [PubMed: 8516329]
9. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* 1996;93:14648–53. [PubMed: 8962108]
10. Peek RM Jr, Thompson SA, Donahue JP, Tham KT, Atherton JC, Blaser MJ, Miller GG. Adherence to gastric epithelial cells induces expression of a *Helicobacter pylori* gene, *iceA*, that is associated with clinical outcome. *Proc Assoc Am Phys* 1998;110:531–44. [PubMed: 9824536]
11. van Doorn L-J, Figueiredo C, Sanna R, Quint W. Genetic organization and molecular typing of the *Helicobacter pylori iceA* gene (Abstract). *Gastroenterology* 1998;114:A320.
12. van Doorn L-J, Figueiredo C, Sanna R, Plaisier A, Schneeberger P, de Boer W, Quint W. Clinical relevance of the *cagA*, *vacA* and *iceA* status of *Helicobacter pylori*. *Gastroenterology* 1998;115:58–66. [PubMed: 9649459]
13. Yamaoka Y, Kodama T, Gutierrez O, Kim JG, Kashima K, Graham DY. Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome. *J Clin Microbiol* 1999;37:2274–9. [PubMed: 10364597]
14. Morgan RD, Campbell RR, Wilson GG, Xu SY. Molecular cloning and expression of *NlaIII* restriction-modification system in *E. coli*. *Gene* 1996;183:215–8. [PubMed: 8996109]
15. Xu Q, Peek RM, Miller GG, Blaser MJ. The *Helicobacter pylori* genome is modified at CATG by the product of *hpyIM*. *J Bacteriol* 1997;179:6807–15. [PubMed: 9352933]
16. Wilson GG. Organization of restriction-modification systems. *Nucl Acids Research* 1991;19:2539–66.
17. Cover TL, Tummuru MKR, Cao P, Thompson SA, Blaser MJ. Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. *J Biol Chem* 1994;269:10566–73. [PubMed: 8144644]
18. Labigne-Roussel A, Courcoux P, Tompkins L. Gene disruption and replacement as a feasible approach for mutagenesis of *Campylobacter jejuni*. *J Bacteriol* 1998;170:1704–8. [PubMed: 2832375]
19. Maniatis, T.; Fritsch, EF.; Sambrook, J. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory; Cold Spring Harbor, NY: 1989.
20. Engstrand L, Nguyen AM, Graham DY, el Zaatari FA. Reverse transcription and polymerase chain reaction amplification of rRNA for detection of *Helicobacter* species. *J Clin Microbiol* 1992;30:2295–301. [PubMed: 1383268]
21. Peek, RM.; Miller, GG.; Blaser, MJ. Reverse transcription-polymerase chain reaction (RT-PCR) for bacterial and host gene expression in human gastric mucosa.. In: Lee, AX.; Megraud, F., editors. *Helicobacter Pylori: Techniques for Clinical Diagnosis and Basic Research*. W.B. Saunders. Co.; London: 1995. p. 163-72.
22. Tomb JF, White O, Kerlavage AR, Clayton RA, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997;389:539–47. [PubMed: 9335485]
23. de Haseth PL, Zupancic ML, Record MT. RNA polymerase–promoter interactions. *J Bacteriol* 1998;180:3019–25. [PubMed: 9620948]
24. Alm RA, Ling LSL, Moir DT, et al. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 1999;397:176–80. [PubMed: 9923682]
25. Pettersson J, Nordfelth R, Dubinina E, Bergman T, Gustafsson M, Magnusson IE, Wolf-Watz H. Modulation of virulence factor expression by pathogen target cell contact. *Science* 1996;273:1231–3. [PubMed: 8703058]

26. Heithoff D, Sinsheimer RL, Low DA, Mahan MJ. An essential role for DNA adenine methylation in bacterial virulence. *Science* 1999;284:967–70. [PubMed: 10320378]
27. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463–7. [PubMed: 271968]

Figure 1.

Nucleotide sequence of a 975-bp DNA fragment containing the 3'-end of the *cysE*, the complete *iceA1*, and the 5'-end of *hpyIM* from *H. pylori* strain 60190. Nucleotide position I corresponds to nucleotide residue 412 of the *cysE-iceA1-hpyIM* sequence deposited in Genbank (accession number U43917). The derived amino acid sequences of the 3' segment of *cysE*, *iceA1* ORF-I and ORF-2, and the 5' segment of *hpyIM* are shown underneath the nucleotide sequence and are boxed and labeled. The arrowhead between nucleotide residues 331 and 332 indicates the site of a I-nucleotide deletion that, if restored, would permit the translation of a 228 amino acid protein with 66% homology (identity + similarity) to *NlaIII* (*iceA1* ORF-1 + ORF-2). The positions of each *iceA1* transcript 5'-termini determined by primer extension analysis are

indicated by vertical arrows. The - 10 and - 35 sequences of the *iceAI* P_I promoter are double underlined and labeled. The location of the *EagI* site used for insertion of the *aphA* cassette is indicated.

Figure 2.

Analysis of *H. pylori* *iceA1* and *hpyIM* transcription. (A) Schematic representation of the *cysE-iceA1-hpyIM* genetic locus in wild type strain 60190. Also shown is the location of the *aphA* cassette in the isogenic mutant strain 60190I⁻. The *cysE*, *hpyIM* and *aphA* ORFs are represented by open boxes with the predicted size of each ORF indicated. The open box labeled *iceA1* corresponds to a 684-bp region that is homologous to *nlaIII*R [14] and is comprised of two potential ORFs (ORF-I and ORF-2) as shown in Fig. 1. The lines between the boxes represent intergenic flanking regions. The direction of transcription is shown by arrows. The location of DNA fragments P-1 and P-2 used to probe the Northern blots shown in (B) are indicated. (B) Autoradiogram of Northern blots of strains 60190 and 60190I⁻ RNA probed with ³²P-labeled

DNA fragments P-1 or P-2. Below each lane is an autoradiogram of a section of the Northern blot showing 16S rRNA. The positions of the 0.24–9.5 kb RNA ladder used to estimate transcript sizes and the estimated sizes of detected transcripts and 16S rRNA are indicated.

A

B

Figure 3.

Localization of the 5'-termini of the major *iceA1* transcripts. (A) Schematic representation of the *H. pylori* *cysE-iceA1-hpyIM* genetic locus. The relative location in *iceA1* and *hpyIM* of DNA fragments P-3 through P-7 (5' the 3' nucleotide residues of each in parentheses; see Fig. 1) used as probes are indicated. The approximate location of the site of insertion of the *aphA* cassette in *iceA1* is indicated by an arrowhead. Other labelling and representations are as described for Fig. 2. (B) Autoradiogram of Northern blots of strain 60190 RNA probed with ^{32}P -labeled DNA fragments P-3 through P-7. The positions of the 0.24–9.5 kb positions RNA ladder used to estimate transcript sizes and the estimated sizes of detected transcripts are

indicated. Below each lane, except for P-7, is an autoradiogram of a section of the Northern blot showing 16S rRNA.

Figure 4.

Primer extension mapping of the 5'-termini of *iceAI* transcripts. (A) Shown is an autoradiogram of an 8% polyacrylamide sequencing gel. The ^{32}P -labeled DNA primer, R7 (Table 1), was annealed to RNA isolated from strains 60190 or J178 and extended with reverse transcriptase. (B) Shown is an autoradiogram of an 8% polyacrylamide sequencing gel. The ^{32}P -labeled R3 oligonucleotide primer (Table 1) was annealed to RNA isolated from strain 60190 and extended with reverse transcriptase. The DNA sequences shown in (A) and (B) were generated by the dideoxy chain-termination method [27] using plasmid pVU1002 (containing the 975 bp DNA fragment shown in Fig. 1) as template and either ^{32}P -labeled R7 or R3 oligonucleotide primers,

respectively. Nucleotide positions are indicated in accordance with the numbering of the complementary sequence shown in Fig. 1.

Figure 5.

RT-PCR analysis *iceA1* transcription. (A) Schematic representation of *cysE-iceA1-hpyIM* genetic locus. The positions of putative promoters located upstream of *cysE* and in the *cysE-iceA1* intergenic region, and *iceA1* P_I promoter are indicated by arrows. The relative locations of the oligonucleotide primers R5, R9, and R10 used for reverse transcription of RNA and the PCR products amplified by primer sets F1/R3, F2/R4, and F6/R8 shown in (B) or (C) are indicated. Other labelling and representations are as described for Fig. 2. (B) Shown is a 1.2% agarose gel stained with ethidium bromide. Reverse transcription of DNase I-treated RNA isolated from *H. pylori* strain 60190 was primed with oligonucleotide primer R10. Primer set F1/R3 PCR amplification of RT reactions with reverse transcriptase or without reverse

transcriptase as a negative control, or strain 60190 genomic DNA as a positive control are indicated above the respective lanes. (C) Shown is a 1.2% agarose gel stained with ethidium bromide. The primer set F2/R4 was used for PCR amplification of 10-fold serial dilutions (10^{-1} – 10^{-4}) of the oligonucleotide R5 primed RT reaction with reverse transcriptase (lanes 1–4) or without reverse transcriptase (10^{-1} dilution) as a negative control (lane 5), and the primer set F6/R8 was used for PCR amplification of 10-fold serial dilutions (10^{-1} – 10^{-4}) of the oligonucleotide R9 primed RT reaction with reverse transcriptase (lanes 6–9) or without reverse transcriptase (10^{-1} dilution) as a negative control (lane 10).

Table 1

Oligonucleotide primers used in this study

Gene	Primer	Sequence (5' to 3')	Strand	Location ^a
<i>cysE</i>	F1	GGGTGCGATTTGCGTGGGCGATG	+	1–23
<i>iceA1</i>	F2	GCGGGTTATTTGAGGTAAATTG	+	152–174
<i>iceA1</i>	R3	CTGTAACGCTTACCCAACGACTCACC	–	287–262
<i>iceA1</i>	R4	CTTTAGCCAAAGCTGAGCTATTC	–	364–342
<i>iceA1</i>	R5	ATTCTATCAATAGAATTTCCTAGA	–	419–396
<i>iceA1</i>	F6	GTGTTTTTAACCAAAGTATC	+	446–465
<i>iceA1</i>	R7	GTGTTTTCAGAGTTGCCACGCACACC	–	539–514
<i>iceA1</i>	R8	AGCTTTCATAAAGCCTGAAAAT	–	639–617
<i>iceA1</i>	R9	GTTTTCTGTATTGTATGGGGTC	–	803–781
<i>hpyIM</i>	R10	GATCATGGCCTACAACCGCATGGA	–	975–952

^aCoordinates are based on the nucleotide sequence of the 975 bp DNA fragment shown in Fig. 1.