PLASMID CONTAINS UNIQUE *HELICOBACTER PYLORI* INSERTION SEQUENCE, IS*605*

A DISSERTATION

SUBMITTED IN PARITAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF BIOLOGY

COLLEGE OF ARTS AND SCIENCES

BY

KARA DAWN BURNHAM, B.A., M.S.

DENTON, TEXAS

AUGUST, 1998

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS

Date July 7, 1998

To the associate Vice president for Research and Dean of the Graduate School:

I am submitting herewith a dissertation written by Kara Dawn Burnham entitled "Plasmid contains unique *Helicobacter pylori* insertion sequence, IS*605*." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Molecular Biology.

Sarah A McIntire
Dr. Sarah A. McIntire, Major Professor

We have read this dissertation and recommend its acceptance:

Department Chairperson

Accepted
Linda M Thompson
Associate Vice President for Research and
Dean of the Graduate School

ACKNOWLEDGMENTS

I would like to thank each member of my committee: Dr. Sarah McIntire, Dr. John Knesek, Dr. Micheal Rudick, Dr. Heather Conrad-Webb and Dr. Robert Benjamin for their knowledge and assistance throughout my years at TWU. Special thanks is extended to my advisor, Dr. Sarah McIntire for her guidance and knowledge of *Helicobacter pylori* and scientific writing. Her continued support and encouragement was greatly appreciated.

I would also like to thank my parents, Dick and Sally Burnham and my sister, Krista for supporting and encouraging me throughout my academic career. Also a special note of thanks to Olivia Jones who lived with me and encouraged me daily to work towards my doctoral degree.

Kara Dawn Burnham
August, 1998

ABSTRACT

PLASMID CONTAINS UNIQUE *HELICOBACTER PYLORI* INSERTION SEQUENCE, *IS605*

KARA DAWN BURNHAM

August 1998

Helicobacter pylori is a human gastric pathogen that causes chronic gastritis, peptic ulcer disease and gastric adenocarcinoma. Other pathogenic bacteria, including *Salmonella* and *Escherichia*, contain stretches of sequence that encode pathogenicity genes, and are referred to as pathogenicity islands (PAI); a 40 kbp chromosomal PAI occurs in pathogenic strains of *H. pylori*. IS*605*, an insertion sequence unique to *H. pylori* is found in the PAI of some isolates and is thought to be involved in the deletion formations that are noted in these strains.

Plasmids have been speculated to be responsible, in part, for the rearrangements of the *H. pylori* PAI. In this study, we describe a 13 kbp plasmid, pHPM186, which contains two copies of IS*605*. In addition, pHPM186 carries 1.6 kbp of DNA that show strong sequence identity to region 86 of the *H. pylori* chromosome, which encodes a PARA protein and 2 hypothetical proteins of unknown functions. Also pHPM186 has a replication gene (*repA*) with strong sequence identity to *repA* genes found in pHPM180, pHPM179 and pHel. The sequence analysis of pHPM186 provided direct evidence that portions of the *H. pylori* chromosome can be carried by plasmids, making these plasmids possible vehicles for movement of chromosomal DNA between *H. pylori* strains.

Creation of a shuttle vector between *H. pylori* and *E. coli* was also attempted in this study. Ligations of the *repA* region of pHPM186 and a kanamycin resistance gene were transformed into both *E. coli* and *H. pylori*. Transformants were recovered only when F^+ strains of *E. coli* were used as recipients. Two different sizes of recombinants were isolated that were resistant to kanamycin, but in both the *repA* region was rearranged.

Small *H. pylori* transformants were initally recovered, but they failed to grow. Attempts to create a shuttle vector showed that the *repA* gene from *H. pylori* is not maintained in *E. coli,* it appears to undergo vast rearrangement. These results provide evidence that a creation of a shuttle vector between this two strains is not possible using this type of recombinant construct.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Helicobacter pylori is a human pathogen that causes chronic gastritis, peptic ulcer disease and gastric adenocarcinoma. The bacterium is a spiral shaped Gram negative organism that infects more than 50% of the human population (3). Due to the high incidence of infection and the virulence of this organism it is important to understand the pathogenesis of *H. pylori*. Consequently, the basic science and clinical aspects of *H. pylori* are being investigated by many laboratories.

The pathology of *H. pylori* is not completely understood. A single mechanism of pathogenesis has not been elucidated, but many factors which may play a role have been described. *H. pylori* produces multiple flagella that allow bacterial motility. Eaton et al. (12) demonstrated that aflagellate *H. pylori* do not induce gastritis, so motility may be a virulence factor. Adherence of *H. pylori* to gastric cells also may be a factor in promoting disease (26). *H. pylori* possess several fibrillar surface antigens which allow the bacteria to adhere to cells. One or more of these antigens may be specific for virulent strains of the organism (13). All disease-causing *H. pylori* strains isolated thus far express urease. The urease enzyme may play a virulence role by producing ammonia from urea, thus producing a micro-change in pH and allowing the bacteria to colonize the gastric mucosa (23). It also has been proposed that urease disrupts the ion flow between cells by disturbing tight cell junctions (33). In vitro vacuolation of human epithelial cells by an *H. pylori* cytotoxin has led to the suggestion that cytotoxin activity may play a virulence role (15), since *H. pylori* strains that cause this vacuolation of cells are found more often in patients with duodenal ulcers than in patients without ulcers (19). The vacuolating cytotoxin is a protein called VacA, encoded by the gene *vacA* (27). Also, the cytotoxin associated gene (*cagA*) has been linked to the occurrence of disease. The *cagA* gene was first identified when a 120-128 kDa protein was isolated from *H. pylori* strains recovered from individuals with peptic

ulcer disease. This protein was present in strains that produce cytotoxin, so that when the gene producing the protein was isolated, the gene was named the **c**ytotoxin **a**ssociated **g**ene (35). This gene has now become a marker for virulence, although the exact function of the CagA protein is not known. Pathogenic strains are *cagA*$^+$ or type I, and non-pathogenic strains are *cagA*$^-$ or type II (37). The protein product of *cagA* ranges from 120 kDa to 128 kDa due to the presence of one or more copies of a 102 bp insert in the gene, which accounts for the variation in protein size (14). A final characteristic of *H. pylori* that may contribute to pathogenicity is the genetic diversity that is present among *H. pylori* strains (21). Such diversity could contribute to the microorganism's ability to adapt to new hosts and to evade the host's immune system. This feature of *H. pylori* may be unique to this species; for example, the order of the genes in *Escherichia coli* strains is fairly constant (10).

Other pathogenic genera of bacteria, including *Salmonella* (29) and *Escherichia* (4), contain stretches of DNA that encode pathogenicity genes. These sequences are referred to as pathogenicity islands (PAI). Recently, Censini et al.(6) described a 40 kbp PAI in *H. pylori*, which includes the gene *cagA*. The *H. pylori* PAI also includes genes *cagB* , *cagC*, *cagD*, *cagE*, *cagF*, *cagG*, *cagH*, *cagI*, *cagL*, *cagN*, *cagM*, *cagO*, *cagP*, *cagQ*, *cagR*, *cagS*, *cagT*.

Many of the *H. pylori PAI* genes have weak identity to known sequences; *cagE*, for example, has identity to several virulence genes from other organisms. It is similar to the *ptlC* gene of *Bordetella pertussis*, the *virB4* gene of *Agrobacterium tumefaciens* and the *traB* gene of the IncN plasmid pKM101. It is suspected that some of the remaining genes will also encode virulence factors (6).

Arrangement of the *H. pylori* PAI differs among type I isolates and is found in two distinct pieces, designated *cagI* and *cagII* (Fig. 1). *CagI* includes *cagA* through *cagR* and *cagII* includes *cagS*, *cagT* and the *tnp* genes. *CagI* and *cagII* are often, but not always, separated by an insertion sequence, IS*605*, which encodes two *tnp* transposases. The *tnp A* gene has high identity to a transposase found in *E. coli* IS*200* and the *tnpB* gene is

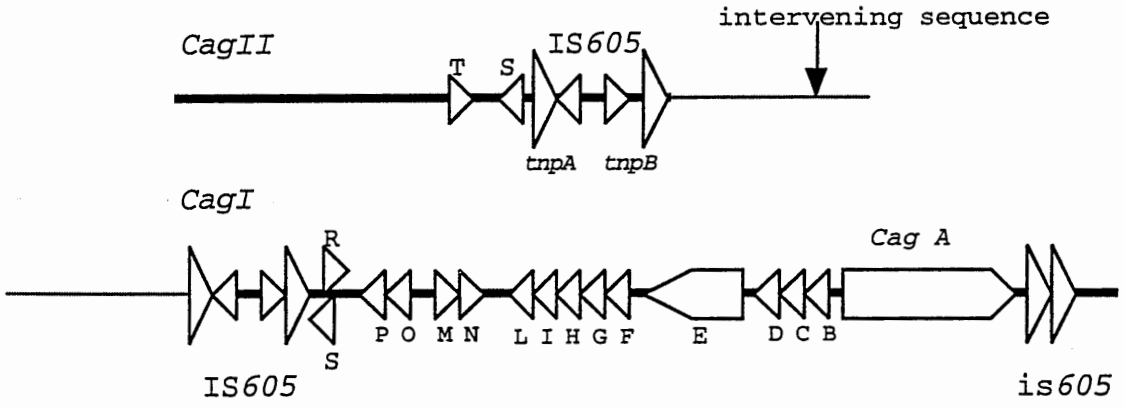

Fig. 1. The *H. pylori* PAI, adapted from Censini et al. (6). The arrows indicate direction of transcription of the gene. The letters were assigned to each gene for identification, i.e. *cagP*, *cagM* etc.

similar to the transposase found on IS*1341* in a thermophylic bacterium *PS3*. *CagI* and *cagII* can also be separated by an intervening sequence of variable DNA sequence and length, flanked by two IS*605* elements. In other strains, portions of the PAI are inverted or deleted. Gene arrangement in the PAI is quite diverse among strains, with the strains carrying *cagI* connected to *cagII* representing pathogenic type I strains. Several intermediate derivatives have been identified: a) those with one to several copies of IS*605*; b) those with partial or total deletion of *cagI* or *cagII* genes; and c) those with all of *cagI* and *cagII* deleted, but retaining IS*605*; and d) a non-pathogenic type II strain with both *cagI* and *cagII* deleted and no copies of IS*605* (6). Therefore, it appears that these rearrangements in the *H. pylori* PAI contribute to pathogenesis.

One possible factor contributing to genetic diversity among *H. pylori* strains carrying the PAI is the presence of plasmids in approximately 80% of *H. pylori* isolates (25). *H. pylori* plasmids differ in size, ranging from 1.5 to 40 kbp in size (20, 22, 25, 28). These extrachromosomal pieces of DNA may recombine into the chromosome and then excise, taking portions of chromosome, as occurs with F plasmids in *E. coli* (36). A possible site for recombination could be the 102 bp repeat or IS*605*, both of which are found throughout the chromosome (14, 6).

Numerous reports suggest that plasmids may be involved in sequence rearrangements. Some of the genes located in the *cagI* fragment of the *H. pylori* PAI show identity to genes which are speculated to have come from plasmids. In addition, genes of the *H. pylori* PAI also have a much higher G+C content than the genes found on the chromosome, suggesting that portions of the PAI are of extrachromosomal origin (10). Evans and Evans (14) recently analyzed published data and describe partial sequence identity (42%) between the 102 bp repeat sequence of *cagA* and portions of plasmids pHPM180 and pHel1 (25, 20). Qasem (28) reported that pHPM179 contains 44 bp of identity to a sequence of DNA outside of the chromosomal *cagA* open reading frame (ORF). Further examination of pHPM179 DNA sequence shows that it occurs in two different sizes, 7.6 kbp or 5.6 kbp. This size difference is due to a 2 kbp transposon that

inserts into the plasmid in a region with strong identity to a portion of the 102 bp insert (32).

Thus, it is possible that portions of the *H. pylori* PAI may be moved around by plasmids either via transformation or conjugation (21). In addition, plasmids may integrate and excise from the chromosome, thereby contributing to the genetic diversity observed in *H. pylori*. As a result, genes normally found on the chromosome may be detected on large plasmids. The first objective of this study was to identify *H. pylori* plasmids that carry segments of *H. pylori* chromosomal DNA. The study began by using a fragment of the *cagA* gene as a DNA probe against *H. pylori* plasmids. *CagA* was selected because it is an indicator of pathogenicity. One such plasmid was discovered, pHPM186, and the DNA sequence of the entire plasmid was determined.

The second objective of this study was to create a recombinant plasmid containing the plasmid *repA* gene from *H. pylori* and a kanamycin resistance marker. This recombinant plasmid would be transformed into *E. coli* strains, as a first step in the creation of a shuttle vector between *H. pylori* and *E. coli*.

CHAPTER II
MATERIALS AND METHODS
Materials

Strains and Plasmids

H. pylori strains HPM180F, HPM179Aa, HPM179Ab, HPM165A, HPM186A, HPM185F, HPMEA4F, HPM182F, HPM184DB, HPM181A were originally obtained from the VA Medical Center in Dallas, TX. The site of the gastric biopsies from which the strain was obtained is indicated by the letters A: antrum, F: fundus, DB: duodenal bulb, at the end of the strain number. Frozen cultures of these strains in trypticase soy broth containing 25% glycerol and 10% horse serum were provided by Dr. Sarah McIntire. All of the strains were grown at 37°C under microaerophilic conditions (5-12% CO_2) on brain heart infusion (BHI) medium with 10% horse serum.

The plasmid pZERO-2.1 was purchased from Invitrogen and was maintained in the *E. coli* strain TOP10F′, also from Invitrogen. *E. coli* TOP10F′ is: F′{*lac*I^{q}Tn10(Tet^R)} *mcr*AΔ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L *end*A1 *nup*G. *E. coli* TOP10F′(pZERO-2.1) was grown on NZY-kanamycin agar medium at 37°C. Three other *E. coli* strains were: DH10B, F$^-$ *mcr*AΔ (*mrr-hsd*RMS-*mcr*BC)ϕ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139Δ(*ara-leu*)7697*gal*U *gal*K *rps*L *end*A1 *nup*G; DH5αMCR, F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ (*lac*ZYA-*arg*F) U169 *deo*R *rec*A1 *end*A1 *pho*A *sup*E44 I$^-$ *thi*-1 *gyr*A96 *rel*A1; DH5αF′, F′ϕ80d*lac*ZΔM15 Δ (*lac*ZYA-*arg*F) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (rk$^-$,mk$^+$) *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1.

Media

BHI medium (Difco) was prepared following the manufacturer's instructions with the addition of 10% horse serum (Sigma). NZY medium (BIO 101, Inc.) was prepared following the manufacturer's recommendations, with the addition of kanamycin to a final concentration of 10, 20, 30, or 50 μg/ml (Sigma). Solid media was made by adding Bacto-Agar (Difco) to 1.5%.

Primers

Primers used for PCR amplification reactions were purchased from Bio-Synthesis, Inc., Lewisville, TX. Primers were designed using OLIGO version 3.4 software for IBM. Primers used for automated sequencing, IRD41-labeled M13 forward and reverse, were purchased from LI-COR, Inc., Lincoln, NE. Primer sequences are given in Table 1; the derivation of each primer is given in the results section.

Enzymes

Enzymes used were lysozyme (Sigma), rTth DNA polymerase, XL (Perkin-Elmer), restriction endonucleases (Promega), RNase A (Sigma), proteinase K (Ambion), shrimp alkaline phosphatase (USB), Fast-Link DNA ligase (Epicentre technologies), Klenow enzyme (Amersham/USB), and T7 sequenase version 2.0 (Amersham/USB).

Commercial kits

The commercial kits used for DNA purification were the PERFECTprep plasmid DNA kit (5 PRIME-3 PRIME, Inc), the Wizard Plus Miniprep DNA purification system (Promega), and the Elu-Quik DNA purification kit (Schleicher & Schuell). Ligations were performed using the Fast-Link DNA Ligation and Screening kit (Epicentre technologies). The GeneAmp XL PCR kit was used for PCR reactions (Perkin-Elmer). Sequencing reactions were performed using the T7 Sequenase version 2.0 DNA sequencing kit

(Amersham). Lastly, the DECAprime II DNA labeling kit was used for DNA probe preparation (Ambion, Inc.). All kits were used as instructed by the manufacturer.

DNA Sequencing Gel

Preparation of the 66 cm, 6% Long Ranger sequencing gel required urea (Amresco), Long Ranger™ 50% gel concentrate (FMC), γmethacryloxypropyltrimeth-oxysilane solution (Sigma), 400 µl of ammonium persulfate (Amresco), and 40 µl of TEMED (Amresco).

DNA Transfer and Hybridization

Zeta-Probe GT genomic-tested nylon blotting membranes (BIO-RAD) and Whatman 3MM paper were used for DNA transfer. Reagents and equipment used for hybridization included QuikHyb solution (Stratagene), sonicated salmon sperm DNA (Stratagene), [α^{32}P]-dCTP (New England Nuclear), NucTrap Probe Purification Columns (Stratagene), Push Column Beta Shield Device (Stratagene), UV Stratalinker model 1800 (Stratagene), hybridization incubator model 310 (Robbins Scientific), 50 ml Autoblot glass bottles (Bellco), and X-OMAT AR film (Kodak).

Buffers and other reagents

Buffers and other solutions used included, 5X TBE, TE (10 mM Tris, pH 7.5, 1 mM Na2EDTA), buffer I (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCL, pH 8.0), buffer II (0.2 N NaOH, 1% SDS), buffer III (3 M potassium acetate, 1.8 M formic acid, pH 4.8), wash buffer (100 mM potassium acetate, 50 mM MOPS, pH 8.0), LiCl solution (10 mM LiCl, 50 mM Tris-HCL, pH 8.0), TES (50 mM Tris, pH 8.0, 5 mM EDTA, 50 mM NaCl), lysis solution (25 µl 5% SDS, 75 µl TE, 50 µl proteinase K, 20 mg/ml), 1X STE (100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM EDTA), 20X SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0), topoisomerase I 5X buffer (1 M Tris, pH 7.5, 1 M KCl, 1 M $MgCl_2$, 1 M DTT, 0.5 M EDTA, 1 mg/ml 10X BSA), and CaRb solution (0.1 M calcium

chloride, 0.01 M rubidium chloride).

Other reagents included 100%, 95%, and 70% ethanol, 1 M Tris-HCl, pH 7.5 and pH 8.0, 0.5 M EDTA, pH 8.0, 1.5 M and 3 M sodium acetate, pH 5.2, phenol : chloroform: IAA, chloroform, 1 mg/ml ethidium bromide, 5X Ficoll dye (0.1 M EDTA, 25% Ficoll, 0.1% bromophenol blue, 0.1% xylene cyanol), SeaKem agarose (FMC), isopropyl–β-D-thiogalactoside (IPTG), and a DNA size standard, 1 kb DNA ladder (BRL).

Methods

A Savant Speedfuge HSC10K with a HSR-24 rotor was used for centrifuging microfuge tubes at room temperature. A Beckman JS-HS centrifuge with either a JS-7.5 swinging bucket rotor for 50 ml conical tubes, or a JA-18.1 fixed angle rotor for microfuge tubes was used for all other centrifugations.

Plasmid preparation

Two methods of *H. pylori* plasmid preparation were used, the method of Minnis et al. (19), and the Wizard Miniprep kit (Promega). The method of Minnis et al. (25) yielded cleaner DNA than did the Wizard kit and this DNA was more suitable for PCR. The Wizard kit DNA could be used for restriction endonuclease digests.

The method of Minnis et al. (25) used 400 ml of bacterial culture. Four 175 cm^2 tissue culture flasks, each containing 100 ml of BHI supplemented with 10% horse serum, were inoculated with *H. pylori* cells from plates and grown at 37°C for three to five days. The cells were transferred to four 50 ml conical tubes and centrifuged at 5,000 rpm at 4°C for 10 min. Pelleted cells were suspended in a total volume of 16 ml TE and aliquoted into sixteen 1.5 ml microfuge tubes. The tubes were microfuged at top speed for 1 min at room temp. Each pellet was suspended in 50 µl buffer I by vortex, placed on ice, and an additional 50 µl buffer I containing 20 mg/ml of lysozyme was added. The tubes were inverted to mix and placed on ice for 10 min. Two hundred microliters of buffer II was added to each tube, the tubes were mixed and left on ice for 5 min. One hundred fifty microliters of buffer III was added, the tubes were mixed and left on ice for 15 min. The tubes were centrifuged at 9,000 rpm at 4°C for 15 min. The DNA was precipitated from the supernatants by adding 1 ml of 95% ethanol and incubating the tubes at -20°C for 20 min. The tubes were centrifuged as in the previous step and the pellets suspended in 100 µl of wash buffer and 200 µl of 95% ethanol. The tubes were incubated at -20°C for 20 min and centrifuged as before. Wash buffer and ethanol were added and the tubes

incubated as done previously. After centrifugation at 9,000 rpm and 4°C for 15 min the pellets were air dried and 100 µl of ultrapure water added along with 100 µl of LiCl solution. The tubes were incubated on ice for 30 min and centrifuged at 9,000 rpm and 4°C for 15 min. The supernatants of two tubes were combined to concentrate the DNA. To each of the eight tubes 28 µl of 1.5 M sodium acetate and 800 µl of 95% ethanol was added. The tubes were incubated at -20°C for 20 min and centrifuged at 9,000 rpm and 4°C for 20 min. Each pellet was suspended in 50 µl wash buffer, mixed and combined to again concentrate the DNA, leaving four tubes. Two volumes of 95% ethanol (200 µl) was added to the tubes, the tubes were incubated at -20°C for 20 min, and centrifuged at 9,000 rpm and 4°C for 20 min. The pellets were washed with 200 µl of 70% ethanol and air dried. Each pellet was suspended in 25 µl of ultrapure water and stored at 4°C. RNA was eliminated from the plasmid DNA prep by using the Elu-Quik DNA purification kit as instructed by the manufacturer.

The Wizard miniprep kit, which was designed for *E. coli,* was used for *H. pylori* with the following modifications to the manufacturer's protocol. One 12.5 cm^2 tissue culture flask containing 20 ml of BHI supplemented with 10% horse serum was inoculated with *H. pylori* cells and grown at 37°C for five days. The cells were divided into 10 ml aliquots in 50 ml conical tubes. The conical tubes were centrifuged at 5,000 rpm and 4°C for 10 min. The cell pellets were suspended in 5 ml of TE and centrifuged again as previously. This washing with TE was repeated twice to ensure that any residue from the BHI media was removed from the *H. pylori* cells. Each cell pellet was suspended in 1 ml of TE and transferred to 1.5 ml microfuge tubes. The two microfuge tubes were centrifuged at 9,000 rpm and 4°C for 10 min and each cell pellet was suspended in 200 µl of Cell Resuspension solution. Two hundred microliters of Cell Lysis buffer was added to each tube and mixed by inversion. Two hundred microliters of Neutralization Buffer was added to each tube and the tubes were centrifuged at top speed at room temperature for 5 min. The supernatants were then combined and the vacuum method described in the manufacturer's protocol was used to bind the DNA to the minicolumn. Ultrapure water

heated to 65°C was used to elute the plasmid DNA in the final step of the protocol.

Recombinant plasmids were purified using the PERFECTprep plasmid DNA kit. *E. coli* TOP10F′ cells containing the recombinant plasmid were grown in 1.5 ml of NZY broth overnight at 37°C. The PERFECTprep protocol was followed and ultrapure water heated to 65°C was used to elute the DNA in the final step.

H. pylori genomic DNA preparation

Genomic DNA was extracted from *H. pylori* cells using the method of Clayton et al. (8) with minor modification. One 12.5 cm^2 tissue culture flask containing 10 ml of BHI supplemented with 10% horse serum was inoculated with *H. pylori* cells and grown at 37°C for three days. The cells were pelleted in a 50 ml conical tube at 5,000 rpm and 4°C for 10 min. The cell pellet was washed with 1 ml TES and centrifuged again as before. The cell pellet was suspended in 1 ml TES and transferred to a 1.5 ml microfuge tube. The cells were again pelleted in the centrifuge at 9,000 rpm and 4°C for 5 min. The pellet was suspended in 200 µl cold 50 mM Tris-25% sucrose solution. One hundred microliters of lysozyme (30 mg/ml in 0.25 M EDTA, pH 8) was added and the tube incubated on ice for 20 min. Following incubation on ice, lysis solution was added and the tube incubated at 56°C for 1 hr. After incubation, 2 µl of RNase A (10 mg/ml) was added and the tube incubated at 37°C for 30 min. The DNA was extracted twice with an equal volume of phenol:chloroform and centrifuged at 9,000 rpm and 4°C for 15 min, followed by an extraction with an equal volume of chloroform. The DNA was precipitated from the aqueous phase by adding two volumes of 95% ethanol and incubating at -80°C for 30 min. The tube was then centrifuged at 9,000 rpm and 4°C for 30 min. The ethanol was removed and the DNA pellet allowed to dry at room temp. for 5 minl. The DNA pellet was suspended in 50 µl of ultrapure water and quantified on the Shimadzu UV160 spectrophotometer (OD_{260} of 1 = 50 µg/ml).

Gel Electrophoresis

Agarose gels were either 30 ml mini-gels or 90 ml standard gels. Mini-gels were 0.8% agarose in 1X TBE, and standard gels were 0.8% (plasmid DNA restriction endonuclease digests) or 1.0% agarose (genomic DNA restriction endonuclease digests) in 1X TBE. Electrophoresis was at 9.6 V/cm for mini-gels and 2.7-3.3 V/cm for standard gels in 1X TBE running buffer. All gels were stained with 0.25 mg/ml ethidium bromide and visualized under UV light. Gels were photographed with an MP-4 camera equipped with an orange filter using Polaroid type 57 film.

Restriction endonuclease analysis

The *H. pylori* plasmid DNAs were digested with *Hin*dIII restriction endonuclease, except for pHPM180, pHPMEA4F and pHPM165A, which were digested with *Dra*I. Recombinant plasmids of pZERO-2.1 and *H. pylori* DNA fragments were digested with various restriction endonucleases as detailed in results. *H. pylori* genomic DNA and pHPM186 DNA were digested with *Dde*I and *Dra*I restriction endonucleases. Restriction endonuclease digests contained 30-40 ng of DNA, 1.5 µl of the appropriate 10X buffer, and 10 units of endonuclease in a total volume of 15 µl and digestions were incubated at 37°C for 1-2 hours.

Competent cell preparation

E. coli cells were grown in 25 ml of NZY broth at 37°C with shaking to OD_{600} = 0.6-0.8 (Shimadzu UV160 spectrophotometer). The cells were pelleted by centrifugation in a 50 ml conical tube at 3,000 rpm and 0°C for 5 min. The pellet was suspended in 15 ml cold CaRb solution and incubated on ice for 30 min. The cells were pelleted again at 2,000 rpm and 0°C for 5 min and the pellet was suspended in 2 ml of cold CaRb solution. The cells were divided into 200 µl aliquots in 1.5 ml microfuge tubes. Thirty microliters of sterile glycerol was added to each tube, mixed and the tubes stored at -70°C.

H. pylori is naturally competent. The method of Haas et al. (18) was used to prepare the cells for transformation. The optimum optical density value for uptake of foreign DNA is 0.2 (A_{550}). A standard inoculating loop was used to transfer *H. pylori* cells from a 2-day old plate. The cells were added to 3 ml of liquid BHI supplemented with 10% horse serum. The cells were diluted to an A_{550} of 0.1 in 1 ml of BHI broth. The bacteria were allowed to recover by incubation for 12 hr in the CO_2 incubator (5% CO_2, 37°C). The optical density was measured, following recovery, to be 0.2 at A_{550}.

Creation of blunt ends

Ligation of two products with non-compatible ends required that the ends of both products be filled. The DNA was first heated to 65°C for 10 min in a volume of 50 µl and was allowed to cool to room temperature. After cooling, 1 µl of 10 mM dNTP's and 1 µl of Klenow enzyme was added and the mixture was then left at room temperature for 30 min. Following the incubation, the DNA was purified using the Elu-Quik kit.

Ligation reactions

The Fast-Link DNA Ligation and Screening kit protocol was used for all ligation reactions. Reactions to ligate blunt end DNA fragments used a ratio of 1:10, vector to insert; reactions to ligate staggered end DNA fragments used a ratio of 1:2, vector to insert. All reactions contained 150 ng of pZERO-2.1 vector DNA, and the appropriate amount of insert DNA as computed using LIGATE software for IBM. The vector was linearized with the appropriate restriction endonuclease and in the case of a staggered end ligation reaction, treated with 2 units of shrimp alkaline phosphatase (SAP). SAP enzyme is compatible with restriction endonuclease buffers and was directly added to the digestion reaction.

DNA Transformation

Three microliters of ligation reaction was added to 200 µl of competent *E. coli* cells and incubated on ice 30 min. The cell/DNA mixture was heat shocked at 42°C for 30 sec. and was placed on ice for 2 min. The cell/DNA mixture was added to 1 ml of NZY broth and incubated at 37°C for one hour. After incubation the cells were pelleted at 3,000 rpm at room temp. for 5 min. Eight hundred microliters of supernatant was removed and the pellet was suspended in the remaining 200 µl. The 200 µl of transformed cells were plated on an NZYM agar plate containing 50 µg/ml of kanamycin and 200 nM IPTG.

Transformation of *H. pylori* competent cells involved adding the foreign DNA to the cells grown to an O.D. of 0.2 (A_{550}). Three micrograms of genomic DNA was added to 1 ml of cells. When a ligation mixture was used for transformation, 5 µl of ligation mixture was used. The transformed *H. pylori* cells were plated on a BHI agar plate containing 10, 20, 30 or 50 µg/ml of kanamycin.

Screening for recombinant clones

The Fast-Link DNA Ligation and Screening kit was used to search for recombinant clones. Transformant colonies were picked using sterile toothpicks and suspended in 15 µl Fast-Link protoplast buffer. The colonies were incubated in protoplast buffer in microtiter plates for 10 min. The colony and buffer were then transferred to the well of an agarose gel which contained 4 µl of Fast-Link lysis buffer. Following electrophoresis the migration rate of DNA from the transformants could be compared to the migration rate of the supercoiled vector DNA to test for the presence of an insert.

PCR amplification

PCR amplification reactions were performed as recommended by the GeneAmp XL PCR kit and were done using the GeneAmp 2400 (Perkin Elmer). Each reaction contained 10 ng of template DNA, 50 pmol of each primer, 30 µl of 3.3XL Buffer, 4.4 µl of $Mg(OAc)_2$ (0.7-1.5 mM), 2 µl of a 200 nM solution of each dNTP (dATP, dTTP, dGTP,

and dCTP), and 2 µl of rTth DNA polymerase (2 units/µl) in a total volume of 100 µl. The amplification cycle for products <5 kb in length was programmed as follows:

Step 1: 94°C for 2 min
Step 2: 94°C for 45 sec
Step 3: 50°C for 45 sec
Step 4: 72°C for 45 sec
Step 5: go to Step 2, 35 times
Step 6: 72°C for 10 min
Step 7: hold at 4°C

The amplification cycle for products >10 kb in length was programmed as follows:

Step 1: 94°C for 1 min
Step 2: 94°C for 15 sec
Step 3: 68°C for 10 min
Step 4: go to step 2, 16 times
Step 5: 94°C for 15 sec
Step 6: 68°C for 10 min
Step 7: go to step 5 12 times
Step 8: 72°C for 10 min
Step 9: 4°C hold

The products from PCR reactions were separated by electrophoresis on agarose gels.

DNA sequencing

The double stranded DNA was prepared for sequencing by alkali denaturation. DNA sequencing reactions were done using the T7 Sequenase version 2.0 DNA sequencing kit. Approximately 1 ug of DNA was used per 1 kb of DNA to be sequenced. For example, to determine the sequence of a recombinant clone of pZERO-2.1 (3.2 kb) with a 1.0 kb insert, 4.5 ug of DNA was needed. The appropriate amount of DNA was

suspended in a total of 18 µl in a 1.5 ml microfuge tube. The DNA was denatured by adding 2 µl of freshly prepared 2N NaOH, 2 mM EDTA and incubating 5 min at room temp. Three microliters of 3 M sodium acetate was added as well as 7 µl of ultrapure water. The DNA was precipitated by adding 75 µl of 100% ethanol and incubating at -70°C for 15 min. The tube was then centrifuged at 9,000 rpm and 4°C for 15 min. The supernatant was removed and 200 µl of 70% ethanol was added. After centrifuging as before, the pellet was air dried.

Annealing of the primer and template was done by adding 2 µl of IRD41 labeled primer, 2 µl 5X Sequenase reaction buffer and 6 µl of ultrapure water, mixing and incubating at 37°C for 30 min. Following annealing, 3.5 µl of ultrapure water and 2.0 µl of diluted Sequenase enzyme (6.5 µl 1X TE, 0.5 µl pyrophosphatase, 1 µl T7 Sequenase version 2.0) was added to the primer/template mixture.

Elongation and termination was done by adding 3.5 µl of the template/primer/ enzyme mix to each termination tube, A, T, C or G. Termination tubes were prepared by adding 1 µl of termination mix (ddATP, ddTTP, ddGTP, or ddCTP) to 1.5 µl of 7-deaza-dGTP Extending Mix (LI-COR). The tubes were incubated at 37°C for 10 min and 4 µl of Stop Solution was added to each tube.

The 6%, 66 cm sequencing gel was prepared by mixing 25.2 g urea, 4.8 ml Long Ranger gel concentrate, and 14.4 ml 5X TBE. The mixture was poured into the gel plates as recommended by LI-COR. The gel polymerized after 2 hours and was placed into the LI-COR scanner.

The sequencing reactions were denatured by boiling for 2 min and placing the reactions on ice. Two microliters of each reaction was applied to the sequencing gel. Operation of the LI-COR 4000L sequencer and scanner was as instructed by LI-COR, Inc.

DNA Transfer and Southern hybridization

DNA to be transferred to a solid support was separated by electrophoresis on an agarose gel as described previously. The agarose gel was denatured in 0.5 N NaOH, 1.5 M NaCl for 1 hr at room temp. with gentle shaking. This was followed by neutralization in 1 M Tris, pH 7.5, 1.5 M NaCl for 1 hr at room temp. with gentle shaking.

Capillary transfer of the DNA was done according to Sambrook et al. (30). The nylon membrane (Zeta-probe) was soaked in ultrapure water for 5 min and then in transfer buffer (10X SSC) for 10 min. Two pieces of 3 MM Whatman paper also were soaked in 10X SSC. The transfer was done in a glass baking dish by assembling the following on a solid block: 3 MM Whatman wick, agarose gel, nylon membrane, 2 pieces of 3 MM Whatman paper, 7 cm stack of paper towels, and a 500 g weight. Transfer was allowed to proceed overnight. The DNA was crosslinked to the nylon membrane in the UV Stratalinker for 30 sec. The membrane was placed into a 50 ml Autoblot glass bottle with 10 ml of QuikHyb hybridization solution and rotated for 30 min at 68°C.

The appropriate DNA probe and 1 kb ladder were randomly labeled with [α^{32}P]-dCTP using the DECAprime II DNA labeling kit. Twenty-five nanograms of linear double stranded DNA was used in making probes using the protocol from Ambion. Unincorporated nucleotides were removed from the probe and 1 kb ladder samples with NucTrap push columns and the Push Column Beta Shield Device (Stratagene) as recommended. Counts per minute (CPM) were determined by placing 1 µl of sample into 7 ml of EcoLume scintillation solution (ICN) and counting in a Beckman LS9000 liquid scintillation counter. The CPM were 175,050 for the probe and 256,715 for the 1 kb ladder. Twenty microliters of labeled probe, 10 µl of labeled kb ladder, and 100 µl of 10 mg/ml sonicated salmon sperm DNA were denatured by boiling for 2 min and placed on ice. The denatured probe solution was mixed with 1 ml of prehybridization solution and placed back into the glass bottle. Hybridization continued for 1 hr at 68°C.

The membrane was washed twice with 200 ml of 2X SSC, 0.1% SDS at room temp for 15 min. The final wash was with 50 ml of 0.1X SSC, 0.1% SDS at 60°C. The

membrane was placed in a plastic bag and exposed to Kodak X-OMAT film for 24 hrs. The film was developed according to manufacturer's instructions. If the membrane was to be used again, it was stripped by placing in a boiling solution of 0.1X SSC and 0.1% SDS for 15 min.

CHAPTER III

RESULTS

PCR amplification of *cagA*

To investigate exchange between chromosomal and plasmid sequences in *H. pylori*, *H. pylori* plasmids were screened for *H. pylori* chromosomal DNA sequence. The first step was to use a fragment of the *cagA* gene as a DNA probe against all of the *H. pylori* plasmids in Dr. McIntire's collection. *CagA* was selected because it is a well characterized member of the PAI and is an indicator of pathogenicity. The primers KB 8 and KB 9 were used to amplify the *cagA* region from HPM180 chromosomal DNA. HPM180 was used because the plasmid sequence of pHPM180 (25) does not have any identity with *cagA*. This ensured that the DNA sequence amplified represented the *cagA* region found on the chromosome of HPM180. Sequences of the 23-28 nt oligonucleotide primers from Xiang et al. (37) are given in Table 1. The primers produced a fragment of 1300 bp (Fig. 2a) which is the expected size for the chromosomal *cagA* gene fragment.

Cloning and sequence determination of the 1300 bp *cagA* fragment

The 1300 bp *cagA* fragment was purified from the PCR reaction using the Elu-Quik kit as previously described. The *cagA* fragment was ligated to pZERO-2.1 which had been linearized with *Eco*RV restriction endonuclease. Recombinant plasmid DNAs were purified and digested with *Xba*I and *Pst*I endonucleases to confirm that the fragment was present. One recombinant containing a 1300 bp insert, pKBT1.3-1, was prepared for automated sequencing as previously described. Dot matrix analysis was used to compare the insert DNA sequence to the *cagA* DNA sequence (Fig. 2b). This analysis compared the two selected sequences by placing a "dot" at each x,y coordinate that corresponded to an

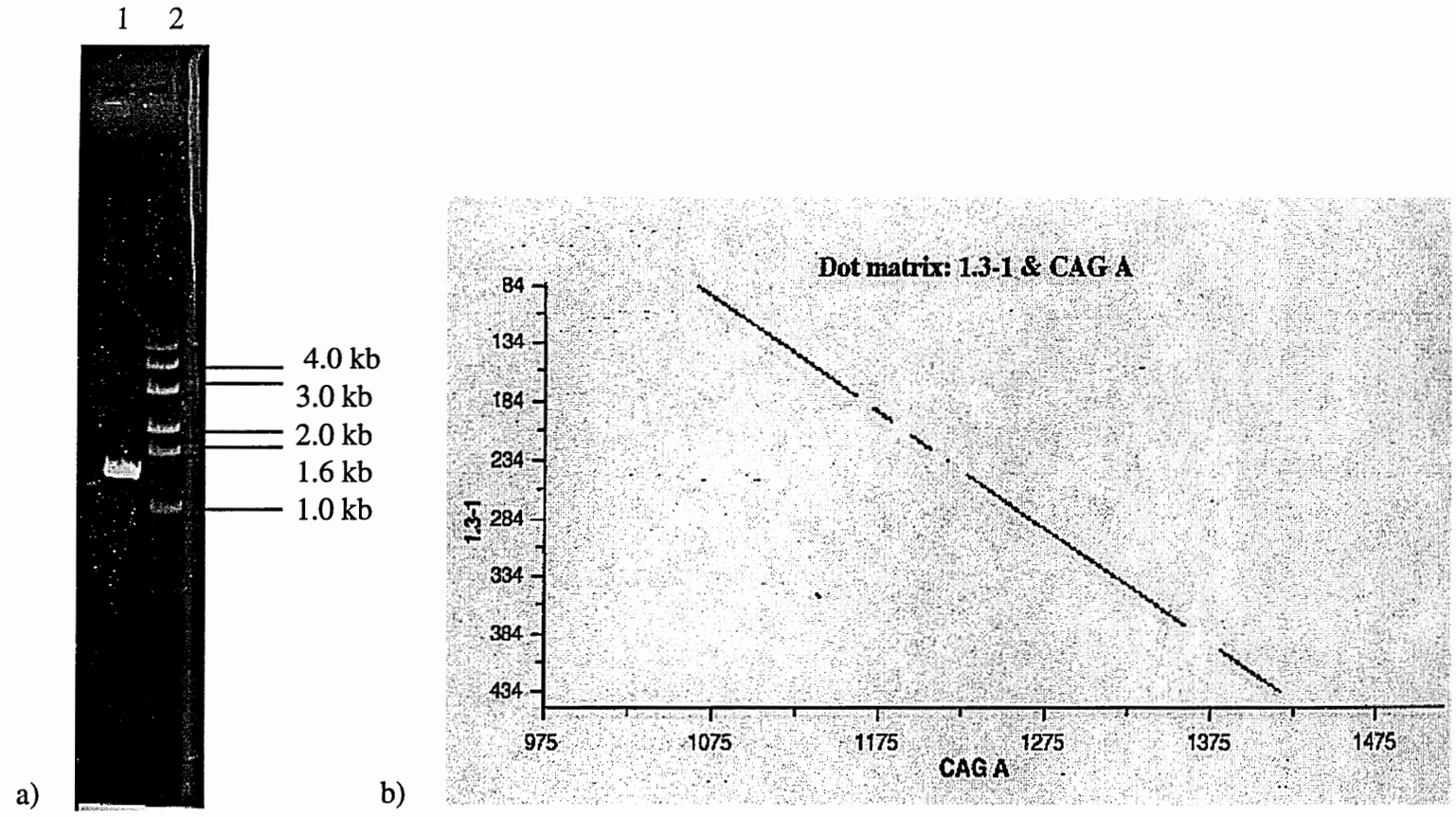

Fig. 2. Comparison of the DNA sequence of the insert from pKBT1.3-1 with the *cagA* gene DNA sequence using dot matrix analysis. a) Analysis of the PCR product produced from primers KB8 and KB9; lane 1, PCR product from produced using primer KB8 and KB9; lane 2, 1 kb ladder marker. b) Dot matrix showing identity between the DNA sequence of the insert from pKBT1,3-1 (y axis) and the DNA sequence of the *cagA* gene.

identity. Thus, when the identity is exact a line is visualized along the diagonal. This allows rapid evaluation of the identity of two DNA sequences. The insert DNA sequence from pKBT1.3-1 showed near perfect identity to the *cagA* region sequenced by Censini et al. (6), thus confirming that the appropriate fragment was cloned.

Analysis of *H. pylori* plasmids for chromosomal DNA

Eleven *H. pylori* plasmid DNAs: pHPM180, pHPM179a, pHPM179b, pHPM165A, pHPM186, pHPM185, pHPMEA4F, pHPM182, pHPM184DB, pHPM181 were purified (25) and screened by Southern blot for similarity to *cagA*. The a and b derivatives of 179A represent the 7.6 and 5.6 kb plasmids of pHPM179, respectively (32). The 11 plasmids were digested with either *Hin*dIII or *Dra*I restriction endonucleases according to Cissell et al. (7). The plasmid pHPM180 functioned as a negative control, as the sequence of this plasmid does not have identity with *cagA*. The 1300 bp *cagA* probe hybridized to one of the 11 plasmids, pHPM186, as shown in Fig. 3, indicating that pHPM186 may contain sequences of homologous DNA normally found on the chromosome of *H. pylori*.

Investigation of a plasmid containing *H. pylori* chromosomal DNA

Southern hybridization indicated that pHPM186 contained DNA sequence with similarity to at least a portion of the chromosomal *cagA* gene. Therefore, to determine whether all of the *cagA* gene was present and whether additional PAI genes were present, pHPM186 was chosen for complete DNA sequencing. Purified plasmid DNA from the 13 kbp pHPM186 was digested with the restriction enzyme *Hin*dIII, resulting in four fragments of 1300 bp, 2100 bp, 4500 bp and 5000 bp (Fig. 4). Digestion of pHPM186 with *Hin*dIII often produced a partially digested 7000 bp fragment seen in Fig. 4, which

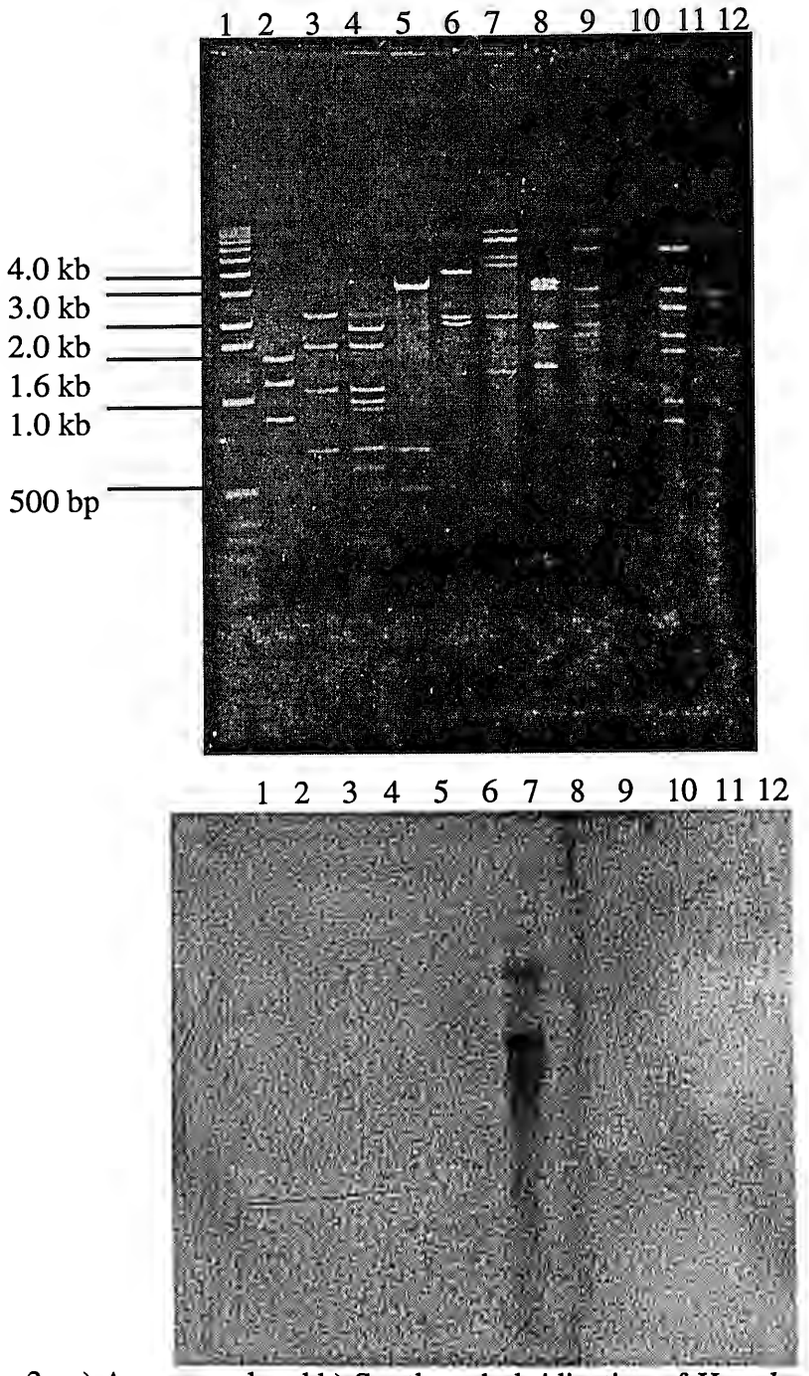

Fig. 3. a) Agarose gel and b) Southern hybridization of *H. pylori* plasmids with *cagA* probe. Lane 1, 1 kb size marker; lane 2, pHPM180; lane 3 pHPM179; lane 4 , pHPM179a; lane 5 pHPMEA8F; lane 6, pHPM181; lane 7, pHPM186; lane 8, pHPM184DB; lane 9, pHPM185; lane 10, pHPMEA4F; lane 11, pHPM182; lane 12, pPHM 165

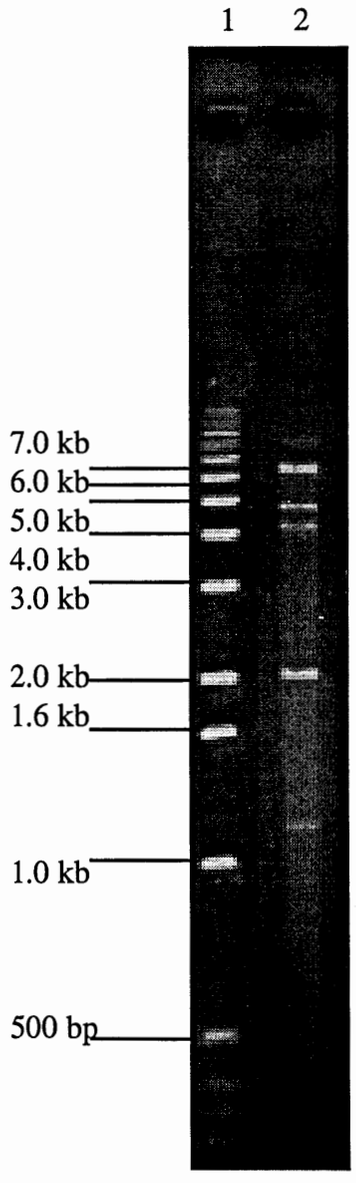

Fig. 4. pHPM186 digested with *Hin*dIII. Lane 1, 1 kb ladder marker; lane 2, pHPM186.

was composed of the 5000 bp and 2000 bp fragments. Each DNA fragment was purified and ligated with pZERO-2.1 DNA which had been linearized with *Hin*dIII and the ligation mixtures were used to transform competent *E. coli* TOP10F′ cells. Plasmid DNA was prepared from transformants and screened by restriction enzyme analysis to ensure that each piece had been cloned (Fig. 5). One recombinant containing each *Hin*dIII fragment was picked and designated pKB1.3, pKB2.1, pKB4.5 and pKB5.0, respectively. Recombinants that contained each fragment in the opposite orientation were identified by digesting each recombinant with *Dra*I and *Sph*I endonucleases. The plasmid pZERO-2.1 contains three *Sph*I sites and no *Dra*I sites; therefore, recombinants with inserts in the opposite orientation gave restriction fragment patterns that differed from the ones already selected. Recombinants with opposite orientation inserts were designated pKB1.3O, pKB2.1O, pKB4.5O and pKB5.0O, i.e. by adding "O" to the clone designation. The two larger fragments, 4500 bp and 5000 bp, each contained an internal *Hin*dIII site, discovered when the recombinants were digested with *Hin*dIII to confirm the presence of an insert. The 4500 bp fragment produced 3,500 bp and 1,000 bp fragments when digested with *Hin*dIII; the 5000 bp fragment produced 4000 bp and 900 bp fragments when digested with *Hin*dIII. Each of these smaller fragments were individually subcloned into *Hin*dIII digested pZERO-2.1. The subclones were named to include the fragment size, i.e. pKB3.5 contains the 3500 bp fragment, pKB1 contains the 1000 bp fragment, pKB900 contains the 900 bp fragment and pKB4.0 contains the 4000 bp fragment.

Sequencing strategy

The sequencing primers M13 forward and reverse (Table 1) were used to determine the sequence of the fragments from pHPM186. Fig. 6 shows the subclones described below that were used to determine the sequence of pHPM186. Sequence of both strands of each fragment was determined as well as the sequence through all restriction sites. Each sequencing reaction, forward or reverse, yielded 700-800 bp. The sequences of the inserts in pKB1.3 and pKB1.3O were determined completely with one set of reactions,

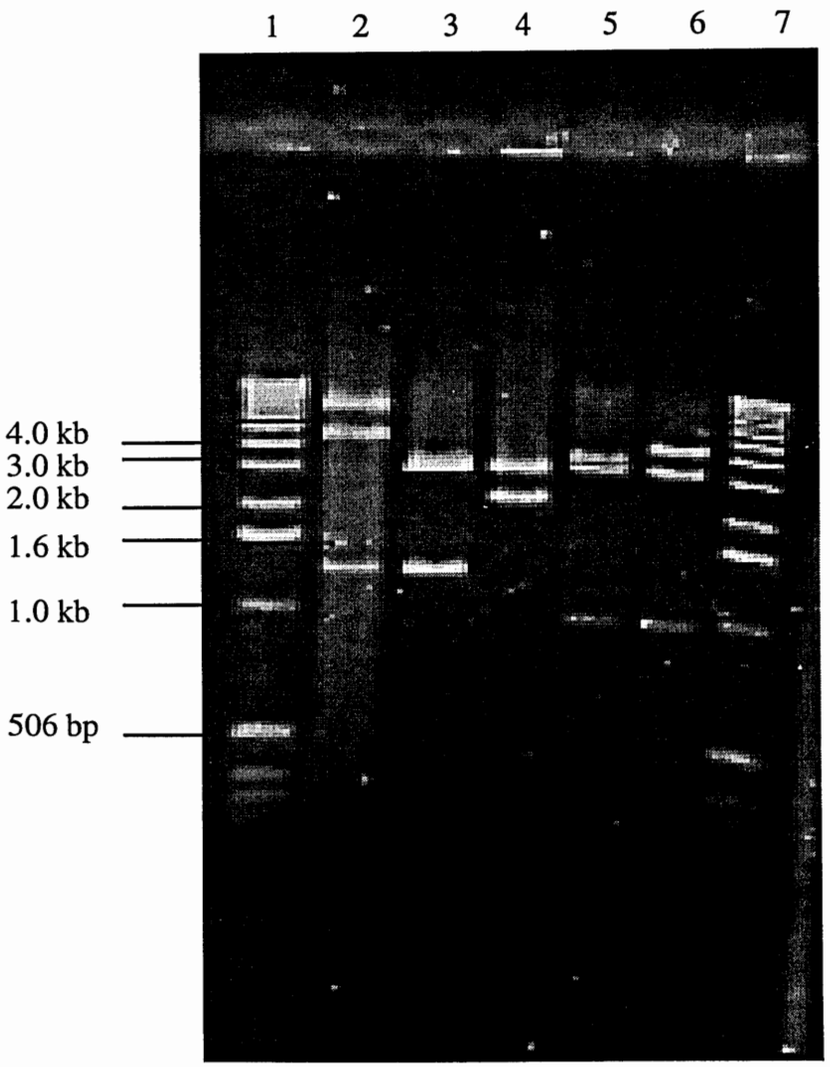

Fig. 5. Clones containing *Hin*dIII fragments of pHPM186. Lanes 1 and 7, 1 kb ladder marker; lane 2, pHPM186; lane 3, pKB1.3; lane 4, pKB2.1; lane 5, pKB4.5; lane 6, pKB5.0. Vector is 3 kb band.

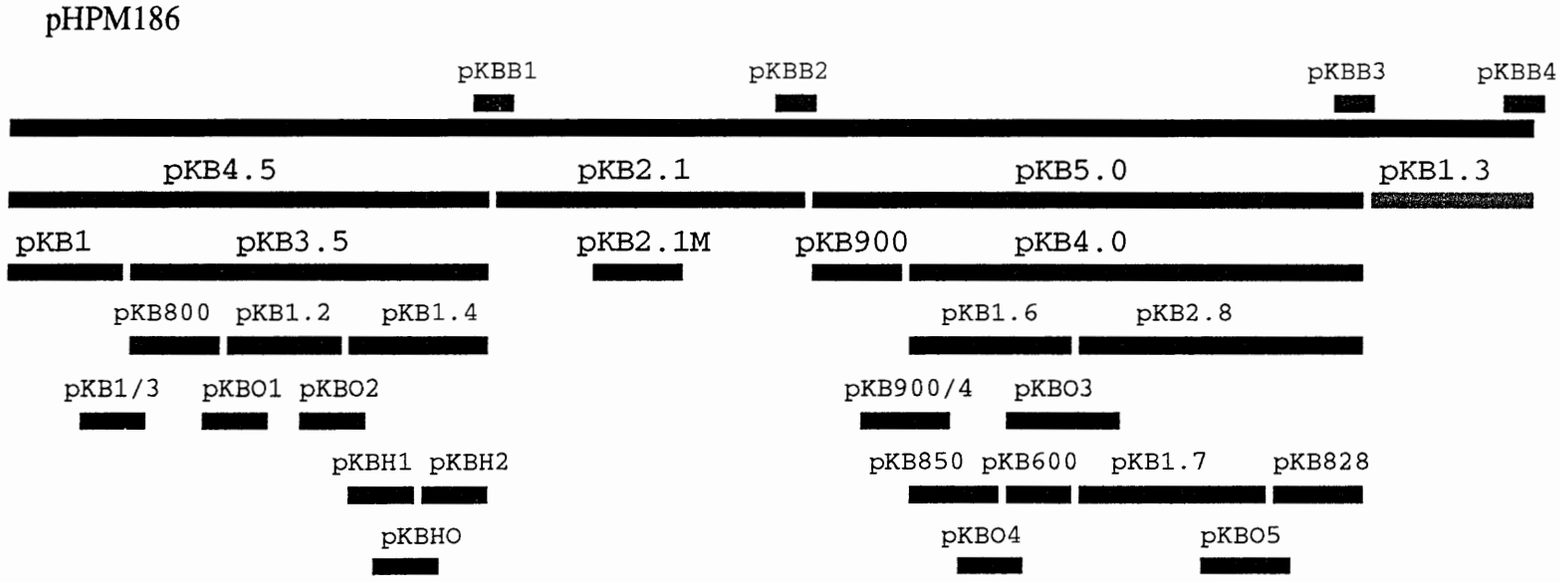

Fig. 6. Schematic representation of cloning strategy. The colored bars represent each insert sequenced. The red bars represented the 4.5kb *Hin*dIII fragment, the dark blue bars represent the 2.1kb *Hin*dIII fragment, the green bars represent the 5.0 kb *Hin*dIII fragment, and the pink bar represents the 1.3 kb *Hin*dIII fragment. The light blue bars represent the PCR amplified fragments created to the bridge the *Hin*dIII restriction sites.

forward and reverse. Sequences of the inserts in pKB2.1 and pKB2.1O were established in the forward and reverse directions, but a portion of sequence remained undetermined between the two ends. PCR primers KB 20 and KB 22 (Table 1) were designed from the sequence of the insert from pKB2.1 to bridge the gap. *Hin*dIII endonuclease sites were added to the primers to facilitate cloning. PCR amplification of pHPM186 using KB 20 and KB 22 resulted in an 800 bp fragment, which was ligated into the *Hin*dIII site of pZERO-2.1 as previously described. The 800 bp insert was generated by PCR, so to ensure that the sequence was correct, two recombinants were selected with inserts in both orientations and were designated pKB2.1M (1 and 2) and pKB2.1MO (1 and 2). The DNA sequences of these four inserts were determined using M13 forward and reverse.

For the 4500 bp fragment, the insert DNA from the two subclones described previously (pKB3.5, pKB1 shown in red in Fig. 6) was purified and the DNA sequences determined. All of the inserts also were subcloned in the opposite orientation, unless otherwise indicated, to facilitate sequencing of both DNA strands. The sequence of the insert in pKB1 was determined entirely using one set of M13 primers. Sequence of each end of the insert in pKB3.5 was established using M13 primers. The insert in pKB3.5 was further digested with the restriction endonucleases *Dra*I and *Hin*dIII, resulting in fragments of 1400 bp, 1000 bp and 800 bp. Two of these fragments had one staggered end (*Hin*dIII) and one blunt end (*Dra*I); the remaining fragment had only blunt ends, but it was unknown which fragment was which. Two ligation reactions were prepared, one with the three fragments and pZERO-2.1 digested with *Eco*RV and *Hin*dIII, and one with the three fragments and pZERO-2.1 digested with *Eco*RV. The 1400 bp and 800 bp fragments were ligated into the *Eco*RV/*Hin*dIII digested pZERO-2.1 producing pKB1.4 and pKB800, while the 1200 bp fragment was ligated into the *Eco*RV digested vector, producing pKB1.2. The sequence of the insert in pKB800 was determined completely on both strands using the M13 primers. To ensure that the three fragments from pKB3.5 were in the proper orientation, recombinants that overlapped each restriction site were obtained by digesting the fragment with *Apa*I (pKBO1, pKBO2). The sequences of the inserts in

pKB1.2 and pKB1.2O were determined completely with the M13 primers. The insert in pKB1.4 was further subcloned by digestion with *Hae*II, yielding pKBH1and pKBH2. Recombinant pKBHO was created by digesting pKB1.4 with *Kpn*I to ensure that the sequence from the inserts of pKBH1 and pKBH2 were complete.

The insert in pKB5.0 (shown in green in Fig. 6) also contained an internal *Hin*dIII site that resulted in fragments of 900 bp and 4000 bp, which were ligated individually into the *Hin*dIII site of pZERO-2.1, resulting in pKB900 and pKB4.0, respectively. The sequence of the insert in pKB900 was determined on both strands using the M13 primers. The insert in pKB4.0 was digested with *Hin*dIII and *Kpn*I endonucleases to produce fragments of 1600 bp and 2800 which were ligated unidirectionally into *Hin*dIII/*Kpn*I digested pZERO-2.1 (pKB1.6 and pKB2.8). To bridge the restriction site between the 1600 and 2800 bp fragments, the 4000 bp fragment was digested with *Sma*I and ligated into *Sma*I-digested pZERO-2.1, producing pKBO3. The insert from pKB1.6 was isolated and digested with *Dra*I, resulting in blunt-ended fragments of 850 bp and 600 bp. These fragments were subsequently ligated in both orientations into the blunt-ended *Eco*RV site of pZERO-2.1 (pKB850 and pKB600) and the sequences were determined. A recombinant (pKBO4) containing a small 300 bp fragment that overlaps the junction between pKB850 and pKB600 was also recovered by digesting pKB1.6 with *Sma*I, and ligating the fragment into *Eco*RV-digested pZERO-2.1. Determination of the sequence of the insert in pKBO4 ensured that the inserts in pKB600 and pKB850 represented the entire 1600 bp fragment. The insert from pKB2.8 was subcloned by digestion with *Xba*I to produce fragments of 1700 bp and 800 bp which were ligated into *Xba*I-digested pZERO-2.1 (pKB1.7 and pKB828). DNA sequence of the insert in pKB828 was determined completely using the M13 primers. Sequencing reactions of the insert in pKB1.7 yielded 1000 bp of sequence, so no further subcloning was required. A recombinant (pKBO5) containing a small 400 bp fragment that overlaps the junction between pKB1.7 and pKB828 was recovered by digesting pKB2.8 with *Hae*III, and ligating the fragment into *Eco*RV-digested pZERO-2.1. Determination of the sequence of the insert from pKB05

*Eco*RV-digested pZERO-2.1. Determination of the sequence of the insert from pKB05 ensured that the inserts in pKB1.7 and pKB828 represented the entire 2800 bp fragment.

The recombinants described above allowed the sequence determination for all four of the original *Hin*dIII fragments of pHPM186. To be certain that no small fragments were missed in the original digestion with *Hin*dIII, four recombinants carrying sequences that overlapped the ends of each of the original *Hin*dIII fragments were constructed. These overlapping fragments were generated by PCR amplification using primers listed in Table 1: KB 27 and KB 33 between pKB1.3/pKB4.5; KB 28 and KB 30 between pKB4.5/pKB2.1; KB 25 and KB 31 between pKB2.1/pKB5.0; and finally, KB 26 and KB 32 between pKB5.0/pKB1.3. The fragments were ligated into the *Eco*RV site of pZERO-2.1 creating pKBB4, pKBB1, pKBB2, and pKBB3, respectively (top line of Fig. 6). The sequence of each insert was determined and the results showed that no additional DNA fragments occurred in pHPM186.

Sequence analysis

Sequencing information from each DNA insert was combined to give the total of 12887 bp for pHPM186. The sequence is shown in Fig. 7. The 12887 bp sequence was analyzed using DNAsis and Gene Inspector and was compared to sequences in the GenBank (1) database using BLAST algorithms (2) through the National Center for Biotechnology Information (NCBI).

DNAsis analysis showed a G+C content of 36.8% and eight open reading frames in pHPM186. Six of the ORFs (ORF2, ORF3, ORF5, ORF6, ORF7, ORF8) were on one strand and two (ORF1and ORF4) on the opposite strand. The amino acid sequences of the eight ORFs are shown beneath the DNA sequence in Fig. 7. The consensus sequences from *E. coli* were used to identify putative promoter and putative ribosome binding sequences. The putative ribosome binding sequences and putative promoter sequences are found in Table 2 and shown in boxes on Fig. 7.

BLAST analysis showed identity to both *H. pylori* plasmid and chromosomal

```
   1 GTGAATTACCCAACCGCTAAAGCGATTGGGCTTTTTCTTGCTTCATCGCTCCATAACTAG
                376 ◄  L  S  Q  A  K  E  Q  K  M  A  G  Y  S  A
  61 CTAGATCCAATATGTTGCCATATTTAGAAATAACCCCATTAGAGGAAGCTCCACAAGCTT
 362   L  D  L  I  N  G  Y  K  S  I  V  G  N  S  S  A  G  C  A  K
 121 TGATGCGCTCATCAGTGTAATGAGCGTAATCAGTTCGGATAATCCCTACCCTACTTTTAT
 342   I  R  E  D  T  Y  H  A  Y  D  T  R  I  I  G  V  R  S  K  D
 181 CTATCTTTATTTTATGCCTGTCATCTAGCATGCCTAAAGCGTAGTTTCTAATATTGACGC
 322   I  K  I  K  H  R  D  D  L  M  G  L  A  Y  N  R  I  N  V  S
 241 TCGCATTGTAATCTCTGTGGTGTGTGATATTGCAATGAGGACAAGTGAATTTAGTGATGT
 302   A  N  Y  D  R  H  H  T  I  N  C  H  P  C  T  F  K  T  I  N
 301 TTTCATGTTTTTTGCCTGTATTGAACCCACAATAAGAACACAATTGAGAGCTAGGGAAAA
 282   E  H  K  K  G  T  N  F  G  C  Y  S  C  L  Q  S  S  P  F  F
 361 ATCTGTCTATGCCTAATAGGGTTTTAGCTTTTCTTTGGGCTTTGTATTTTAGCATAGTAA
 262   R  D  I  G  L  L  T  K  A  K  R  Q  A  K  Y  K  L  M  T  L
 421 GGAATTTCCCCCAACTCGCATTAGCAAGGCTTTTAGAATGATAGGTTCTCATAAGCCCCT
 242   F  K  G  W  S  A  N  A  L  S  K  S  H  Y  T  R  M  L  G  K
 481 TAACATTCAAAGTTTCTACCCCTATCAAATCGTATTGATTGGTTATCTCATTACTGATTT
 222   V  N  L  T  E  V  G  I  L  D  Y  Q  N  T  I  E  N  S  I  K
 541 TATGCAAGTAGTCATCTCTAGTGTTTGAACAAGCTAGATGCAATCTAGCTACCTTTTTAG
 202   H  L  Y  D  D  R  T  N  S  C  A  L  H  L  R  A  V  K  K  A
 601 CTTGTTTTTTCCTGTTGTTGGAGCCTTTTACTTTTTTACTTAACCTCCTTTGCGCTTTAG
 182   Q  K  K  R  N  N  S  G  K  V  K  K  S  L  R  R  Q  A  K  T
 661 TGAGTTTCTTTTCTAATTTTTGATAAAAACGGATGTGTGGGTATTCTATTTTATCGCTTG
 162   L  K  K  E  L  K  Q  Y  F  R  I  H  P  Y  E  I  K  D  S  T
 721 TAACAATGAGCGTTCTTAAGCCCATGTCTAAACCTACCGCTTTTTTAATGATGGTAGGTT
 142   V  I  L  T  R  L  G  M  D  L  G  V  A  K  K  I  I  T  P  K
 781 TAGGGATAGGCTCTTTGGTTTCATAGGATATAGAACAAAAATATTGATCGGCTATGCAAG
 122   P  I  P  E  K  T  E  Y  S  I  S  C  F  Y  Q  D  A  I  C  S
 841 AAATAAAAGCCTGTTTGATAACGCTATCTTTAGGCAATTCTCTGTGTAATTTAGCCTTAA
 102   I  F  A  Q  K  I  V  S  D  K  P  L  E  R  H  L  K  A  K  I
 901 TGCCCTCTTTGAATTTAGGGAGAGCGATGGTTTGAGTCTCTGTTTTGATTTCTATGTTTT
  82   G  E  K  F  K  P  L  A  I  T  Q  T  E  T  K  I  E  I  N  Q
 961 GGGGGATTGCAAAAGATTGTTTAGCGTTTTTTTTAGATTTGAATTTAGGGTATCTCGCTC
  62   P  I  A  F  S  Q  K  A  N  K  K  S  K  F  K  P  Y  R  A  R
1021 TTTTGCTAAAGAAATTATCATAAGCACTCACAAGCTGTCTTAACGCCATTTGCAAGCTTT
  42   K  S  F  F  N  D  Y  A  S  V  L  Q  R  L  A  M  Q  L  S  Q
```

```
1081 GAGAATTGCATTCATTGAGGTAATGGTATTTTTCTTGGTGCTTGATTTGGGTTAAGACTT
  22   S  N  C  E  N  L  Y  H  Y  K  E  Q  H  K  I  Q  T  L  V  K

1141 TTTGCATGGTGAAGTAAGTTTCTTTAATGCCTTTTGCGTATTGCTTTTGTCGGTAATCTA
       Q  M     ◄ 1

1201 AAAAGTAGTTATACACGACCCTAGAACAGCCAAAATGTTTAGAAATAAGCTCTTTTTGTT
1261 GAGCGTTAGGATAAATTCTAAACTTGATAGCGTTAAGCAAAAATAACTCCTTATCTTTAT
1321 CTTGTATTATTATACAATGTTTTTATAAATAATGAATAAGGATAGAGGAATGAGAAAAAA
                              1 ►      M  N  K  D  R  G  M  R  K  N
1381 CCATTACCCATTAAGGGGGTATATTTCAACCAACCGAAGCAAGCATAATCTAAAAGCCCA
  11   H  Y  P  L  R  G  Y  I  S  T  N  R  S  K  H  N  L  K  A  H
1441 TTTGATTTTAGTGTGTAAGTATAGAAAAAAGTTGTTGCAAGGGGATTTGAATAACTTTAT
  31   L  I  L  V  C  K  Y  R  K  K  L  L  Q  G  D  L  N  N  F  I
1501 TAAGTCTGTTATAGATGAGATAGCCACCCAAAGTCATTTCATTATCATTGCGATGGAAAG
  51   K  S  V  I  D  E  I  A  T  Q  S  H  F  I  I  I  A  M  E  S
1561 CGATATAGATCATTTGCACTTAATGGTTCAATATATCCCTAGAATGTCTATTAGTTCCAT
  71   D  I  D  H  L  H  L  M  V  Q  Y  I  P  R  M  S  I  S  S  I
1621 TATTTCTAAAATCAAACAAATCACCACTTATAGAGTTTGGCACGATAAAAGATTTATCCC
  91   I  S  K  I  K  Q  I  T  T  Y  R  V  W  H  D  K  R  F  I  P
1681 CTTATTGCAAAAACACTTTTGGAAAGAAAAAACTTTTTGGACTGATGGATTTTTTGTTTG
 111   L  L  Q  K  H  F  W  K  E  K  T  F  W  T  D  G  F  F  V  C
1741 CTCTATCGGTCAAGCTAATCCTGAAACGATCAAGGCGTATATAGAAAATCAAGGTTAGTT
 131   S  I  G  Q  A  N  P  E  T  I  K  A  Y  I  E  N  Q  G     ►
1801 TCACTCATAAGGTTTTTATATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACACTAA
1861 AGATATTTGGGATTTCTGCTTGGGTGGTTTAAGACCACGCTTTAGTTGGAACATACAACG
1921 GCTGTAGAGAGTGCCACATTAGGCCTGATGTTTTGCTTGTGTATAGAGTGAAAGGCAATA
1981 TTTTGGCTTTGGTTAGGCTTGGCAGTCATAGCGAGTTGTTTTAGACAAATTTCTTTCTAA
2041 AATGGCCTAGAATCAATTTTCAATTCTAATCCAACCTAATCATCAGTTTGCGAATGAAAG
2101 GATTTTTTCGCTCATATTTTCGTTTTAACCGTTTTTTCGCTTGTAGTTAATTCATTAATC
2161 CGTTTTTTCCTTTCCCTTGATTGTAGGCGATGTGGTCAACGAAAATAGTCCCATTGTGCC
2221 AACTGATAACCATCAAAGGTTTTTAGCTCAAGAAAACCTAGCTTTCTATAGTGGCATTTA
2281 GCGTTCTTTTAGCGATTTGGTTGTGGCCTAGCGGGCGTTAGCGGGATCAAGGGTTTGCCC
2341 TATTGTTTGCTGTCAAGCAAACACCAACAAGCGAAGCATGGACGAAAGGGCATGGCAGTT
2401 TGAAAGCGTAGGCGTTAGCCGAAGCTGTTTTGCGTTAGCAAATCAAACAAGATAGCGGAA
2461 ACCTGCGGTTAGCCGAAAAAATCCCTAAAAACTAAACCTCAAAATATGTAGGCGTCATGC
2521 GCGTTGTTTTTAATTACATTTTTTAACAACTATGTTGTTTTTACATGTTTTTACAACGGG
2581 GGCGCATGTGAGGGATTGGGGGTTGAAACCCCCTAAATAACGAAGCTGTAGGGTTTCTCA
2641 TTTTTGGGTAAAAATGAATATAACAGAACTTCTTGCCAACGATAACAGAACTTCTTGCCA
```

```
2701 ACGATAACAGAACTTCTTGCCAACGATAACAGAACTTCTTGCCAACGATAACAGAACTTA
2761 AAAAGTTAATAGTTATAATATTAACAATTTTTAGCTATAATAACAACCTAAGGCGGTATA
2821 ACATGGTTGCTAATCCTAGTGTTAACAAATTTGGAGCAATAAGCTTTAAAAGCTAGTTTG
2881 TTGGGAGTTTGTAGTGGTTTCACTCCGTTAGAGCCACACCATGAAAGCATTTTTTATAGT
2941 AGTTATTTTAGTTGTAATCTTAACACAGCCACTATATTAAAACCTTATCTTTTTAATAAC
3001 CCTTATAATTTCTCCAAGACTTTTTAATTGTTTCACTCCTATCATCATATCGACTTTTTA
3061 AAAATAAGCATTAAAACCCTCTCAAACCCCCATGAATACCAATTTTTAACAGCTTAGAAA
3121 ACAAGAATAGGAATTACCAAAATTATTAGAAGAATCAGACACCCCCCCACAAACCCCACA
3181 AATTAAACTACAAAAACAAGAAACACAAACTGACAGAGACAAGACAACTCCAAGCACCCC
3241 CCACACCCCCCAGCACAAAGTCAAAGAAAGCACAGAAAAACCGACGCTACTAAGGTCACT
3301 CTAGGGGGTGTGGTACTCACGCCCGTCCTGTTGTCGCGATTTGTCAGAGATTGAGACTCA
3361 AGGGACACCCTCATTCGTTTGACTGCTATAAGTGAAACGCAGGAGTAACATAGGAGGTTC
3421 TAATGAGCGCTATCAGAAGTGGTATTAAGCTGTGGGATAGCGTCAAGGCCGCTTTACTTT
   1 ►M  S  A  I  R  S  G  I  K  L  W  D  S  V  K  A  A  L  L
3481 GCGGGATTAGCGAGACAGAAACTTCAATCATTCAAGCAGATTACATGCTTTTTAGTCGGT
  20 C  G  I  S  E  T  E  T  S  I  I  Q  A  D  Y  M  L  F  S  R
3541 GTAAAATTGAATTGAACAAACCGAGTAAAGATTTGAAGTATTTAGAAATCCAACTCAATG
  40 C  K  I  E  L  N  K  P  S  K  D  L  K  Y  L  E  I  Q  L  N
3601 ATAACTACCAATACTTGCTCAACAATCTAGGCATGGGGCAATACACTTCTTTCAATCTGT
  60 D  N  Y  Q  Y  L  L  N  N  L  G  M  G  Q  Y  T  S  F  N  L
3661 TAGAATTTCAAAGAGTGAGAGGTAAATACGCTAAAACGCTCTATCGCTTGCTCAGGCAAT
  80 L  E  F  Q  R  V  R  G  K  Y  A  K  T  L  Y  R  L  L  R  Q
3721 CCAAAAGCACAGGGATTTTGAGCGTGGAATGGACTCAATTCAGGGAACTTTTAGACATTC
 100 S  K  S  T  G  I  L  S  V  E  W  T  Q  F  R  E  L  L  D  I
3781 CAAAAGATTACAAAATGGAAAACATCGATCAAAAAGTCTTAACCCCAAGCCTCAAAGAAC
 120 P  K  D  Y  K  M  E  N  I  D  Q  K  V  L  T  P  S  L  K  E
3841 TCAGAAAAATCTATCCTTTTGAACACTTGAGCTACAAGAAAGAACGCCGAAACAGCCATG
 140 L  R  K  I  Y  P  F  E  H  L  S  Y  K  K  E  R  R  N  S  H
3901 ACAAACGCAAAGTAACGCATATTGATTTCTATTTTGAACAATTGCCACAGAGCGAAACCA
 160 D  K  R  K  V  T  H  I  D  F  Y  F  E  Q  L  P  Q  S  E  T
3961 AGCATCAAATTCAAAAAGACAAGCAACGCGCTAAAAGGGACATCAAGCTCATAGCATGGG
 180 K  H  Q  I  Q  K  D  K  Q  R  A  K  R  D  I  K  L  I  A  W
4021 ATATTCACAAACTCAAAGCTATCAAAATTGAAGCATACAAGAATAAAACCAAATCACTAA
 200 D  I  H  K  L  K  A  I  K  I  E  A  Y  K  N  K  T  K  S  L
4081 AAGTTTCGCAAAATCCCTTATGGAAGCTAGGTTTCTTGAATTCAAAACCTTGGTCGGCTA
 220 K  V  S  Q  N  P  L  W  K  L  G  F  L  N  S  K  P  W  S  A
```

```
4141 TCAGTTCAGGCACAACAATGGGACTATTTTGCAAATTAACAACGCCACTTTTGAAAAGAA
 240 I  S  S  G  T  T  M  G  L  F  C  K  L  T  T  P  L  L  K  R
4201 TCAAATGTTTTTGCATGTTTCAACCAACAAAAACTCTCAAAAATTCCTTGTGTCCAACAA
 260 I  K  C  F  C  M  F  Q  P  T  K  T  L  K  N  S  L  C  P  T
4261 GACATTCGCTTTGGAACTTCTGTTTGAGAATGGATACAGCCGGAAAAAAGACAATTTGCG
 280 R  H  S  L  W  N  F  C  L  R  M  D  T  A  G  K  K  T  I  C
4321 AGAAGAAATTGATCCCCCTAAAATCCACCCTATCACTAACGAATCTATTAAGGAATTTGC
 300 E  K  K  L  I  P  L  K  S  T  L  S  L  T  N  L  L  R  N  L
4381 AGAATATATCGGCAAAACGATCCACATCACTAATTTCAATGTGGATCAATGCCCTGAGGG
 320 O  N  I  S  A  K  R  S  T  S  L  I  S  M  W  I  N  A  L  R
4441 ATCAATACCTACCTGAAAATCACTCGGATTGTGAAATTGAATTACAATCGGATCTGTGTT
 340 D  Q  Y  L  P  E  N  H  S  D  C  E  I  E  L  Q  S  D  L  C
4501 TCAGTCCAAGATGTGGATACTCCTGCGAAACTGCTCAAACCTTTCATTGCTACAGATGCA
 360 F  S  P  R  C  G  Y  S  C  E  T  A  Q  T  F  H  C  Y  R  C
4561 AAACATTTGAATATTTGGTTCAAGAAACACTATTGGTGAAAAAGACAGATGCATTTGACT
 380 K  T  F  E  Y  L  V  Q  E  T  L  L  V  K  K  T  D  A  F  D
4621 ATATTTATATACTAGGACAAGTGATGGATTATAAAGAAATTGCTAGAATTCGAAGATTAC
 400 Y  I  Y  I  L  G  Q  V  M  D  Y  K  E  I  A  R  I  R  R  L
4681 GCTATGGATCTGACCATTCGCATGGCACACACCACAACTCCGCTATGGAGGGAACAAACT
 420 R  Y  G  S  D  H  S  H  G  T  H  H  N  S  A  M  E  G  T  N
4741 CACTTTGGGCAGAGCATGAGTATGTTAATAGAACGAAAGAGCCCTATGAAGTGGGTGAGC
 440 S  L  W  A  E  H  E  Y  V  N  R  T  K  E  P  Y  E  V  G  E
4801 TTAGAGGAAGTGCACGAGATAGAAAATTATGGCAATTTTGTTCCCCTTATTTTAGAATTT
 460 L  R  G  S  A  R  D  R  K  L  W  Q  F  C  S  P  Y  F  R  I
4861 TTAAAGAGAGAAAAGATACTAGAAGAACCTTTGATTGGCCATTTGCATTCTGTTTTGATG
 480 F  K  E  R  K  D  T  R  R  T  F  D  W  P  F  A  F  C  F  D
4921 CGCAATATCCTCGCTCATTTTGGAAAATTCAAAACCACTTACAAAGAGATCATAGGGGCA
 500 A  Q  Y  P  R  S  F  W  K  I  Q  N  H  L  Q  R  D  H  R  G
4981 AAAAAACCCAGCTCGAAGTCCACATAATGGTGCAACTAGAATCAATAATGTGTGTTTGAA
 520 K  K  T  Q  L  E  V  H  I
5041 AATACCAGATGAAGTGTCATTGAATTTGAGCCCTGAAGAAAAAATAAAAAAGATAGCAGA
5101 AGCCCACATAGAATTTGAAGAGATCCCCCTTTTAGCCATGGCAATGGTTGCACCCGAAGA
5161 GTCTTGATGTTCTACGACAGCATAAAAGCTAATTTAAGGCCCTTTGTGATTGAAGTAAGC
5221 ACTAGAAGTCAATACATGCATGCTATGAGGGAAACCGAGCCTAAACCTCTAGTTGCGCTG
5281 CCATTTTGAAGAAAGGCGATCGGCAAATTTTTCTGTCAATAAATCATTGAGCACTTTGAA
5341 AAATTGGGTCAAGAAACACTACAAGTGAAAAGACAGATGCATGCTAACTAAACTCATAGA
5401 AAAACTGAATCACGAAAGAAAGAATGCTATCAAAAATGGCATTTACCACCTGATCCAAAT
```

```
5461 CAAATTTTCTTACAACTCTAATCGCATTGAAGGAAGCGGTTTGACCTACGAACAAACCGC
5521 TCATATTTTTGACAAATTCGTTCTCATAACTGAAAAAAACGCCAATATCAAACTTGATGA
5581 TATTTTTGAAACTATCAATCATTTTGAATCGTGAATCACTTGCTTTAAATCATCATATAA
5641 AAAACATTTTAGTTTAAAATACTTTAGAACTTTACACAAAATCTTGAAAAAAAATTGTTC
5701 TGATGAAGTTATTGGTAGTTTCAAAAAACTATCGAATTTTGTAGACAATAGCGCCACAAC
5761 AAGACCCAAATTAGTTGAAAGCGAATTGACAAATCTTGTGAAAAATTACCAACGCAACCT
5821 TGAAGTGAGTTTGGAAAACATCATAGATTTTCATGTGGCTTTTGAAGAGATCCACCCTTT
5881 TAGCGATGTTAATGGTAGGGCATCAAAGAATATGACAATACAAAAGGTTATTTGAAAGAC
5941 ACGATTTTGCAAAGTCAAGACAATTTCAATGAAATGGTTAGCTATTTCTTTTGCGAATGA
6001 TTTCGCGCATATTTTCGTTTTGAGACTTTTTTATCTTGTAGGTAGTATCATTAGTCGCTT
6061 TTTTGCTTTCGTTGGTTGTAGGCGAAGTTAGATAGCAAATCGTGGGATTGAATATAGTGA
6121 ATTACCCAACCGCTAAAGCGATTGGGCTTTTTCTTGCTTCATCGCTCCATAACTAGCTAG
          376 ◀  L  S  Q  A  K  E  Q  K  M  A  G  Y  S  A  L
6181 ATCCAATATGTTGCCATATTTAGAAATAACCCCATTAGAGGAAGCTCCACAAGCTTTGAT
 361  D  L  I  N  G  Y  K  S  I  V  G  N  S  S  A  G  C  A  K  I
6241 GCGCTCATCAGTGTAATGAGCGTAATCAGTTCGGATAATCCCTACCCTACTTTTATCTAT
 341  R  E  D  T  Y  H  A  Y  S  T  R  I  I  G  V  R  S  K  D  I
6301 CTTTATTTTATGCCTGTCATCTAGCATGCCTAAAGCGTAGTTTCTAATATTGACGCTCGC
 321  K  I  K  H  R  D  D  L  M  G  L  A  Y  N  R  I  N  V  S  A
6361 ATTGTAATCTCTGTGGTGTGTGATATTGCAATGAGGACAAGTGAATTTAGTGATGTTTTC
 301  N  Y  D  R  H  H  T  I  N  C  H  P  C  T  F  K  T  I  N  E
6421 ATGTTTTTTGCCTGTATTGAACCCACAATAAGAACACAATTGAGAGCTAGGGAAAAATCT
 281  H  K  K  G  T  N  F  G  C  Y  S  C  L  Q  S  S  P  F  F  R
6481 GTCTATGCCTAATAGGGTTTTAGCTTTTCTTTGGGCTTTGTATTTTAGCATAGTAAGGAA
 261  D  I  G  L  L  T  K  A  K  R  Q  A  K  Y  K  L  M  T  L  F
6541 TTTCCCCCAACTCGCATTAGCAAGGCTTTTAGAATGATAGGTTCTCATAAGCCCCTTAAC
 241  K  G  W  S  A  N  A  L  S  K  S  H  Y  T  R  M  L  G  K  V
6601 ATTCAAAGTTTCTACCCCTATCAAATCGTATTGATTGGTTATCTCATTACTGATTTTATG
 221  N  L  T  E  V  G  I  L  D  Y  Q  N  T  I  E  N  S  I  K  H
6661 CAAGTAGTCATCTCTAGTGTTTGAACAAGCTAGATGCAATCTAGCTACCTTTTTAGCTTG
 201  L  Y  D  D  R  T  N  S  C  A  L  H  L  R  A  V  K  K  A  Q
6721 TTTTTTCCTGTTGTTGGAGCCTTTTACTTTTTTACTTAACCTCCTTTGCGCTTTAGTGAG
 181  K  K  R  N  N  S  G  K  V  K  K  S  L  R  R  O  A  K  T  L
6781 TTTCTTTTCTAATTTTTGATAAAAACGGATGTGTGGGTATTCTATTTTATCGCTTGTAAC
 161  K  K  E  L  K  Q  Y  F  R  I  H  P  Y  E  I  K  D  S  T  V
6841 AATGAGCGTTCTTAAGCCCATGTCTAAACCTACCGCTTTTTTAATGATGGTAGGTTTAGG
 141  I  L  T  R  L  G  M  D  L  G  V  A  K  K  I  I  T  P  K  P
```

```
6901 GATAGGCTCTTTGGTTTCATAGGATATAGAACAAAAATATTGATCGGCTATGCAAGAAAT
 121  I  P  E  K  T  E  Y  S  I  S  C  F  Y  Q  D  A  I  C  S  I
6961 AAAAGCCTGTTTGATAACGCTATCTTTAGGCAATTCTCTGTGTAATTTAGCCTTAATGCC
 101  F  A  Q  K  I  V  S  D  K  P  L  E  R  H  L  K  A  K  I  G
7021 CTCTTTGAATTTAGGGAGAGCGATGGTTTGAGTCTCTGTTTTGATTTCTATGTTTTGGGG
  81  E  K  F  K  P  L  A  I  T  O  T  E  T  K  I  E  I  N  Q  P
7081 GATTGCAAAAGATTGTTTAGCGTTTTTTTTAGATTTGAATTTAGGGTATCTCGCTCTTTT
  61  I  A  F  S  Q  K  A  N  K  K  S  K  F  K  P  Y  R  A  R  K
7141 GCTAAAGAAATTATCATAAGCACTCACAAGCTGTCTTAACGCCATTTGCAAGCTTTGAGA
  41  S  F  F  N  D  Y  A  S  V  L  Q  R  L  A  M  Q  L  S  Q  S
7201 ATTGCATTCATTGAGGTAATGGTATTTTTCTTGGTGCTTGATTTGGGTTAAGACTTTTTG
  21  N  C  E  N  L  Y  H  Y  K  E  Q  H  K  I  Q  T  L  V  K  Q

7261 CATGGTGAAGTAAGTTTCTTTAATGCCTTTTGCGTATTGCTTTTGTCGGTAATCTAAAAA
     M      ◄  1
7321 GTAGTTATACACGACCCTAGAACAGCCAAAATGTTTAGAAATAAGCTCTTTTTGTTGAGC
7381 GTTAGGATAAATTCTAAACTTGATAGCGTTAAGCAAAAATAACTCCTTATCTTTATCTTG
7441 TATTATTATACAATGTTTTTATAAATAATGAATAAGGATAGAGGAATGAGAAAAAACCAT
                          1 ►    M  N  K  D  R  G  M  R  K  N  H
7501 TACCCATTAAGGGGGTATATTTCAACCAACCGAAGCAAGCATAATCTAAAAGCCCATTTG
  12  Y  P  L  R  G  Y  I  S  T  N  R  S  K  H  N  L  K  A  H  L
7561 ATTTTAGTGTGTAAGTATAGAAAAAAGTTGTTGCAAGGGGATTTGAATAACTTTATTAAG
  32  I  L  V  C  K  Y  R  K  K  L  L  Q  G  D  L  N  N  F  I  K
7621 TCTGTTATAGATGAGATAGCCACCCAAAGTCATTTCATTATCATTGCGATGGAAAGCGAT
  52  S  V  I  D  E  I  A  T  Q  S  H  F  I  I  I  A  M  E  S  D
7681 ATAGATCATTTGCACTTAATGGTTCAATATATCCCTAGAATGTCTATTAGTTCCATTATT
  72  I  D  H  L  H  L  M  V  Q  V  I  P  R  M  S  I  S  S  I  I
7741 TCTAAAATCAAACAAATCACCACTTATAGAGTTTGGCACGATAAAAGATTTATCCCCTTA
  92  S  K  I  K  O  I  T  T  Y  R  V  W  H  D  K  R  F  I  P  L
7801 TTGCAAAAACACTTTTGGAAAGAAAAAACTTTTTGGACTGATGGATTTTTTGTTTGCTCT
 112  L  Q  K  H  F  W  K  E  K  T  F  W  T  D  G  F  F  V  C  S
7861 ATCGGTCAAGCTAATCCTGAAACGATCAAGGCGTATATAGAAAATCAAGGTTAGTTTCAC
 132  I  G  O  A  N  P  E  T  I  K  A  Y  I  E  N  O  G
7921 TCATAAGGTTTTTATATAGTTCCTAGCGGAACTAAAGCATTCATCCCAAACACTAAAGAT
7981 ATTTGGGATTTCTGCTTGGGTGGTTTATCGTGTCGATCGCGATATTTTTGGGCGTAGCTG
8041 ACCCCCGCTAAAAAGTGCTAGCTAGCCCTAGCTTTAGTCGAAATTTCGGGGAGCGCTGTA
8101 ATATATATATTTTTCTCTCGATGGTGTCATATAGCTGATCGATCTAGAAATCACTTTTAC
8161 CAAATAAATCCTAGTTAATACGTCGATGGATCTCGATTTAAATCGGGATCAGCTAGTATA
```

```
 8221 AAAATCCATCCTAGAGTATATAGGCACTATCAAATAGCAATATGTTCATACTTAGCAACA
 8281 GTGAACTAGCAAGTCCAGTCAAAAAAATCATGGTAATAGTCATATATCTGTTGTACGAGT
 8341 TGTCGTTACAATAATGAACGAGAATTCCTATAATTTGAAATGATATATAAAATATAAACA
 8401 AGTAGTTTTCTGTAATTTGTTTGTCTAGCAAGTAGGCTTGCCATAATTGAAAATGACTTT
 8461 GAGAAAAGATTTGTAAAATTAGACTGAGTAAAATTAGTTTTTTAAGAAGTGGTTTATCTT
 8521 TTAGCTCTCTAAAACTTTCCCTTACATGTTGTTTTAACATTTGTAACTGTTGCTTTTTTT
 8581 CTAGTTTATGAGAAGAATCATCTTTAAAAAATAAAATAATGACCAAGACACACAGTAATG
 8641 CAAGCAAGACTGATATGTAGTACATGGAGACACCATATTTTAAATACAAGAATGAGCCTG
 8701 AGCAACTTCCAATTATCAACCCTATAAAACTGATTTGGCGACTTTTAGCAATGAATGGTG
 8761 ATAAATCTTCTTTTTGGTCTTTGATGTTGGTAATGAGGTAAAATAGCGACTTTCGAATAC
 8821 GGGGGCGAATTTCGCGCGCCCCTCTATCGACTGGCATAGCTCAATCAGATGTCCGCATAC
 8881 GTAGTGATGGGATCTCTGCCGATCACCACGCTCACTAGCCGGCCGTTACCTAGTGGAGTC
 8941 CGACGCCTCGGTATCCTGAAGTCTATGAAGCCGTTTGAATAAACCCACTAAAGTAGCCCC
 9001 ACTAGCTAAGAGATCCACTAGCTAAACTAAATACTCGTATCTAAACCACTACGAAGAGCC
 9061 AAAATAAAAATAGCAGCCCTAGAAGCTAACTTTTAGAAAGCAGACTAAAAAGAGCAGTGC
 9121 TGTATACTACTTTAGTTAAAGTGCTAGCTAAACACCAGCCACCTCGGGTAGTTCTTGAAG
 9181 AGATAAAAGCATCACCAACAGCTGTATGGTGGCTTGTAGTGTACTAAACGTGCTGCATCC
 9241 CTTTAGCAGTGCAGCTGCAGCTGCAGATGAGACGTGCATTTTAAGCATGCAGCTCGAGCT
 9301 AACACCCTAAAATGCTTTCAGTGCAGTCAGTGCAGTGCAGCTTCTAGAGCAAAACTATTA
 9361 TGACAAAAAGGACAACAAGACAATGAAGGGATATAAAAAGGACCAATAAGAGATATGTCA
 9421 TTTAAATAGCCAATATTTAAAAATGGAATGCGACACTTACAAAAAAATCTTGTTGCCAAT
 9481 TGCACCGTGTTTGGATTATCTCCAGAAACAATTCCAAACAAATTTTCTTTTTTATGTTTA
 9541 GAAAAAATTTCTTCAAGATTGTCTTCACTAGCAAAATCATCTACTGTCTCAATATTGATA
 9601 CTAAATCTACTAGAAAGAGCCTGCTTTATGGCAGATGTTTTATACTGTGAAATGTAATTC
 9661 TTGTCAAACAAGAACTGGCGGTTAAGATTACTGGAATCAACAGTATCTTTATCTAACAAT
 9721 ATTAGTTTTTTTGGATAAAAACTTGCTAAAGCATAGCTTACAAAATTTCCAATACCCCCA
 9781 CAACCAATAATGATGAAAATAGTTCTTTTAAAATTATCTATTGTTAGTTCTGGTTTACTT
 9841 GACACTAAATCGATAAATAGGTGGTTCTTAAAGTTTTGTTCGTCATTTGGATTTAGTATG
 9901 AATGAGGTAATAAGTTTATTGGCAACTAGCTTGTCAAAAATATGATGTTCTATATTGCTA
 9961 GTTTGCATAAAACTCGGCAGCTCAGCATCGGCGCATCTCAGAGAACAGAGGTCGAGATTC
10021 TCAATACATAGACAGCGCAAACGAAGCACAGCACATAGAACGCGTACGCGCACACGAGCC
10081 CACACAGTGAGACACGCCACGCACTCCTCGCACGCCTCGCGAGCCTGACACTCCGACCGC
10141 GCGTGATGATATGACATGGGATGTGATTCACTATGCTTCATACATGTTTAACATGTATGA
10201 TTACGTATGGCTTTTAATATTATTTCTTTTAAACTTTCCATGCATTTACGTACGTAGCAT
10261 GTACATGACTAGGGTAAACCATACACAAATTAACGTAGCAATTAAGTCAGTAGATCATAT
10321 AGACTAGCTTAGTATTTAGGAGCATGACTGTACGTTTAATGAGCAAAGACAGAGATTATA
                    1 ►  M  S  K  D  R  D  Y
```

```
10381 AGCCTGTGTTCAACCCTAAATCTAAAGAAGTCTTAGAGAAAATGAACGCTGAAAAAAGAG
    8 K  P  V  F  N  P  K  S  K  E  V  L  E  K  M  N  A  E  K  R
10441 CGAGTTTAGTGGATGAGAGGGAAGAGGTAAAAGAGCAAGCCAACCAAGAAGCGTATAGAC
   28 A  S  L  V  D  E  R  E  E  V  K  E  Q  A  N  Q  E  A  Y  R
10501 CAATGAAAAAAGCGGCTAATGATGATTATGACATGGGGATGTGAGTTCCACTAATATTTT
   48 P  M  K  K  A  A  N  D  D  Y  D  M  G  M
10561 CAATAACATGTTTTTAAATAATGTTATTGGATTTAGCGTAATGCCTTTTAATATTCTTTT
10621 TCTCTAAACTTTTCCTTTTTAAAAACACACTAAAAAACCACCACTAATAAAATTATACTA
10681 ACCATTTCAAAAAGAACTAACAGATAACATAACTATTTGACAATAACATAGTATTAAGTT
10741 TTTCTATTTTATATACTTTTAGCAGGTTTATCTTTAAAATATGTTTAACAAAATGATAAA
                                                    1  ►  M  I  K
10801 GGAAAAAGCGATGATAATCACCATAGCCAATGAAAAAGGAGGGAGCGGTAAATCCACTTT
    4   E  K  A  M  I  I  T  I  A  N  E  K  G  G  S  G  K  S  T  L
10861 GTGTTTGAACTTGTGCGTTCAATTATTGTTAGACAAGAAAGATATTGCTGCATTAGACAC
   24   C  L  N  L  C  V  Q  L  L  L  D  K  K  D  I  A  A  L  D  T
10921 TGATAGCCAAAAGAGTTTGGAAGTGTTTAATAACATTAGGAGTGAAACGAGTTTGCCCAA
   44   D  S  Q  K  S  L  E  V  F  N  N  I  R  S  E  T  S  L  P  N
10981 TTTCACGCTCTTTAATCGCACAGGCAATATCACAGACACCTTAAAACAGATGATGGATAA
   64   F  T  L  F  N  R  T  G  N  I  T  D  T  L  K  Q  M  M  D  K
11041 ATATGAATACATCCTTATTGATACTAAAGGCGAACATTCCAAAGAAAGCCAAAGGGCTAT
   84   Y  E  Y  I  L  I  D  T  K  G  E  H  S  K  E  S  Q  R  A  M

11101 GCTATTGAGCGATTGGGTGCTAATACCCACCACGCCAAGCCAACTAGACACAGCGGTGCT
  104   L  L  S  D  W  V  L  I  P  T  T  P  S  Q  L  D  T  A  V  L
11161 ATTAGACATGCTAGAAAGGATTAGAGACATCCAAGCGTTGAATGAAAACTTGAAAGCTTG
  124   L  D  M  L  E  R  I  R  D  I  Q  A  L  N  E  N  L  K  A  C
11221 TATTGTGATGAACCGTATCCCTACTATCCCCACTCTTAAAGAGAAAAAAGCTCTCATTGA
  144   I  V  M  N  R  I  P  T  I  P  T  L  K  E  K  K  A  L  I  D
11281 TTTTATCAACCAAAATAACGCTAATGAAAGCGTGTTTTTAATGGATAATCTATTAAGCGA
  164   F  I  N  Q  N  N  A  N  E  S  V  F  L  M  D  N  L  L  S  E
11341 AAGGATAGCTTATAAGCGTTCAGTGAGTGAAGGCATGGGCGTGATGGAATACAATGACAA
  184   R  I  A  Y  K  R  S  V  S  E  G  M  G  V  M  E  Y  N  D  N
11401 TAAAGCTAAAAACGAATGGTCTCAATTCTATGATGAATTGAACGGGTATTTAAGGATCAA
  204   K  A  K  N  E  W  S  Q  F  Y  D  E  L  N  G  Y  L  R  I  K
11461 AAAATGATGAAATGGTAGAATTTATAGTTTAATTCTAAATAACGCTTAATTGTGTTATGA
  224   K
```

```
11521 TATTAAAAACAATCTACTTTAAAGGATAAATATGGAATTTAAAAACACCAAAAAAGACAG
                                1 ►    M  E  F  K  N  T  K  K  D  R
11581 GCTGAGCGATTTAGAAAATCGCTTTGAGAATGCCAACAAGCATGAAACAAACAAACATGA
   11   L  S  D  L  E  N  R  F  E  N  A  N  K  H  E  T  N  K  H  E
11641 AAAAGAAGACAGGAAAAAAGCCCACACGCTCTATATTTCAGAGAAAGTGATGAACAGCGT
   31   K  E  D  R  K  K  A  H  T  L  Y  I  S  E  K  V  M  N  S  V
11701 AGAAGAGTATTTGAATGAATTTGGAGCGTTCAGAGAAAATAAAAGCGTGTTTGTTCAAGA
   51   E  E  Y  L  N  E  F  G  A  F  R  E  N  K  S  V  F  V  Q  D
11761 CGCTATTATTTTTTATTTAGAATACAAGAAAAAAGAAATGAAGCAAATGCTTTTAGATAA
   71   A  I  I  F  Y  L  E  Y  K  K  K  E  M  K  Q  M  L  L  D  K
11821 GGCGAGCAAGTTGTGATTTTATGTCATGCTCTTTTGTAAGGCTATCTTTCAACCTCATCA
   91   A  S  K  L
11881 TCTCTAGAGCATTGCTTCAAATTGTGTTGTGGTTTGGTATTTTTTGAAACTTTTTGTTGC
11941 GATTTTTGCGTATTTTGTAAAGACACCATTTCTTGCAACACTCTTAGTGTTATATTAGCC
12001 ACACCCTTACTCACACTCTTTTGCAATTCTAATTCTTTTTCTAAAAATTGCAACGACTTT
12061 TTAAGGCGAGCTAATTTTTTGCCATTATCTTCACATTCTAAAATCTCAATTTGCAGTCGC
12121 TCTTTTCTTTCAATACAGAAATTTTTTCGCTTTTCTAAAAGCTCCATGGTTTCATTTTCT
12181 CTTTGGGCAAGATTTTTTAGAAGTTCAATGCAAGAGTGGTTAGTCTCATTAATATTTTCA
12241 GCTCTTAAAGTATCGCTTGCAATCAATGCAATATTGTTTAAATATTTTAAGTGTCTATTA
12301 AGAGCGAATGAATTTGTTTTGTCATTGAGTGCTAACTGCGAATAAGACAGATAGCTATAC
12361 ACAATACCTAAAGCATTTTTAGTGTCATTGTTATGTTCTTTAATTTCTTTTTCCAACAAA
12421 ATAAGATTTTTAGTCATCGTTTTGTTTGCACCATGATGCTGTAGTGGCTTTAATAATTCA
12481 TAAGCTTTTCTTTTTCCAGCAATTGCATGCATAACCCCTATTGAGAGCATGAAATTTATT
12541 AGGTATATTGATGTAGGTGGAGCAGCAAGCCTTATTGAGTGCATTTATTTTTTACTCTAG
12601 TGTATGGATTTTCTTAGCGCTATTATAAAAATTTTGGAGCTCTTTATGCGCTATGAGTGA
12661 ATGCTAGGGCTTTTTGTTTTAGAGCATTCAAGCGTTGTTTTTCTAAAAGAGAAGTCTTTT
12721 GGGTTCTTAAAGCTTTTTCAAAATTGTGATATGGGAATTATTTGCCATAACTTGGGGGGA
12781 AGTGGCTTGGGGGGAGTGACCCAACGGCGAAGGCGATTGGCTTAAACCTTTTCCAAAATT
12841 GTTGATATGGTAATTATTTTCCATAACTTTTTTGTAAGTTCCTTTCC
```

Fig. 7. Single-stranded sequence of pHPM186. Putative promoter sequences and putative ribosome binding sequences are boxed. Direct repeats are underlined. ORFs are marked by an arrowhead that indicates direction of transcription.

	Putative ribosome binding site		Putative promoter sequences			
	Nucleotide sequence (5′ to 3′)	Nucleotide position	Nucleotide sequence (5′ to 3′)		Consensus sequence -10	-35
ORF 1	AGAACA	1290	TAGAAAT	TTGAGC	1307	1326
ORF 2	not found	not found	TATTAT	TTGATA	1392	1351
ORF 3	AGGAG	3468	TATAA	TTGAG	3454	3418
ORF 4	AGAACA	7406	TAGAAAT	TTGAGC	7423	7444
ORF 5	not found	not found	TATTAT	TTGATA	7508	7467
ORF 6	AGGAGC	10404	ATATA	TTGCCA	10383	10357
ORF 7	AGGAGC	10827	TATATA	TTGACA	10816	10783
ORF 8	not found	not found	TTTAAT	TTGAAC	11554	11504

Table 2. Putatitive ribosome binding sites and putative promoter sequences for pHPM186. *E. coli* consensus sequences were used, and are given as: RBS, AGGAGG; -10, TATAAT; -35, TTGACA.

sequence. ORF3, bases 3423 to 5007, had strong identity to *repA* genes found in the plasmids pHPM180 (87%), pHPM179 (85%) and pHel (93%). Four direct repeats of 22 bases (2662- 27050) were found upstream of the *repA* gene (ORF3). These directly repeated sequences had 82% identity to the iterons found in pHel (20), 68% identity to iterons from pHPM180 and 73% identity to iterons from pHPM179. These plasmids apparently replicate via the theta-type mechanism, thus suggesting that pHPM186 probably replicates via the theta-type mechanism as well.

Six thousand, two hundred bp of pHPM186 had identity to genes in the *H. pylori* chromosome, the sequence having been recently published (34) and available in the Genbank database (1). Included in this region were two copies of the insertion sequence, IS*605*, bases 1-1880 and bases 6136-8016, where the identity was 97% and is shown in the dot matrix analysis in Fig. 8, with the IS*605* sequence from pHPM186 on the Y axis and the chromosomal IS*605* sequence on the X axis. Both copies of the insertion sequence contained two ORF's which encode transposase (*tnp*) genes. ORF1 (bases 20 to 1147) and ORF2 (bases 1352-1795) were in one copy of IS*605*, with ORF1 identical to *tnp*B and ORF2 identical to *tnp*A. ORF4 (bases 6136-7263) and ORF5 (bases 7468-7911) were in the second copy of IS*605*, with ORF4 identical to *tnp*B and ORF5 identical to *tnp*A. Thus the copies of IS*605* were directly repeated.

The three remaining ORFs also showed identity to the *H. pylori* chromosome on section 86 as shown in Fig. 9. ORF6 (bases 10359-10544) and ORF8 (11552-11836) showed identity to the putative ORFs HP0999 and HP1001, respectively. ORF7 (bases 10793-11467) showed identity to the chromosomal gene *parA* (also called HP1000 in Fig. 9) for which Tomb et al. (34) described 50.7% identity with the *parA* gene from the *Agrobacterium tumifaciens* plasmid, pTAR. Again, the identity between pHPM186 sequences and those on the chromosome was very high (98%). The dot matrix analysis in Fig. 10 shows section 86 of the chromosome on the Y axis and pHPM186 DNA on the X axis.

Sequence identity to three other *H. pylori* genes was present also in pHPM186: 1)

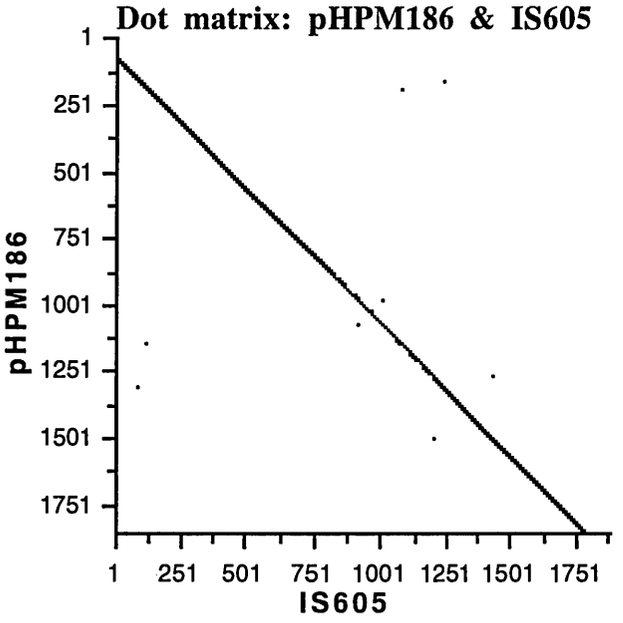

Fig. 8. Dot matrix comparison between the IS*605* found on pHPM186 (Y axis) and the IS*605* on the chromosome of *H. pylori*.

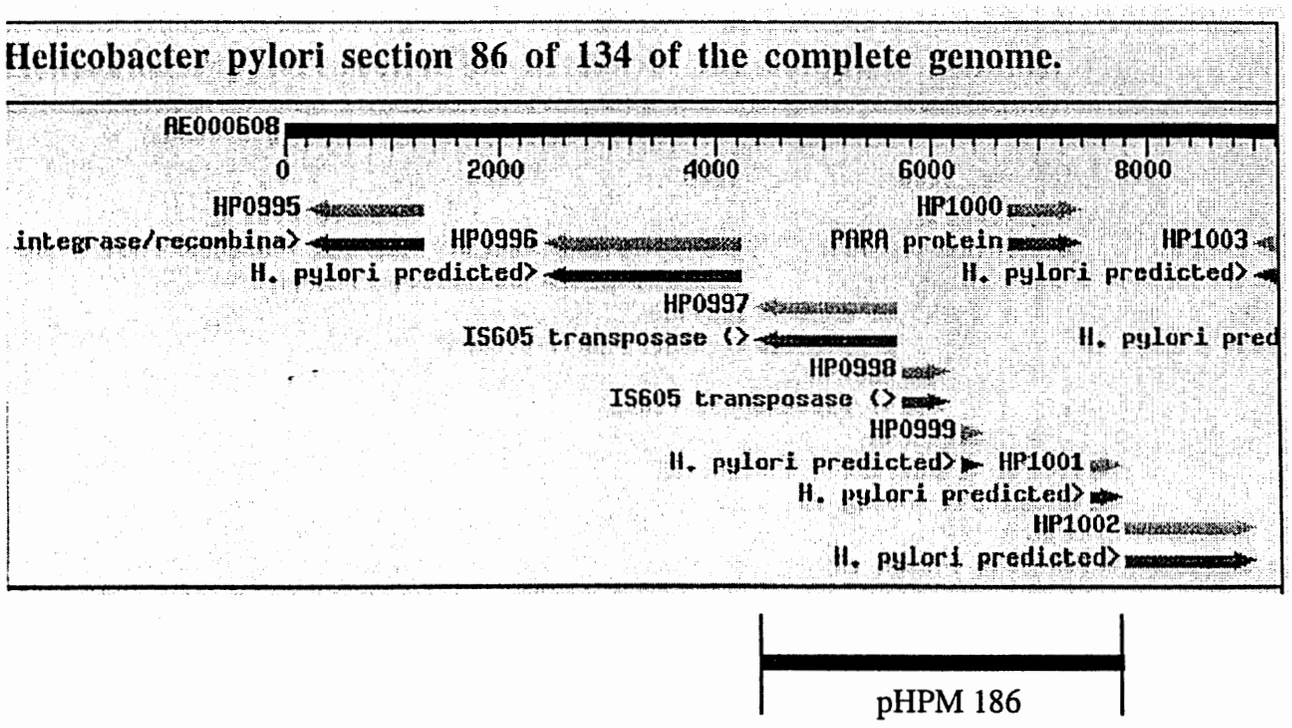

Fig. 9. Section 86 of the *H. pylori* chromosome (34). The genes found on pHPM186 are HP0999, HP1000, and HP 1001.

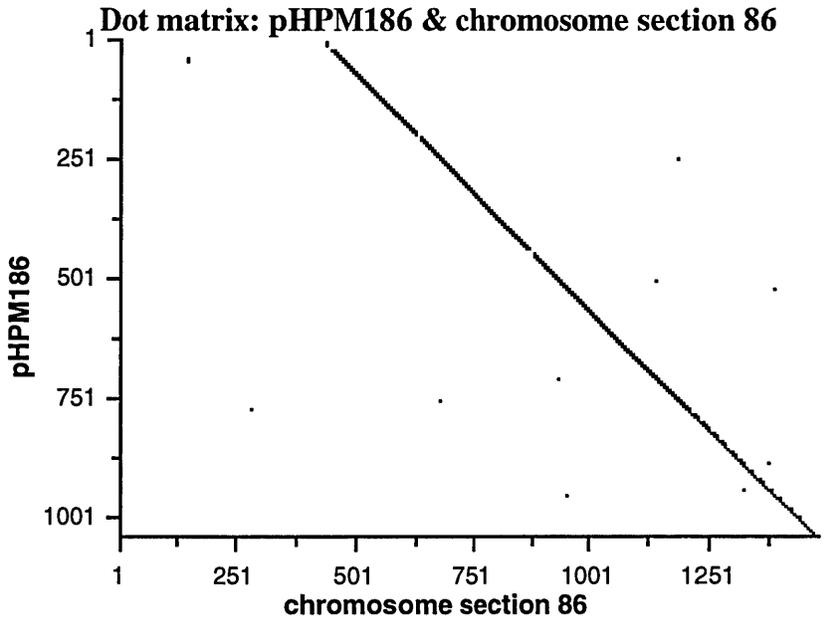

Fig. 10. Dot matrix comparsion between section 86 of the *H. pylori* chromosome (X axis) and the sequence of pHPM186 with identity to section 86 (Y axis). #1 of pHPM186 is base #10359. #1 of chromosome is base #6025 from Fig. 9.

bases 8015 to 8055 and bases 1907 to 1947 showed 90% identity to a 40 bp region of the chromosomal *cagA* gene (6); 2) bases 1999 to 2041 showed 90% identity to the *vapD* chromosomal gene (5) and; 3) bases 8016 to 8059 and bases 1908 to 1951 showed 90% identity to the *alpA* and *alpB* chromosomal genes (2, Z82988). The *cagA* gene was identified in *H. pylori* as the cytotoxin associated gene, and is a marker for pathogenesis. The *vapD* gene was initially identified in the microorganism *Dichelobacter nodosus* (5), and since it is thought to be related to virulence it is called the virulence-associated protein D gene. The *alp* genes were found in *H. pylori*, but have an unknown function. Note that the same directly repeated sequence in pHPM186 showed identity with both the *cagA* gene and the *alp* genes. The database search did not reveal any significant sequence identities for the pHPM186 DNA between coordinates 8055 and 10359. A graphic representation of pHPM186 is shown in Fig. 11.

Southern hybridizations to confirm that pHPM186 is plasmid DNA

The presence of chromosomal DNA sequences in pHPM186 made it necessary to provide evidence that the DNA was plasmid in origin and not chromosomal contamination. Both plasmid and total genomic DNA from HPM186 were digested with *Dde*I and *Dra*I endonucleases. The plasmid pHPM186 DNA was digested also with *Hin*dIII. All fragments from endonuclease digestions as well as samples of undigested plasmid and genomic DNA were separated on a 1.0% agarose gel (Fig. 12) and transferred to a nylon membrane as previously described.

The recombinants pKB2.1, pKB1.3, pKB4.5 and pKB5.0 were digested with *Hin*dIII, separated on a 0.8% agarose gel and the insert fragments excised from the gel and purified using the Elu-Quik kit. These fragments were then labeled and used as probes against the plasmid and genomic DNA digests from above (Fig. 13 and Fig. 14).

The labeled 2100 bp fragment from pKB2.1 hybridized only with plasmid DNA (Fig. 13a), as expected since this fragment contains the plasmid *repA* gene. The 1,300 bp fragment (Fig. 13b), the 4500 bp fragment (Fig. 14a), and the 5000 bp fragment

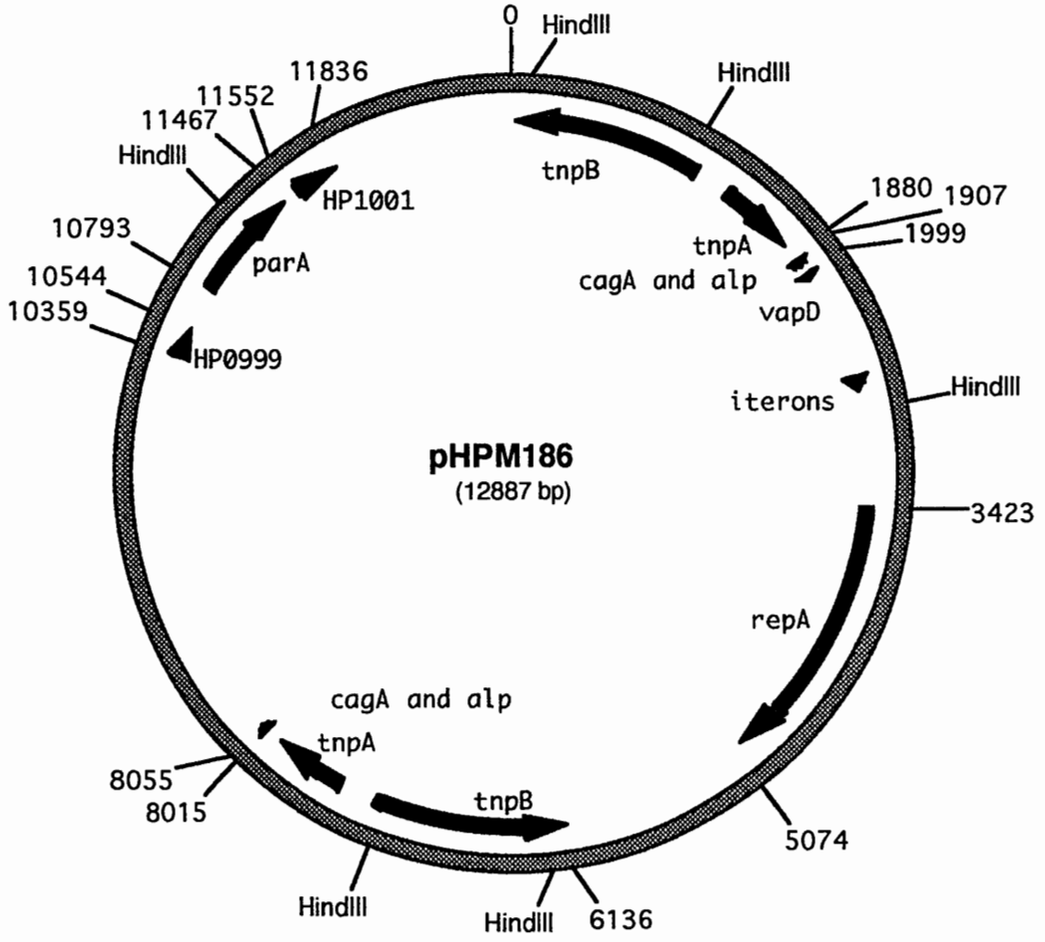

Fig. 11. Graphical representation of pHPM186.

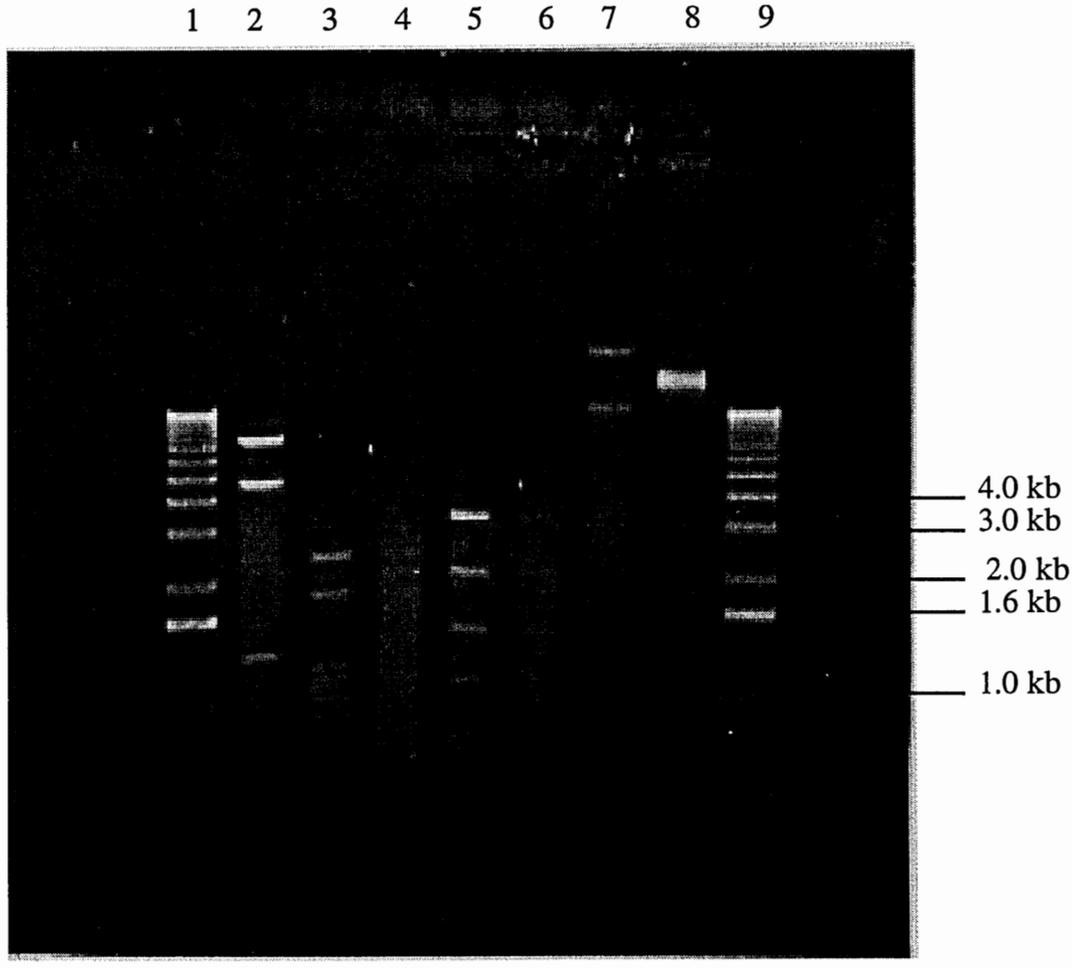

Fig. 12 . Restriction enzyme digestions of pHPM186 and HPM186 for Southern hybridization. Lanes 1 and 9, 1 kb ladder size marker; lane 2, pHPM186 /*Hin*dIII; lane 3, pHPM186/*Dde*I; lane 4, HPM186/*Dde*I; lane 5, pHPM 186/*Dra*I; lane 6, HPM186/*Dra*I; lane 7 undigested pHPM186; lane 8, undigested HPM186.

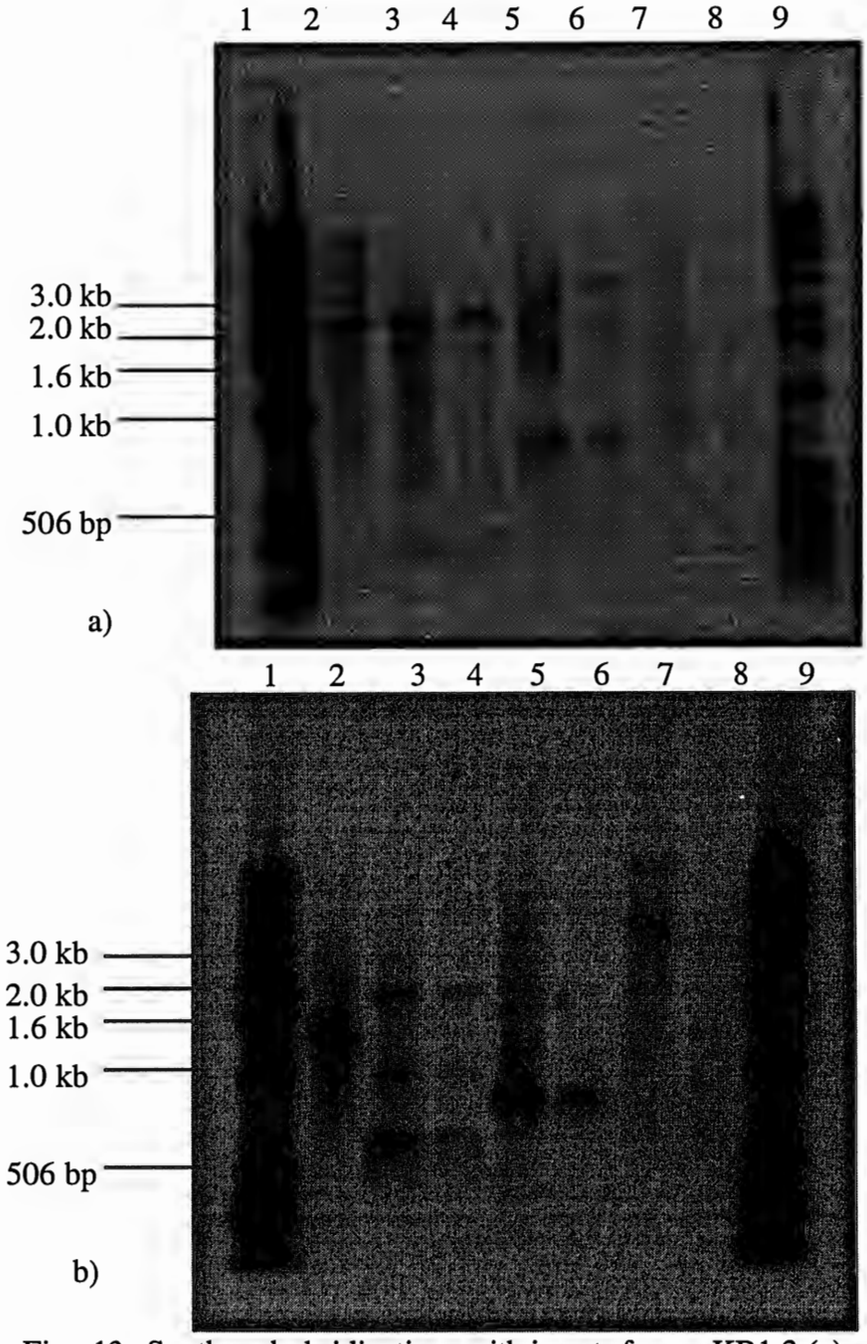

Fig. 13. Southern hybridizations with inserts from pKB1.3 (a) and pKB2.1 (b). Lanes 1 and 9, 1 kb ladder size marker; lane 2, pHPM186 /*Hin*dIII; lane 3, pHPM186/*Dde*I; lane 4, HPM186/*Dde*I; lane 5, pHPM 186/*Dra*I; lane 6, HPM186/*Dra*I; lane 7 undigested pHPM186; lane 8, undigested HPM186.

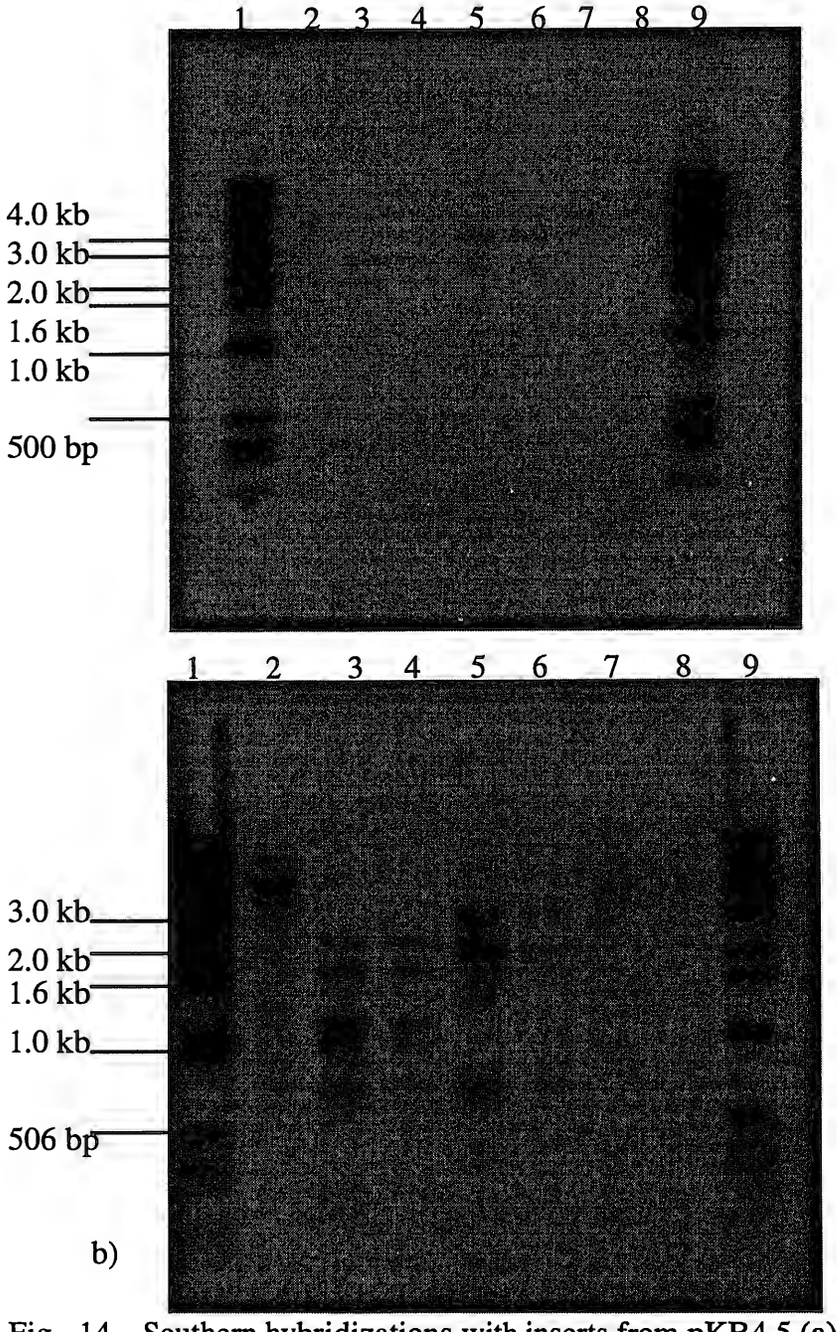

Fig. 14 . Southern hybridizations with inserts from pKB4.5 (a) and pKB5.0 (b). Lanes 1 and 9, 1 kb ladder size marker; lane 2, pHPM186/*Hin*dIII; lane 3, pHPM186/*Dde*I; lane 4, HPM186/*Dde*I; lane 5, pHPM 186/*Dra*I; lane 6, HPM186/*Dra*I; lane 7 undigested pHPM186; lane 8, undigested HPM186.

(Fig. 14b) hybridized only with plasmid DNA. The presence of additional bands in tracks 4 and 6 of these Figs. would have indicated that these plasmid sequences were present on the chromosome of HPM186. Since these additional bands were not observed, the conclusion was that the sequenced DNA was of plasmid origin and was not chromosomal.

Southern hybridization using the urease gene as a probe

One difficulty with the above hybridizations was the possibility that a single copy gene in the chromosome would not be detected. To confirm that a single copy gene, like the *parA* gene, could be detected by Southern hybridization, a probe was made using the urease gene which is a single copy gene found in all *H. pylori* strains. Primers (ure 1 and ure 2) were constructed to amplify a 600 bp fragment from the highly conserved *ureA* and *ureB* region (Table 1). This fragment was PCR amplified, purified, and labeled for use as a probe. The DNA from pHPM186 and HPM186 was separated on a 0.8% agarose gel (Fig. 15a) which was denatured and the DNA transferred to a nylon membrane only under alkaline conditions, ensuring that the DNA remained single stranded, and increasing the probability of hybridization. The urease probe hybridized to HPM186 (Fig. 15b, lane 3) but not to pHPM186 (Fig. 15b, lane 2). In addition, the previous nylon membrane (used for Fig. 12 and Fig. 13) was stripped (see methods) and exposed to the urease probe. The urease probe hybridized to HPM186 DNA (Fig. 15c, lanes 4 and 6). The results of both Southern hybridizations with the urease probe showed that a single copy gene could be detected using this method.

Southern hybridization using IS*605* as a probe

The Southern hybridizations shown in Fig. 14 a and b utilized the 4500 and 5000 bp fragments as probes against pHPM186 and genomic HPM186 DNA. The sequencing results indicated that each of these fragments contain a copy of IS*605*. The hybridization results showed that IS*605* was found only on pHPM186 and not on the chromosome of HPM186. To confirm that no copies of IS*605* were on the chromosome, a Southern

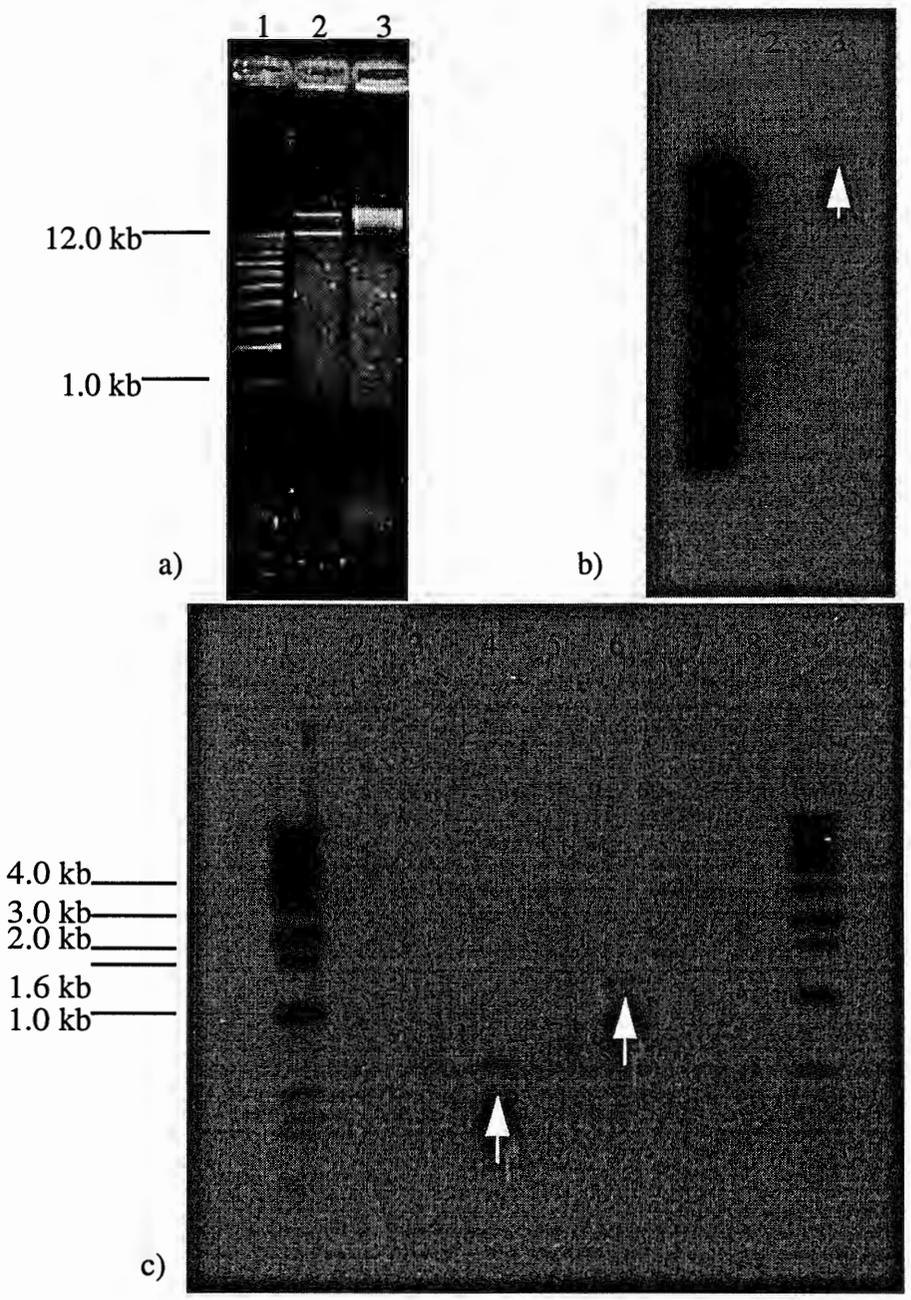

Fig. 15. Southern hybridization using urease gene as a probe. a) Agarose gel Lane 1, 1kb ladder marker; lane 2, pHPM186; lane 3, HPM186.
b) Autoradiograph Lane 1, 1kb ladder marker; lane 2, pHPM186; lane 3, HPM186. c) Autoradiograph Lanes 1 and 9, 1 kb ladder size marker; lane 2, pHPM186/*Hin*dIII; lane 3, pHPM186/*Dde*I; lane 4, HPM186/*Dde*I; lane 5, pHPM186/*Dra*I; lane 6, HPM186/*Dra*I; lane 7 undigested pHPM186; lane 8, undigested HPM186.

hybridization was performed using IS*605* as a probe. Primers for amplifying IS*605* DNA were designed (KB 23 and KB 24), and PCR amplification was performed using pHPM186 as template DNA. The resulting blunt-ended fragment was ligated into the blunt-ended *Eco*RV site of pZERO-2.1 and transformed into *E. coli* TOP10F′. A recombinant was selected, and DNA sequence of the insert was determined to confirm the presence of IS*605*. This recombinant was called pIS8. The DNA from pHPM186 and HPM186 was separated on a 0.8% agarose gel (Fig. 16a); the agarose gel was denatured and transferred to a nylon membrane under alkaline conditions, ensuring that the DNA remained single-stranded and increasing the probability of hybridization.

The pIS8 DNA was digested with *Xba*I and *Hin*dIII restriction endonucleases and separated on an agarose gel; the IS*605* fragment was excised, purified using the Elu-Quik kit, and labeled for use as a probe. The hybridization results are shown in Fig. 16b. These results confirmed that the IS605 insertion sequence was found only on plasmid DNA. The hybridization signal was very strong against isolated pHPM186 DNA (lane 2), and weaker from the plasmid DNA present in the genomic DNA in lane 3.

PCR of pHPM186

One further test was performed to confirm that the sequenced 12887 bp were from an autonomous plasmid. PCR primers that abut each other and allow DNA synthesis in opposite directions should produce a PCR product of approximately 12000 bp. The primers, KB 24 and LC 2 comp (Table 1), were used to amplify the entire plasmid from pHPM186 DNA. The native plasmid and the PCR product were digested with *Hin*dIII. The agarose gel in Fig. 17 was a comparison of *Hin*dIII digests of native pHPM186 and the PCR-amplified pHPM186 DNA and showed they are identical. To ensure

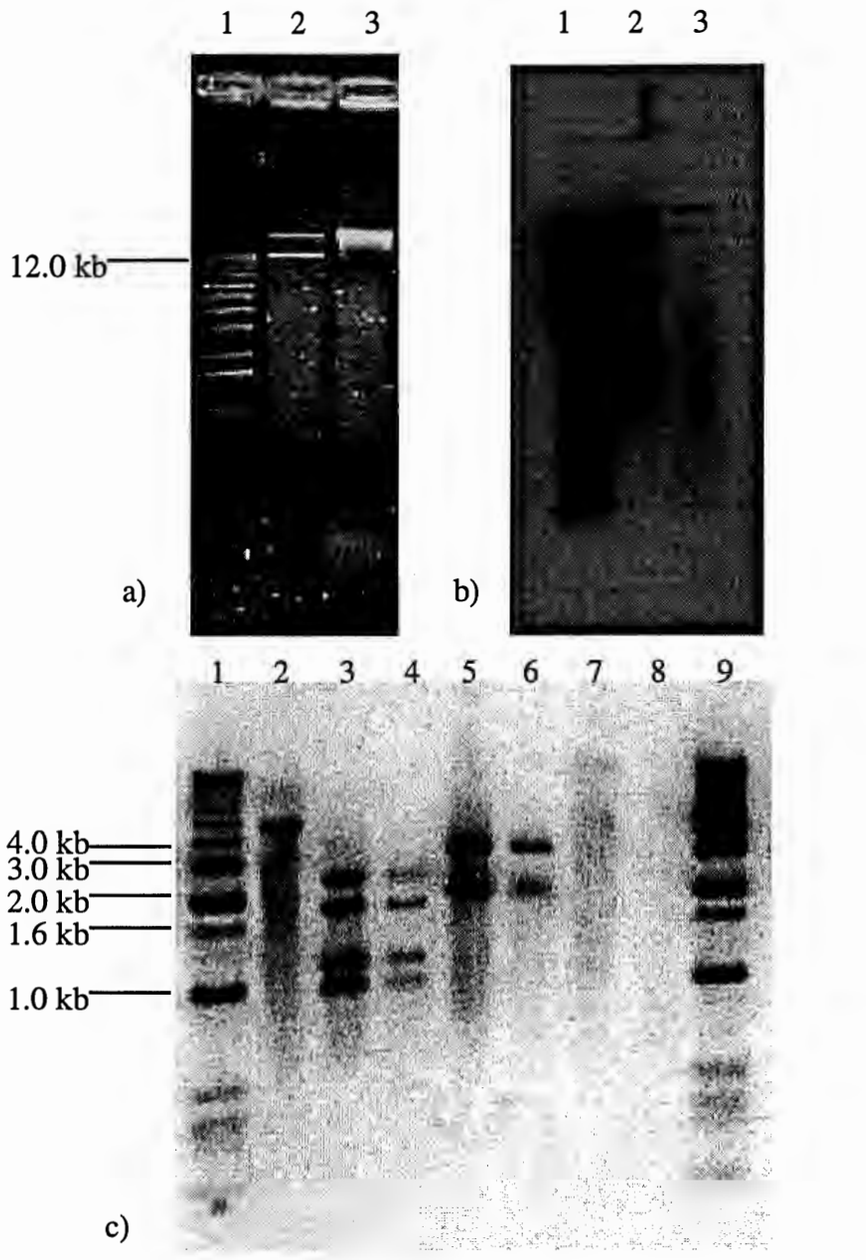

Fig. 16. Southern hybridization using IS*605* as a probe. a) Agarose gel Lane 1, 1 kb ladder marker; lane 2, pHPM186;lane 3, HPM186. b) Autoradiograph Lane 1, 1 kb ladder marker; lane 2, pHPM186; lane 3, HPM186. c) Autoradiograph Lanes 1and 9, 1 kb ladder size marker; lane 2, pHPM186/*Hin*dIII; lane 3, pHPM186/*Dde*I; lane 4, HPM186/*Dde*I; lane 5, pHPM186/*Dra*I; lane 6, HPM186/*Dra*I; lane 7 undigested pHPM186; lane 8, undigested HPM186.

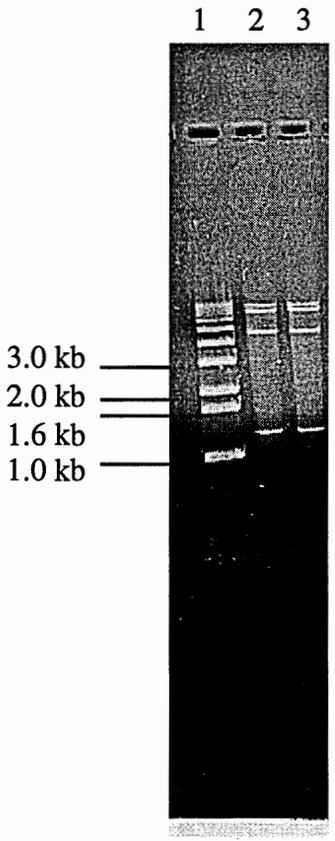

Fig. 17. PCR of pHPM186. Lane 1, 1 kb ladder marker; lane 2, PCR pHPM 186 digested with *Hin*dIII; lane 3, native pHPM186 digested with *Hin*dIII.

that this PCR product represents a circular plasmid, primers also were used that overlapped the outward primers, but pointed inward and would synthesize a smaller fragment. The inward primers were LC 1 and LC 2 (Table 1), which produced a product of 1100 bp. Primer LC 2 comp was the complementary sequence to LC 2. Primer KB 24 was on the strand complementary to primer LC 1 and overlaps that sequence. This stragedy is illustrated in Fig. 18. The synthesis of both the 12000 bp and 1100 bp products from the presumed plasmid template confirmed that pHPM186 DNA was a circular plasmid.

Construction of *H. pylori* shuttle vector

A recombinant plasmid carrying the plasmid *repA* gene and a selectable marker could be the initial construct needed to make a shuttle vector for *H. pylori.* Recombinant plasmids containing the pHPM186 *repA* gene and a kanamycin resistance (Kan^R) gene were created by amplifying the appropriate fragments using PCR. The kanamycin resistance gene was amplified from pZERO-2.1 DNA using the primers KB 10 and KB 11 (Table 1) which were designed to amplify the promoter regions and the ORF encoding kanamycin resistance. This reaction produced the expected 1200 bp fragment.

To ensure that all regulatory sequences necessary for *repA* expression were included in the recombinant plasmids, primers KB 34 and KB 35 (Table 1) were designed to amplify a 3000 bp fragment of pHPM186 that excluded all of the chromosomal sequences found on pHPM186. To confirm that the PCR-generated product was indeed the fragment required, two additional PCR reactions were performed using the 3000 bp PCR product as template. First, to verify that the internal sequences of *repA* were intact the primers KB 28 and KB 30 (Table 1) were used in a PCR reaction and a fragment of the expected size, 300 bp was produced (Fig. 19, lane 2). Second, the primers LC 1 and LC 2 (Table 1) which lie just inside the *repA* ORF and should amplify a 1100 bp fragment from *repA* (7), produced the expected product (Fig. 19, lane 3).

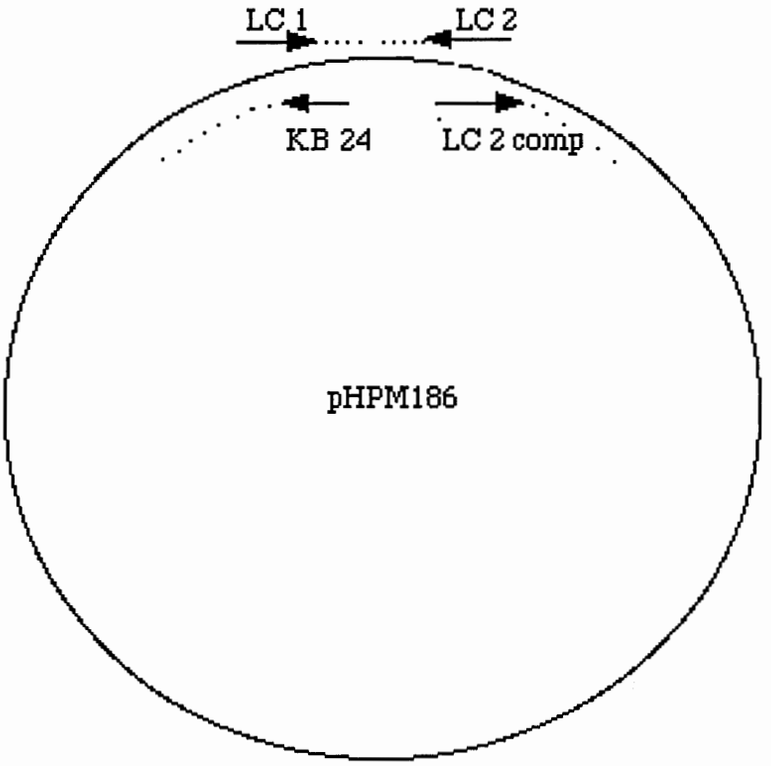

Fig. 18. Schematic diagram of primers used for PCR amplification of pHPM186. Primers LC 1 and LC 2 amplify an 1100 bp PCR product (7). Primer LC 2 comp was the complementary sequence to LC 2. Primer KB 24 was on the strand complementary to primer LC 1 and overlaps that sequence.

Ligation of the *repA* and kanamycin resistance fragments should produce a recombinant plasmid of approximately 4.3 kb (Fig. 20). To ligate the fragments, both the kanamycin resistance and the *repA* fragments were treated with Klenow enzyme to create blunt ends. The ligation reaction was a blunt end ligation with a ratio of 1:10, *repA*:Kan^R piece. The ligation mix was transformed into four restriction and modification deficient *E. coli* strains, DH5αMCR, DH5αF′, TOP10F′ and DH10B.

The *E. coli* strains DH5αMCR and DH5αF′ have the same chromosomal genotype, differing only in the presence of the F′ plasmid and the same is true of DH10B and TOP10F′. Transformants occurred only in strains that contained the F′ plasmid, with the plates of DH5αMCR and DH10B having no colonies. Transformants of the F′ containing recipients were purified, plasmid DNA was isolated using the PERFECTPrep protocol, and the recombinants were linearized with *Eco*RI. Two different transformants were recovered, a 4500 bp and a 3500 bp (Fig. 21).

Analysis of the transformants indicated that some rearrangements in the ligated sequences had occurred. One recombinant plasmid of each size was chosen for analysis. Restriction digestion showed that none of the expected fragments were present, for example, digestion of the original PCR products with *Sph*I or *Hin*dIII enzymes resulted in the expected fragments (Fig.22, lanes 2, 3, 6 and 7), but neither recombinant contained fragments of the same size (Fig. 22, lanes 4, 5, 8 and 9). PCR amplifications were performed with the recombinants using the two sets of *repA* PCR primers described above. Neither recombinant yielded a PCR product using LC 1 and LC 2 primers and recombinant #2 did not produce a product using KB 28 and KB 30 primers. Recombinant #1, however, produced the expected 300 bp fragment (Fig. 23, lane 2), indicating that the primer sites for KB 28 and KB 30 were not altered. Lastly, the *Sph*I fragments from both recombinants were cloned into pGEM-3Z and sample sequencing was performed. As expected the Kan^R gene was maintained in both recombinants, but the *repA* region in recombinant #1 was partially deleted (300 bp fragment) and in recombinant #2 was totally deleted. In recombinant #1, portions of *repA* were replaced with DNA showing identity to

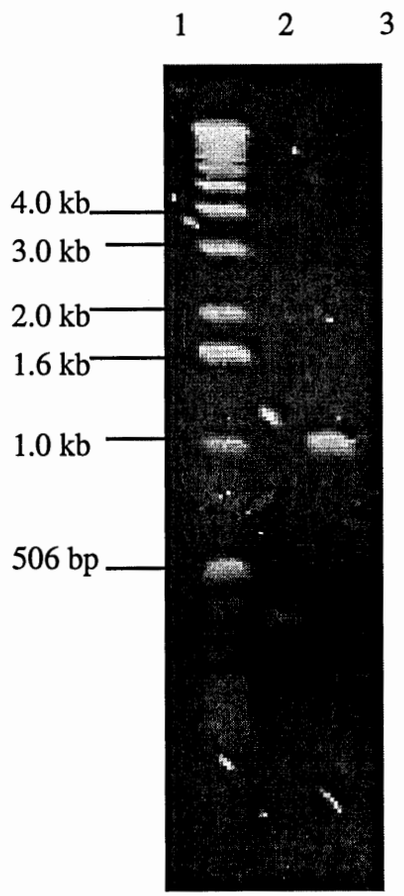

Fig. 19. PCR to verify that the *repA* portion of pHPM186 had been produced. Lane 1, 1kb ladder marker; lane 2, PCR with KB 28 and KB 30; lane 3, PCR with LC 1 and LC 2.

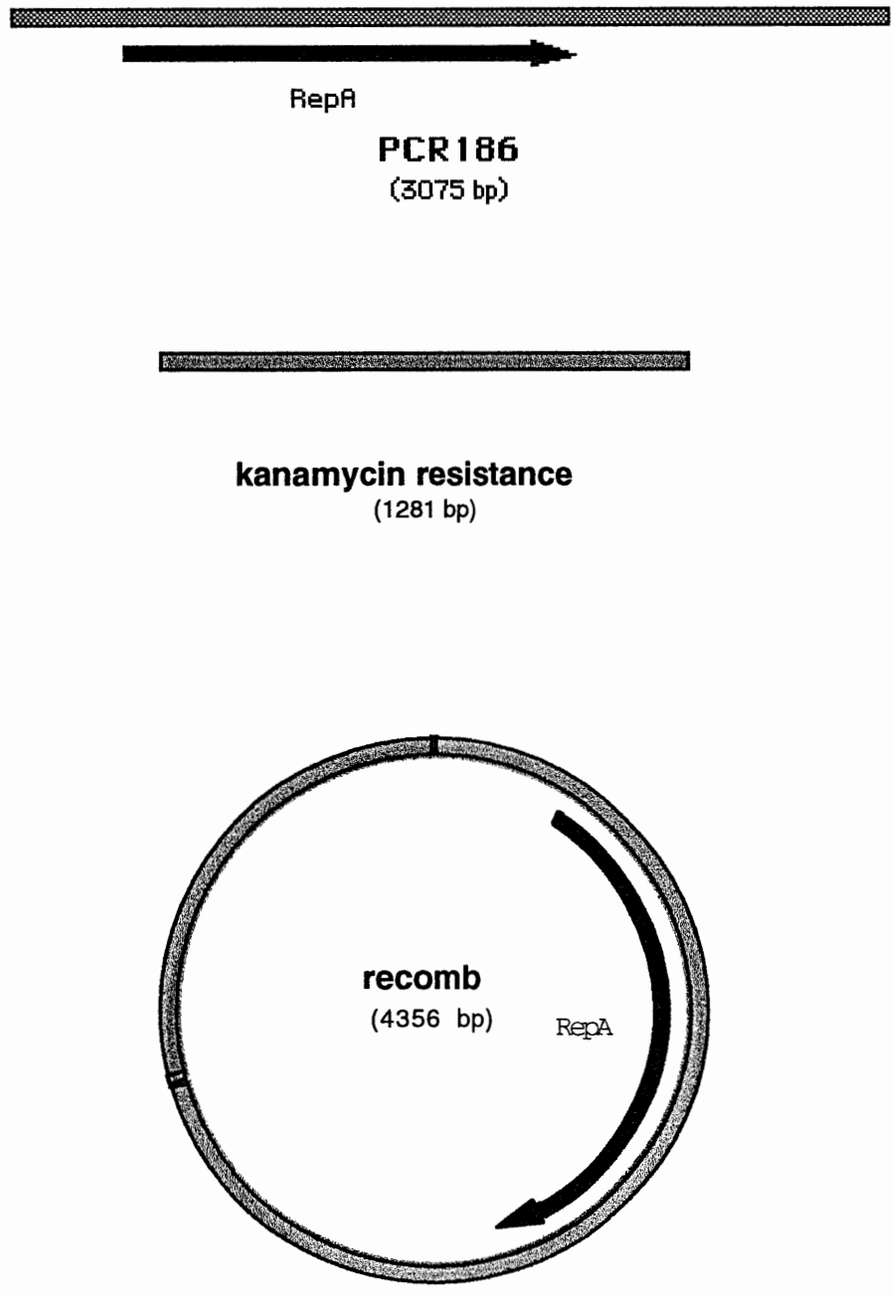

Fig. 20. Schematic representation of construction of recombinant shuttle vector. The green bar represents the *repA* gene, the blue bar represents the kanamycin gene and the red bar represents the DNA sequence on either side of the *repA* gene.

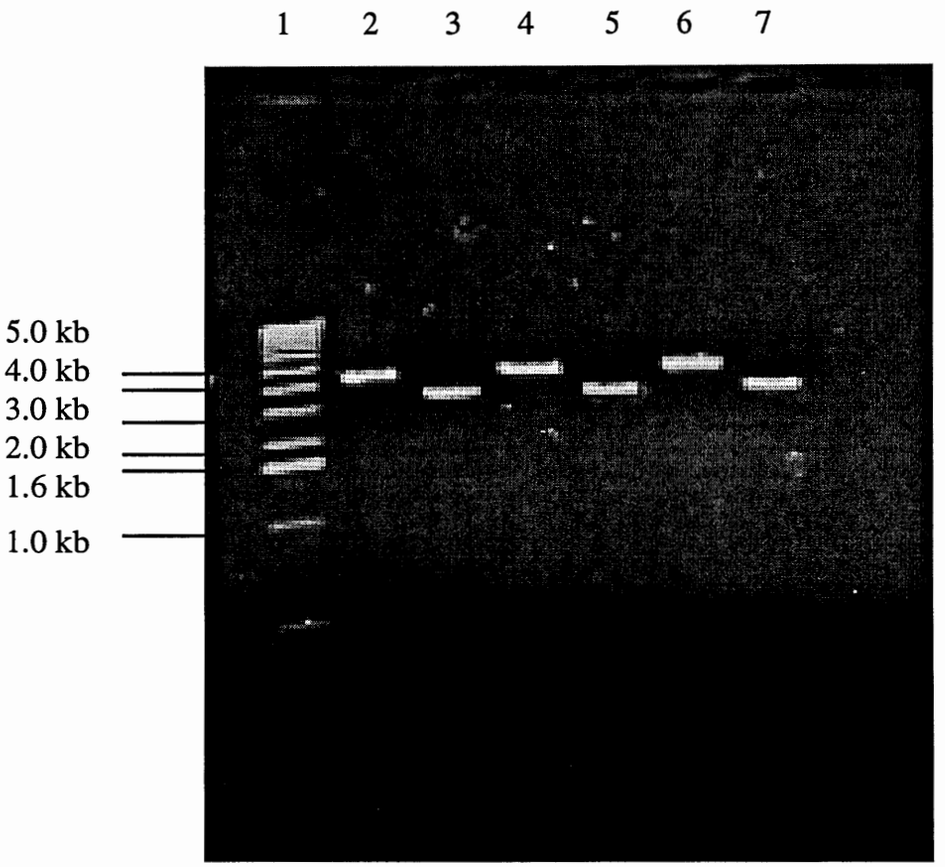

Fig. 21. Recombinant clones from tranformation of *E.coli* with ligation of pHPM186PCR and kanamycin cassette. Lane 1, 1 kb ladder marker; lane 2, recombinant 1TOP10F'; lane 3, recombinant 2 TOP10F'; lane 4, recombinant 3 TOP10F; lane 5, recombinant 4 TOP10F'; lane 6, recombinant 1DH5αF'; lane 7, recombinant 2 DH5αF'.

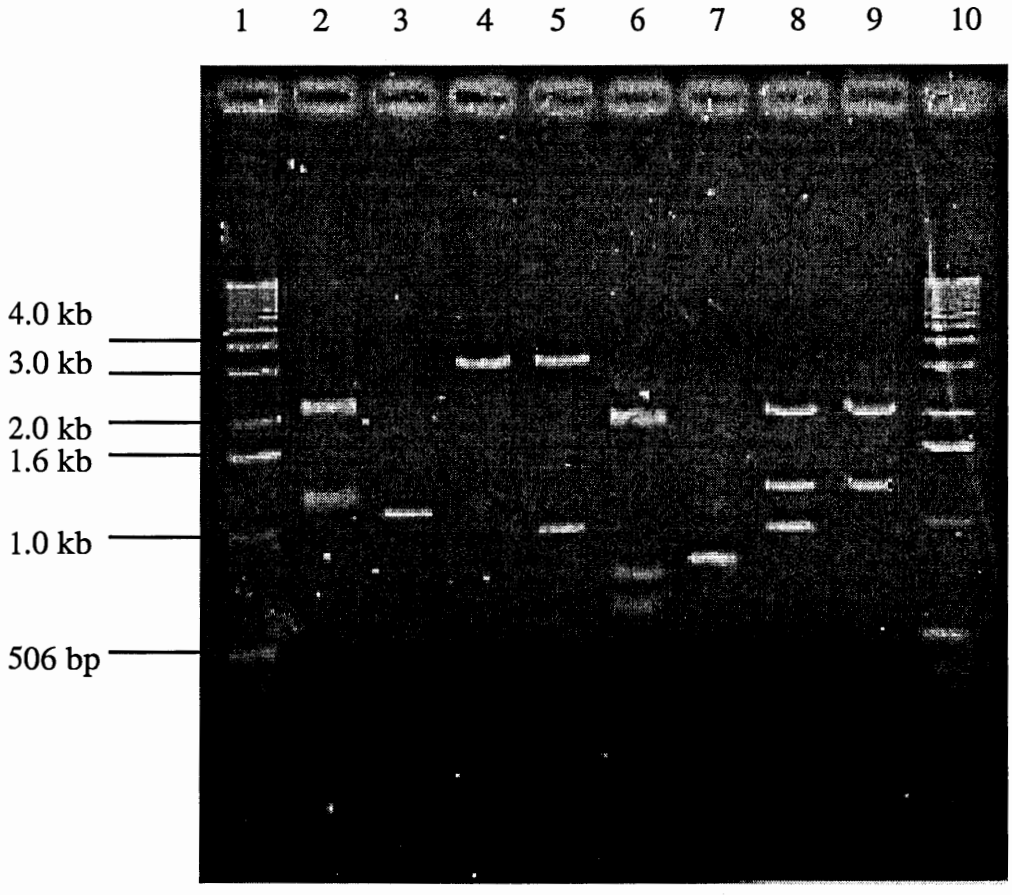

Fig. 22. Recombinant clones digested with *Hin*dIII (2-5) or *SphI* (6-9). Lane 1 and 10, 1 kb ladder marker; lane 2, PCR of pHPM186 *rep*A; lane 3, Kan cassette; lane 4, recombinant #2; lane 5, recombinant #1; lane 6, PCR of pHPM186 *rep*A; lane 7, Kan cassette; lane 8, recombinant #1.; lane 9, recombinant #2.

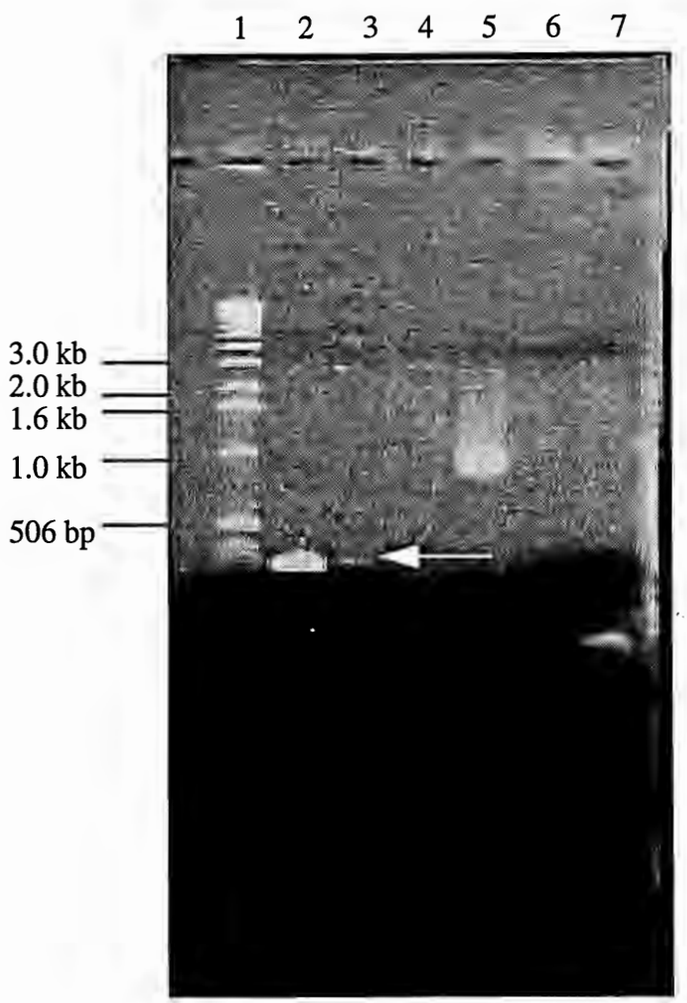

Fig. 23. PCR of recombinant clones using primers from *repA*. KB 28 and KB 30, lanes 2-4; LC 1 and LC 2, lanes 5-7. Lane 1, 1 kb ladder marker; lane 2, pHPM186 *repA*; lane 3, recombinant #1; lane 3, recombinant #2; lane 5, pHPM186 repA; lane 6, recombinant #1; lane 7, recombinant #2.

E. coli chromosomal DNA sequences. In the smaller recombinant, the *repA* gene was replaced with *E. coli* chromosomal, but different chromosomal DNA than found in recombinant #1. No DNA sequence had identity to the plasmid *repA* gene. Comparison of the extraneous *E. coli* DNA sequences in both recombinants did not yield information as to a possible mechanism for scrambling the *H. pylori repA* gene.

Transformation of *H. pylori*

H. pylori cells are naturally competent and were prepared as described in methods. Two recipient strains were used for transformation, HPM186 and HPM7 which does not contain detectable plasmid DNA. Transformation was attempted with four different DNA preparations: 1) chromosomal DNA which carried rifampin resistance; 2) the ligation mixture of the *repA* region and the kanamycin cassette; 3) recombinant #1; and 4) recombinant #2 (as described in the previous section). A rifampin resistant strain of *H. pylori* was created to be used as a control by plating HPM167 on 50 μg/ml of rifampin. Resistant colonies were chosen and replated. Genomic DNA was isolated from the rifampin resistant strain as previously described. HPM167 does not contain detectable plasmid DNA, so isolated genomic DNA contained only chromosomal DNA.

Transformation was successful only when the *H. pylori* chromosomal rifampin resistant DNA was used. The Rif^R DNA transformed both HPM186 and HPM7, indicating that the presence of plasmid DNA in the recipient was not critical for transformation. The other DNAs, ligation mix and recombinants #1 and #2, were plated on a series of Kan concentrations: 10 μg/ml; 20 μg/ml; 30 μg/ml; and 50 μg/ml. Transformants were found only on the 10, 20 and 30 μg/ml concentrations and these transformants were small colonies that appear after 2-3 days, but failed to continue growing. These colonies could be replated within the 2 or 3 days, but again small colonies appeared and failed to grow beyond 3 days. Neither of the recombinants yielded transformants, which was not surprising since the *repA* region was rearranged.

CHAPTER IV
DISCUSSION

The pathogenicity island (PAI) of *H. pylori* has a significant amount of genetic variation and portions of the PAI are often deleted in various strains (6). It has been suggested by several investigators that these rearrangements may be due, in part, to plasmids (10, 14, 21). If plasmids are involved in these rearrangements, then plasmids might carry portions of chromosomal DNA. The goal of this project was to determine whether such plasmids exist. *cagA*, a gene found in the PAI, has been well characterized and thus a portion of the *cagA* gene was chosen to probe various plasmid DNAs for the presence of that chromosomal sequence (35). The plasmid pHPM186 was chosen for sequencing following Southern hybridization which indicated that pHPM186 carried *cagA* DNA. Subsequent sequence determination of pHPM186 did in fact reveal two small stretches of DNA with 90% identity to the *cagA* gene.

Identification of presumed chromosomal DNA on pHPM186 made it necessary to prove that the sequenced DNA was actually of plasmid origin. Two separate approaches were used to obtain this proof: 1) Southern hybridizations; and 2) PCR.

First, both purified plasmid DNA and genomic DNA were tested with two sets of probes, the four purified *Hin*dIII plasmid fragments and a PCR amplified IS*605* fragment. If the sequenced plasmid DNA was chromosomal, then chromosomal bands should have hybridized with the *Hin*dIII probes. Such hybridization did not occur, indicating that the sequenced DNA was an autonomous plasmid. Since two of these *Hin*dIII fragments contained IS*605* DNA, these results further indicated that IS*605* was not present on the chromosome of HPM186. To confirm this observation, an IS*605* probe was tested against the same plasmid and genomic preparations. No hybridizations to chromosomal DNA were observed, confirming that IS*605* was only present in the plasmid.

Second, two sets of PCR primers were used to show that pHPM186 was a circular plasmid. The first set of primers (KB 24 and LC 2 comp in Fig. 18) primed DNA synthesis in opposite directions and a produced a 12000 bp PCR product. The second set

of primers (LC 1 and LC 2 in Fig. 18) primed DNA synthesis in convergent direction, yielding a 1100 bp PCR product. Since these sets of primers overlapped, these amplifications resulted in a circular map of pHPM186. Restriction profiles of the PCR products also gave the expected results which again verified that pHPM186 is an autonomous plasmid.

In addition to the identity with the *cagA* gene, pHPM186 had short stretches of identity with other chromosomal genes, for example, bases 1999 to 2041 had identity to the *vapD* gene, which is a gene of unknown function originally described in *Dichelobacter nodosus* (5). It is suspected to be related to virulence and thus is called the **v**irulence-**a**ssociated **p**rotein D gene. The *D. nodosus* gene is closely related (66%) to the *vapD* gene found on the rolling circle plasmid, pVT736-1 of *Actinobacillus actinomycetemcomitans*. A cryptic plasmid, pJD1, of *Neisseria gonorrhoeae* also encodes a gene similar (48%) to the *vapD* gene of *D. nodosus* (5). In other genera, including *H. pylori*, the *vapD* gene is found on the chromosome. For example, *Haemophilus influenzae* contains a chromosomal gene with 53% identity to the *vapD* gene of *D. nodosus*. Since this gene is found in both plasmid and chromosomal locations, it is possible that the *vapD* sequence on pHPM186 originated on a plasmid (5). This would suggest that the majority of the *vapD* gene sequence, presumably found on the chromosome of HPM186, is of plasmid origin.

Identity was also noted to genes known as *alpA* and *alpB* (2, Genbank # Z82988). This identity was found twice, due to a directly repeated element from bases 8016 to 8059 and bases 1908 to 1951. The *alp* gene identity was found only in Genbank, as no publication has been released detailing the characteristics of these genes. Therefore, it is impossible to comment on the significance of this finding. It was noted however, that the bases 8016 to 8059 and 1908 to 1951 had identity to both the *cagA* gene and the *alp* genes.

The most significant region of identity between pHPM186 and the *H. pylori* chromosome occurred from bases 10359 to 11836 in pHPM186. This region, which contains three ORFs, was 98% identical to three ORFs found in section 86 of the chromosome (34). The first and third ORFs (ORF6 and 8, respectively) encode

hypothetical proteins. Neither of these protein sequences had identity to known protein sequences. The second ORF (ORF7) had 98% identity to the chromosomal gene *parA*.

The *parA* gene is found in P1 plasmid, in the pTAR plasmid from *Agrobacterium tumifaciens*, and in many other low copy plasmids (17). It is part of the *par* partitioning region encoded on these plasmids, which is essential for the partitioning of plasmids into daughter cells. The *par* system is found in low copy number plasmids, because high copy plasmids rely on passive distribution of plasmid DNA into daughter cells. Partitioning of low copy plasmids depends on two genes, *parA* and *parB*, although it has been shown that the *A. tumefaciens* plasmid only produces a single protein product, ParA (16). Although the role of the ParA protein in partitioning is not well understood, ParA is involved in regulating the *par* operon through a feedback mechanism. For example, Davis et al. (1992) showed that ParA can reduce the expression of a reporter gene up to fivefold. However, if either ParA or ParB are overexpressed, partitioning is disrupted. The *parA* gene has been found on the chromosome of *H. pylori*, which is unusual, since this is normally a plasmid-encoded gene. This may suggest that *parA* is found on the chromosome as a result of past recombination with a plasmid.

Two directly repeated copies of IS*605* occurred (bases 1-1880 and bases 6136-8016) on pHPM186. Between these copies of IS*605* were the three ORFs described above (ORF6, 10359-10544; ORF7, 10793-11467; and ORF8, 11552-11836) and an additional 2300 bp of DNA that lacked identity to any known sequence, including the *H. pylori* sequenced strain #26695 (34). With the known diversity of *H. pylori* DNA, it is possible that these 2300 bp were originally part of the chromosome of HPM186, probably residing next to ORF6, ORF7 and ORF8, between two copies of IS*605*.

Region 86 of the *H. pylori* genome (Fig. 9) contains the three ORFs found on pHPM186: HP0999, *parA*, and HP1001 (34). All of the ORFs on the *H. pylori* genome are numbered sequentially, so the *parA* gene is HP1000. Genes HP0997 and HP0998 are the two transposase genes, *tnpB* and *tnpA*, that are part of the IS*605* element. Gene HP1002 is not another transposase gene, as one might expect from studying the

arrangement of the genes in pHPM186, but is instead another hypothetical ORF. This is not surprising, due to the lack of consistent gene order among strains of *H. pylori*. As suggested above, it is possible that the chromosome of the strain containing pHPM186 was not identical to strain 26695 and may have had an IS*605* element where the chromosome of the strain 26695 has an hypothetical ORF, and it is this region of HPM186, between the two IS*605*s, that is now found in pHPM186.

IS*605* is also a chromosomal element found in the PAI as well as at other sites throughout the chromosome. Logan and Berg (1996) suggestted that the mechanism by which genetic diversity arises in *H. pylori* may involve IS*605*, and a related insertion sequence, IS*606*. The two copies of IS*605* were found in a direct repeat pattern in pHPM186, which may allow the sequences to be more stable in the plasmid. The discovery of IS*605* elements on plasmid DNA supported the concept that insertion sequences may move between plasmid and chromosome, contributing to the immense genetic diversity found in *H. pylori*.

Although plasmid DNA is found in approximately 80% of *H. pylori* isolates, to date the function of these plasmids is unknown (25). This study provided information as to the possible function of such plasmids by showing that pHPM186, isolated from the strain HPM186, contained DNA sequences normally found on the chromosome of *H. pylori*. This discovery lends support to the idea that an *H. pylori* plasmid may have recombined into and out of the chromosome of *H. pylori*, taking with it some chromosomal sequences.

Absence of the region 86 DNA in the chromosome of HPM186 indicated that the region moved from chromosome to plasmid, probably due to the activity of IS*605*. The chromosomal DNA in pHPM186 could be viewed as a composite transposon, with IS*605* on either end. However, the presence of 40 bp of *cagA* sequence on the "outside" of one IS*605* (bp 1907 to 1947) would seem to discredit this view. A more likely model is that in the parental strain, IS*605* was inserted in the middle of *cagA* in the chromosome and the excision mechanism incorporated 40 bp of *cagA* sequence into the plasmid. Presumably

the second copy of IS*605* on pHPM186 arose by duplication of this sequence, perhaps reflecting the mechanism of transposition.

The plasmid pHPM186 contained one ORF (ORF3) with identity to the plasmid replication gene, *repA*. This gene has been found in the *H. pylori* plasmids pHPM180, pHPM179, and pHel, as well as in several plasmids isolated from various bacterial genera (25). In addition to *repA* , pHPM186 contained four directly repeated 22 bp sequences upstream of the ORF which had 82% identity to the iterons found in pHel (20), 68% identity to iterons from pHPM180 and 73% identity to iterons from pHPM179 (20, 25, 28). Because Minnis et al. (1995) showed that the *repA* ORF in pHPM180 was transcribed in HPM180 and that the protein had identity with other Rep proteins from microorgansims that replicate via a theta-type mechanism, so based on the identities observed it is likely that pHPM186 also replicates via this mechanism.

The construction of a vector containing a *repA* gene from *H. pylori* and an antibiotic resistance gene would be the first step towards creating a shuttle vector between *E. coli* and *H. pylori*. Since the DNA sequence of pHPM186 was determined and the location of the *repA* gene and iterons known, it should be possible to use these sequences to create such a vector. The *repA* portion of pHPM186 was PCR amplified, ligated to a Kan^R gene, and transformed into *E. coli*. Transformants appeared and recombinants survived only in F^+ strains of *E. coli*; however, when recombinant plasmids from the F^+ *E. coli* were analyzed, none of the intact *repA* sequence was maintained.

All of these observations suggest that *E. coli* is intolerant of the *H. pylori* plasmid *repA* DNA sequence. It may be that the plasmid *repA* gene interferes with replication of the *E. coli* chromosome, so that the only plasmids that survive are those with rearrangements of the *repA* gene. The rearrangement does not appear to be random, however, as the plasmids are of two kinds. The recombinants with the scrambled *repA* gene cannot survive in an *E. coli* strain without another plasmid present to provide a replication function.

The *repA*/KanR recombinant construct is not maintained in *E. coli* , although recombinants with rearranged or deleted repA survive in cells containing an F plasmid.

This suggests that the *repA* gene is transcribed in *E. coli* and that the translated protein product is lethal to *E. coli* cells. Thus any recombinant with an intact *repA* gene does not survive. Rearrangement or deletion of *repA* produces a recombinant that can survive only if the F plasmid provides some replication function for the recombinant.

Recovery of a *repA*/Kan^R recombinant in *H. pylori* also was unsuccessful. The ligation mixture was transformed into HPM186 and into HPM7, a strain lacking plasmids. The Kan^R *H. pylori* transformants grew as very small colonies for 2 days and then stopped growing. If replated the transformants grew, but again only for a few days. These results were unexpected since the *H. pylori repA* region should be expressed by these cells. A series of kanamycin concentrations was used and this did not change the results. It is possible that the kan^R gene was not tolerated by *H. pylori*, since previous attempts at creating a shuttle vector using pHPM180 *repA* gave the same results. When pHPM180 was used, both kanamycin and ampillicin resistance genes were tried with similar negative results.

The conclusions from this work are: 1) pHPM186 is a 12887 bp plasmid that carries approximately 8000 bp of DNA sequence with identity to the *H. pylori* chromosome, including two copies of IS*605*; 2) the plasmid probably replicates via the theta-type mechanism; and 3) the construct suggested in this work will not function as a shuttle vector between *H. pylori* and *E. coli.*

This study demonstrated that IS*605*, normally found on the chromosome of *H. pylori*, was contained on a plasmid along with portions of the chromosome. Therefore, IS*605* may be involved in chromosomal rearrangements, which are common in *H. pylori.* The model proposed above suggested that the IS*605* moved from the chromosome of *H. pylori* to the plasmid. Dissemination of this plasmid to other strains could provide a means of moving this DNA to another chromosome, thus generating the genetic diversity that is observed in *H. pylori.* The implication from this study is that *H. pylori* plasmids are involved in recombination events both with the *H. pylori* chromosome and between *H. pylori* strains.

REFERENCES

1. Altschul, S.F., W. Gish, W. Miller, E.W. Meyers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

2. Benson, D.A., M. Boguski, D.J. Lipman, and J. Ostell. 1994. GenBank. Nucl. Acids Res. 22:3441-3444.

3. Blaser, M.J. 1993. *Helicobacter pylori*: microbiology of a "slow" bacterial infection. Trends Microbiol. 1:255-260.

4. Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H.Tschape, and J. Hacker. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. 62:606-614.

5. Cao, Ping, and T.L. Cover. 1997. High-level genetic diversity in the *vapD* chromosomal region of *Helicobacter pylori.* J. Bacteriol. 179:2852-2856.

6. Censini, S., C. Lange, Z. Xiang, J.E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci. 1996. *Cag*, a pathogenicity island of *Helicobacter pylori*, encodes type-I specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. USA. 93:14648-14653.

7. Cissell, L.K. 1996. Hybridization and PCR homologies among *Helicobacter pylori* plasmids. M.S. thesis. Texas Woman's University, Denton TX.

8. Clayton, C.L., B.W. Wren, P. Mullany, A. Topping, and S. Tabaqchali. 1989. Molecular cloning and expression of *Campylobacter pylori* species-specific antigens in *Escherichia coli* K-12. Infect. Immun. 57:623-629.

9. Covacci, A., S. Falkow, D.E. Berg, and R. Rappuoli. 1997. Did the inheritance of a pathogenicity island modify the virulence of *Helicobacter pylori.* Trends in Microbiol. 5:205-208.

10. Cover, T.L. 1997. Genetic diversity in *Helicobacter pylori. Helicobacter* 2:108.

11. Davis, M.A., K.A. Martin, and S.J. Austin. 1992. Biochemical activities of the ParA partition protein of the P1 plasmid. Mol. Microbiol. 6:1141-1147.

12. Eaton, K.A., D.R. Morgan and S. Krakowka. 1989. *Campylobacter pylori* virulence factors in gnotobiotic piglets. Infect. Immun. 57:1119-1125.

13. Evans, D.G., D.J. Evans, Jr., J.J. Moulds and D.Y. Graham. 1988. N-acetylneuraminyllactose-binding fibrillar hemagglutinin of *Campylobacter pylori*: a putative colonization factor antigen. Infect. Immun. 56:2896-2906.

14. Evans, Jr., D. and D.G. Evans. 1997. Direct repeat sequences in the *cagA* gene of *Helicobacter pylori*: a ghost of a chance encounter? Mol. Microbiol. 23: 409-411.

15. Fitura, N.P., P. Guglielmetti, A. Rossolini, A. Barberi, G. Cusi, R.A. Musmanno, M. Russi, and S. Quaranta. 1989. Cytotoxin production by *Campylobacter pylori* strains isolated from patients with peptic ulcers and from patients with chronic gastritis only. J. Clin. Microbiol. 27:225-226.

16. Gallie, D.R., and C.I. Kado. 1987. *Agrobacterium tumefaciens* pTAR *parA* promoter region involved in autoregulation, incompatibility and plasmid partitioning. J. Mol. Biol. 193:465-478.

17. Gallie, D.R., D. Zaitlin, K.L. Perry, and C.I. Kado. 1984. Characterization of the replication and stability regions of *Agrobacterium tumefaciens* plasmid pTAR. J. Bacteriol. 157:739-745.

18. Haas, R., T.F. Meyer, and J.P.M. van Putten. 1993. Aflagellated mutants of *Helicobacter pylori* generated by genetic transformation of naturally competent strains using transposon shuttle mutagenesis. Mol. Microbiol. 8:753-760.

19. Harris, P.R., T.L. Cover, D.R. Crowe, J.M. Orenstein, M.F. Graham, M.J. Blaser and P.D. Smith. 1996. *Helicobacter pylori* cytotoxin induces vacuolation of primary human mucosal epithelial cells. Infect. Immun. 64:4867-4871.

20. Hauermann, D., and R. Haas. 1995. Genetic organization of a small cryptic plasmid of *Helicobacter pylori.* Gene 165:17-24.

21. Jiang, Q., K. Hiratsuka, and D.E. Taylor. 1996. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. Mol. Microbiol. 20:833-842.

22. Kleanthous, H., C.L. Clayton, and S. Tabaqchali. 1991. Characterization of a plasmid from *Helicobacter pylori* encoding a replication protein common to plasmids in Gram-positive bacteria. Mol. Microbiol. 5:2377-2389.

23. Labigne, A., V. Cussac, and P. Courcoux. 1991. Shuttle cloning and nucleotide sequences of *Helicobacter pylori* genes responsible for urease activity. J. Bacteriol. 173:1920-1931.

24. Logan, R.P.H., and D.E. Berg. 1996. Genetic diversity of *Helicobacter pylori.* The Lancet 348:1462-1463.

25. Minnis, J.A., T.E. Taylor, J.E. Knesek, W.L. Peterson, and S.A. McIntire. 1995. Characterization of a 3.5 kbp plasmid from *Helicobacter pylori.* Plasmid 43:22-36.

26. Neman-Simha, V. and F. Megraud. 1988. In vitro model for *Campylobacter pylori* adherence properties. Infect. Immun. 56:3329-3333.

27. Phadnis, S.H., D. Ilver, L. Janzon, S. Normark, and T.U. Westblom. 1994. Pathological significance and molecular characterization of the vacuolating toxin gene of *Helicobacter pylori.* Infect. Immun. 62:1557-1565.

28. Qasem, J.A. 1995. Cloning and sequence analysis of plasmid DNA from *Helicobacter pylori.* Ph.D. dissertation. Texas Woman's University, Denton, TX.

29. Stein, M.A., K.Y. Leung, M. Zwick, F.G. Portillo, and B.B. Finlay. 1996. Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. Mol. Microbiol. 20:151-164.

30. Sambrook, J., E.T. Fritsch, and T. Maniatis. 1989. Molecular Cloning. A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

31. Taylor, D.E., M. Eaton, N. Chang, S.M. Salama. 1992. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. J. Bacteriol. 174:6800-6806.

32. Taylor, T., J. Knesek, and S. McIntire. 1996. Unusual structure of a *Helicobacter pylori* plasmid. 96th General Meeting of the American Society for Microbiology. New Orleans, LA, May, 1996.

33. Tjia, T.N., W.E.S. Harper, C.S. Goodwin, and W.B. Grubb. 1987. Plasmids in *Campylobacter pyloridis*. Microbios. Lett. 36:7-11.

34. Tomb, J.F., O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.D. Fleischmann, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty, K. Nelson J. Quackenbush, L. Zhou, E.F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H.G. Khalak, A. Glodek, K. McKenney, L.M. Fitzegerald, N. Lee, M.D. Adams, and J.C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 338: 539-547.

35. Tummuru, M.K.R., T.L. Cover and M.J. Blaser. 1993. Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. Infect. Immun. 61:1799-1809.

36. Willetts, N. and R. Skurray. 1987. Structure and function of the F factor and mechanism of conjugation, p. 1110-1129. *In* F. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. American Society for Microbiology. Washington, D.C.

37. Xiang, Z., S. Censis, P.F. Bayeli, J.L. Telford, N. Figura, R. Rappuoli and A. Covacci. 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. Infect. Immun. 63:94-98.