CLONING AND SEQUENCE ANALYSIS OF PLASMID DNA FROM *HELICOBACTER PYLORI*

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

BY

JAFAR ABDULRIDA ABDULLA QASEM, B.S., M.S.

DENTON, TEXAS

MAY 1995

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS

April 5, 1995
Date

To the Associate Vice President for Research and Dean of the Graduate School:

I am submitting herewith a dissertation written by Jafar Abdulrida Abdulla Qasem entitled "Cloning and Sequence Analysis of Plasmid DNA from *Helicobacter pylori*." 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:

Richard J. Ludwick

F. Sellin

John E. Kneek

J. S. Hardcastle

Accepted

Leslie M Thompson
Associate Vice President for Research and
Dean of the Graduate School.

ACKNOWLEDGMENT

I would like to express my sincerest thanks to my committee members for their input and patience: Dr. John Knesek, Dr. Fritz Schwalm, Dr. Micheal Rudick and Dr. James Hardcastle. Special thanks to Dr. Sarah McIntire advisor and chairman of the committee, for her patience, valuable guidance and helpful criticisms during my research work. I am also greatly indebted for her going over the early draft of the manuscript with a fine tooth comb, many of the improvements in various sections are a direct result of her efforts.

I am also indebted to Dr. Louise Higgins for her excellent photographic assistance. Special thanks to Dr. Jennifer Minnis and Dr. Ahmed Yanouri, my good friends and colleagues, for always being there when I needed them, and for always telling me "I could". And it would have been a lot less fun without them.

To two amazing people who waited in Kuwait. They wondered what in the world I was doing, how may years more.?. And through it all, they never stopped praying for me. These two people are my mother and father, and without them, I never would have found my way to this world . To my brothers and sisters. Special thanks to my wife Eqbal for her patience, support and understanding.

Last, and most of all to God who has been my guidance and my source of inner strength through my study. I offer this humble accomplishment.

Jafar Abdulrida Abdull Jasem

Denton, Texas 1995

ABSTRACT

CLONING AND SEQUENCE ANALYSIS OF PLASMID DNA FROM *HELICOBACTER PYLORI*

JAFAR ABDULRIDA ABDULLA QASEM

May 1995

The bacterium *Helicobacter pylori* is recognized as the cause of some types of gastritis and gastric ulcers. Many bacterial species contain plasmid DNA which encodes virulence factors, including antibiotic resistance, adherence factors, degradative enzymes, and toxins. The majority of *H. pylori* strains contain plasmid DNA, although the size varies from strain to strain. No phenotype has been assigned to the plasmids of *H. pylori*. The purpose of this study was to characterize the plasmid, pHPM179, isolated from strain HPM179.

Three *Hin*dIII fragments (2.2, 1.6, and 1.1 kb) were ligated into the vector pTZ19R and transformed into *E. coli* DH5αF' and DH5αMCR for sequence determination. The DNA sequence was determined using the dideoxy ribonucleotide chain termination method. This sequence was added to the sequence of the 684 bp *Hin*dIII fragment sequenced previously by Chen (Chen, MS thesis 1993, Denton, TX).

The complete DNA sequence was analyzed using DNAsis and DNA inspector *II*e computer programs, as well as BLASTP and BLASTN programs provided by the NIH. Three open reading frames (ORFs) were determined: ORF1 could encode a protein of 52 kDa; ORF2 could encode a protein of 10.8 kDa; and ORF3 could encode a protein of 31.9 kDa.

Comparison of the pHPM179 DNA sequence with those in the GenBank revealed matches with other *H. pylori* sequences. Found were regions of near identity with *H. pylori* pHPM180, as well as regions highly homologous to another *H. pylori* plasmid pHPK255. In addition a short stretch of near identity to the chromosomal *cagA* gene (*c*ytotoxin *a*ssociated *g*ene) was detected.

The results also indicated that the ORF1 encoded protein was homologous to replication proteins found in nine other plasmids isolated from various organisms, including *Campylobacter*, *Pediococcus*, *Pseudomonas*, *Lactococcus*, *Neisseria*, and *Klebsiella*. Similarly to these other plasmids, pHPM179 contained four 22 bp iterons, i.e. direct repeats that are thought to be the binding site for these replication proteins. Based on this analysis, pHPM179 was predicted to replicate via a theta type mechanism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Helicobacter pylori (previously designated *Campylobacter pylori*) is a curved, 3.5 μm long and 0.5-1 μm wide bacterium with 1-6 sheathed flagella at one end. It is a Gram negative, microaerophilic bacillus that grows best in an atmosphere containing 5% CO_2 at 37^0C, and is found in the stomach of patients with active chronic gastritis and duodenal ulcers. Since its discovery by Warren and Marshall in 1982 (68) and successful isolation by Marshall et al. in 1984 (42), clinical, histological, and bacteriological investigations have been conducted world wide in an attempt to determine the role of the bacteria as a causative agent in gastroduodenal diseases. *H. pylori* is recognized as the etiological agent of active gastritis and peptic ulceration (6).

H. pylori is most likely transmitted from person to person (61), although the organism has not been isolated from food, water, or animals with which humans typically come in contact. Transmission of the bacteria is believed to be via the oral route, and the risk of infection is known to increase with age (66). The evidence of person-to-person transmission comes from clustering within families (15) and from preliminary reports of higher than expected prevalences in residents of custodial institutions (33) and nursing homes (30). *H. pylori* has not been isolated from stool, although it was isolated from the oral cavity (56).

Antibiotic susceptibility testing indicates that most *H. pylori* isolates are sensitive to tetracycline, erythromycin, kanamycin, gentamicin and penicillin and are resistant to nalidixic acid (6).

H. pylori infections can be treated successfully with various anti-microbial agents (43). Therapy with amoxicillin plus metronidazole, without omeprazole, has a 70% cure rate. Amoxicillin/omeprazole therapy has a cure rate of 75% to 85% (40), but the cure rate can be enhanced by the addition of metronidazole to the amoxicillin/omeprazole combination. The combination of high-dose omeprazole with clarithromycin results in an 80% eradication rate (22, 37).

Patients who had been treated with the triple therapy of metronidazole, amoxicillin, and bismuth, resulting in documented eradication of *H. pylori* and cure of the ulcer for at least 4 years, were recalled and of the 94 patients restudied, with a follow-up period range of 48-96 months, only two (2.2%) were again *H. pylori* positive. This gives an effective reinfection rate of only 0.36% per patient year (5).

The presence of *H. pylori* in the gastric mucosa is strongly associated with chronic active gastritis. *H. pylori* gastritis may progress to chronic atrophic gastritis over several decades, and this lesion may eventually be a precursor of gastric carcinoma (27). Several epidemiological studies indicate that *H. pylori* infection may be a risk factor for gastric carcinoma (50, 60). The presence of *H. pylori* also was reported in a large number of gastric lymphomas of the mucosa associated lymphoid tissue (MALT lymphomas); *H. pylori* was detected in 92% of cases (14, 71).

Several properties commonly associated with *H. pylori* are suspected to play a role in the pathogenic process of gastritis as well as ulcer formation. *H. pylori* strains produce a high level of urease, an enzyme that is responsible for hydrolysis of urea (16, 19, 67). Presumably, the production of ammonia from urea allows *H. pylori* to survive the acidic environment and penetrate the mucus lining (17, 41). Urease production is a common trait in most *H. pylori* strains so far isolated (26) and it is chromosomally encoded (9).

In addition *H. pylori* strains have high catalase activity that protects against the damaging effects of oxygen metabolites. Such activity could possibly protect the bacterial cells from endogenous hydrogen peroxide produced by polymorphonuclear leukocytes. This too could be considered a virulence factor (25).

H. pylori strains produce a mucin protease that degrades the polymeric structure of the gastric mucus glycoprotein. Such activity could allow *H. pylori* cells access to the stomach epithelial layer (55, 58).

Other potential virulence factors include production of a weak cell-free hemolysin that is produced by some strains of *H. pylori* and found to be active on human, horse, rabbit and sheep erythrocytes (69). Also, *H. pylori* isolates express hemagglutinins (47), a property that correlates with adhesion to the gastric epithelial layer (26).

Several investigators reported cytotoxin activity in *H. pylori* isolates (18, 28, 35, 70). An 87 kDa extracellular protein, called vacuolating cytotoxin, induces the formation of intracellular vacuoles in eukaryotic cells *in vitro*, probably via an autophagic mechanism (11, 12, 34). Expression of the vacuolating cytotoxin activity is found in 50-60% of *H. pylori* isolates (*vac*$^+$ strains), and it has been associated with duodenal ulcer disease (20).

In view of the partial homology between the amino-terminal sequence of the 87 kDa protein and the sequence of various ion-transporting ATPases, Cover and Blaser speculate that the cytotoxin protein may interact with the function of parietal cell (H^+, K^+)-ATPase and thereby alter gastric acid secretion (11). Further studies show that bafilomycin A1, a specific inhibitor of eukaryotic H^+-ATPases prevents the cytotoxin mediated vacuolation of HeLa cells, suggesting that an ATPase proton pump is the direct or indirect target of the cytotoxin in epithelial cells (49).

It has been postulated (10, 64) that expression of vacuolating cytotoxin activity is linked to the presence of a 128 kDa major surface antigen of *H. pylori*, encoded by the *cagA* gene (cytotoxin associated gene A). A recent report however, shows that the *cagA* gene product does not affect the cytotoxin activity (65) and thus, the relationship between the toxin and the *cagA* gene product remains unknown.

In many bacteria, plasmid mediated virulence factors play essential roles in bacterium-host interactions and pathogenesis (36). Plasmids may encode virulence factors such as enterotoxins, antibiotic resistance factors, heavy metal resistance and proteins involved in cell adhesion.

Plasmids have been visualized in *H. pylori* and it was initially thought that strains could be typed based on plasmid content. Early studies found that 58% of clinical strains contain one or more plasmids ranging in size from 1.8 kb to 40 kb (57, 58, 62); Minnis et al. (in press) found plasmids in 80% of isolates, probably because of an improved plasmid isolation method (44, 45). Since plasmid size and restriction profiles vary, most investigators do not use either chromosome or plasmid restriction profiles for strain typing (51). Although the restriction enzyme profiles of *H. pylori* plasmids differ among isolates from different patients they are mostly the same within one individual. This information has been useful in following the recurrence patterns of *H. pylori* in the same individual with gastritis or duodenal ulcers (57).

The DNA sequence of a 1.5 kb plasmid (pHPK255) was reported (31). One open reading frame was identified which could encode a 25 kDa protein similar to a plasmid replication protein found in bacteria that replicate by a rolling circle mechanism (RCR). Using low stringency hybridization and PCR analysis the authors reported that plasmids

from some strains of *H. pylori* show no homology with the replication regions of pHPM255 and propose that these *H. pylori* plasmids replicate via a different mechanism.

Minnis (46) determined the DNA sequence of an additional *H. pylori* plasmid, pHPM180 (3.5 kbp). It contains two open reading frames (ORFs), neither of which shows significant homology with the 25 kDa protein of pHPK255. ORF1 and ORF2 could encode proteins of 54.5 kDa and 28 kDa, respectively. Nucleotide sequence analysis of pHPM180 DNA revealed a 232 bp direct repeat interrupted by 39 bp of non repeating DNA, and also five short direct repeats of 22 bp each. These short direct repeats are usually found in plasmids that replicate by a theta type mechanism.

This study reports the sequence of a larger *H. pylori* plasmid, pHPM179, which is approximately 5.6 kb. One 684 bp *Hin*dIII fragment had previously been cloned and its sequence determined (8). Analysis of this fragment showed that it probably came from the middle of an ORF. The goal of this project is to complete the DNA sequence analysis of pHPM179 and to determine whether additional proteins might be encoded by this larger *H. pylori* plasmid.

CHAPTER II

MATERIALS AND METHODS

Materials

Strains and vectors

H. pylori strain HPM179 was originally isolated from an antral biopsy from a patient with an active duodenal ulcer at the VA Medical Center in Dallas. A frozen culture of this bacterium in trypticase soy broth containing 25% glycerol and 10% horse serum was provided by Dr. S. McIntire. HPM179 was cultured under microaerophilic conditions (5-12% CO_2) on brain heart infusion (BHI) medium with 10% horse serum at 37^0C.

E. coli DH5αF' is *supE*44, Δ *lacU*169 *(ϕ80dlacZ*Δ M15*) hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1 and contains F' *traD*¯, *proAB*. This bacterium was obtained from Dr. J. Knesek and it was cultured at 37^0C on NZY agar medium .

E. coli DH5αMCR is F¯ *mcrA*, Δ*(mrr-hsdRMS-mcrBC)*, *ϕ80dlacZΔM*15, Δ *(lacZYA-argF)*, Δ *lacU*169, *deoR*, *recA*1, *endA*1, *supE*44, λ¯, *thi*-1, *gyrA*96, *relA*1. These competent cells were used as initial transformation recipients because they are suitable for cloning methylated DNA (DNA with methylated cytosine or adenine residues). In this MCR strain, obtained from Dr. J. Knesek, the systems that restrict DNA containing methylated cytosine and adenine residues (*mcrA*, *mcrB*, *mcrC*, and *mrr*) have been eliminated (4). When plated on selective medium containing X-gal as substrate, cells

containing recombinant plasmids produce colorless colonies due to interruption of the β-galactosidase gene in the vector; cells with nonrecombinant plasmids produce blue colonies.

The pTZ19R vector contains a T7 RNA polymerase promoter within the *lacZ* gene, adjacent to the polylinker sequence of pUC19. It also contains the intergenic (origin) region of phage f1, oriented in the "R" orientation, and the β-lactamase gene which confers ampicillin resistance. Infection with helper phage will produce "virions" containing single stranded DNA with the inverse complement. The helper phage M13KO7 (obtained from Pharmacia) is required for propagation of pTZ19R as single-stranded DNA (ssDNA) viral particles.

Primers

Initial sequencing primers were Universal primer (-40 forward, 24-mer) supplied in Sequenase Version 2.0 kit (USB) and Reverse primer (-50 reverse, 21-mer) obtained from Pharmacia. Additional primers for sequencing and PCR were obtained from Bio-Synthesis Inc., Lewisville, Texas.

Enzymes

Enzymes utilized were T4 DNA ligase (USB), lysozyme (Sigma), Sequenase Version 2.0 DNA polymerase (USB), Pyrostase thermostable DNA polymerase (Bio-Synthesis), and restriction endonucleases (Gibco BRL and Promega).

Commercial Kits

The commercial kits were Elu-Quik DNA purification system (Schleicher & Schuell), Magic Maxipreps DNA purification system (Promega), Super Comp media and TXN salts (Bio101), and Sequenase Version 2.0 DNA sequencing kit (USB). These kits were used as specified by the manufacturers.

Hybridization

Solutions and equipment for hybridization were QuikHyb solution (Stratagene), [α^{32}P]-dCTP (Dupont), Nick translation system (BRL), PhotoGene 0.45 mm nylon membrane (BRL), Sephadex G50 Nick Columns (Pharmacia), PosiBlot pressure blotter and UV Stratalinker model 1800 by Stratagene, Hybidization incubator model 310 (Robbins Scientific), and Hyper film-MP (Amersham).

Media

NZY medium (Gibco BRL), YT medium (Bio 101), and BHI (DIFCO) were prepared according to the manufacturers recommendations. In addition BHI medium contained 5-10% horse serum (Sigma). B broth contained 10 g tryptone (Difco), 10 g NaCl, and 1 ml 1% thiamine in 1 L ultrapure H_2O.

Buffers and other reagents

Buffers and solutions used were TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0), TE (10 mM Tris, pH 7.5, 1 mM Na_2 EDTA), lysis buffer (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 M LiCl, 50 mM Tris-HCl, pH 8.0), solution A (50 mM Tris-HCl, pH 8.0, 4% Triton X-100, 2.5 M LiCl, 62.5 mM EDTA), 5X ligation buffer (439 mM Tris-HCl, pH 7.6, 43 mM MgCl2, 66 mM DTT, 439 mM ATP, 133 mg/ml PEG, 33 mg/ml BSA), PEG solution (20% PEG-8000, 3.5 M ammonium acetate), TXN Salts (Bio 101), 1% thiamine (vitamin B_1), and plasmid reaction buffer (400 mM Tris-HCl, pH 7.5, 100 mM MgCl2, 250 mM NaCl).

Other reagent grade chemicals were 100%, 95%, 80%, and 70% ethanol, 1.5 M sodium acetate, 1 mg/ml ethidium bromide, Tris-HCl buffered phenol/chloroform (1:1), 3 M sodium acetate, methanol, acetic acid, agarose, 10% X-Gal, 400 mM IPTG, isopropanol, 7.5 M ammonium acetate, acryl/bis (19:1) solution (Amresco), TEMED (Amresco), ammonium persulfate (Bio-Rad), and [α^{35}S]-dATP and [α^{32}P]-dCTP (Dupont). Antibiotic stock solutions were kanamycin, 50 mg/ml and ampicillin, 100 mg/ml.

Methods

Centrifugations were carried out in a Beckman JS-HS centrifuge using either a JS-7.5 swinging bucket rotor (for 15, 50 and 250 ml tubes) or a JA-18.1 fixed angle rotor (for microcentrifuge tubes).

Plasmid preparation

The pHPM179 DNA was purified from *H. pylori* isolate HPM179 using a modification (46) of the alkaline/SDS extraction procedure of Birnboim and Doly (3). This modified procedure increases the lysozyme concentration of the preparation to

30 mg/ml, adds a LiCl RNA precipitation step, and includes a final purification step using the S&S Elu-Quik DNA purification kit.

Restriction enzyme analysis

Restriction enzymes whose sites are found in the multicloning region of pTZ19R vector were tested to determine which would digest pHPM179 DNA. *Hin*dIII enzyme produced four fragments and was used to produce linear fragments with ends compatible to the *Hin*dIII digested vector pTZ19R. These digestion mixtures contained 100 ng of plasmid DNA with 1 µl of restriction enzyme (~6 units) and 2 µl of 10X restriction buffer, in a total volume of 20 µl, and were incubated at 37^0C overnight.

Digested DNA was analyzed on a 0.7% agarose gel. Electrophoresis was carried out at 7-10 volt/cm for 1.5 h in TBE buffer containing 0.25 mg/ml ethidium bromide . The DNA fragments were visualized on a long wave UV transilluminator (302 nm) and photographed using type 57 film in a Polaroid MP-3 camera equipped with an orange # 15 filter.

In preparation for ligation of vector DNA (pTZ19R) to insert DNA (pHPM179), 5 µg of pTZ19R was digested with 1 µl of *Hin*dIII (12 unit/µl), 5 µl 10X restriction enzyme buffer, and 2 µl of shrimp alkaline phosphatase (SAP) at 5 unit/µl in a final volume of 50 µl for 60 min at 37^0C. This digestion resulted in linearized vector that should not religate because the 5' phosphates have been removed. Insert DNA was digested as follows : 600 ng of pHPM179 DNA was combined with 1 µl of *Hin*dIII (10 unit/µl), and 2 µl of 10X restriction buffer in a final volume of 20 µl and incubated overnight.

The restriction enzyme and SAP were inactivated as follows: each digest was incubated for 15 min at 65^0C, extracted once with an equal volume of phenol/chloroform

(1:1), and centrifuged in Beckman J2-HS centrifuge at 12,000 rpm, 4^0C, for 5 min. The aqueous phase was removed and placed in a clean tube. DNA was precipitated with 0.1 volume of 3 M sodium acetate and 2.5 volumes of 100% ethanol. Tubes were placed at -70^0C for 30 min, then centrifuged at 9,000 rpm and 20^0C for 20 min. Pellets were washed twice with 70% ethanol, dried in a JOUAN vacuum evaporator, and dissolved in ultrapure water such that vector DNA was at a final concentration of ~50 ng/µl and insert DNA was at final concentration of ~15 ng/µl.

Ligation reaction

Vector DNA was ligated to insert DNA as follows: 80 ng digested vector DNA, an equimolar amount of insert DNA (122 ng), 4 µl of 5X ligation buffer, 1 µl T4 DNA ligase (1unit/µl), and ultrapure water to 20 µl were incubated at 20^0C for 2 h and then stored at -20^0C until ready for transformation reactions.

Preparation of competent cells

E. coli DH5αMCR and *E. coli* DH5αF' were grown in 50 ml Super Comp medium at 37^0C with shaking to an OD_{600} of 0.6. The bacterial cells were made competent for transformation by placing 25 ml of the suspension into a 50 ml conical tube and centrifuging at 6,000 rpm, 4^0C, for 15 min. The pellet was suspended in 15 ml 1X TXN salts and incubated on ice for 30 min. After incubation, the ice cold mixture was centrifuged at 6,000 rpm, 4^0C, for 15 min. The resulting pellet was suspended in 5 ml 1X TXN and was ready for transformation.

Transformation and screening of recombinant plasmids

The transformation protocol was as follows: 10 µl of ligated DNA (or control DNA) was added to 200 µl DH5αMCR competent cells. The mixture was incubated on ice for 45 min and heat shocked for 2 min at 42^0C. To this tube 300 µl of 2X YT was added and the tube was incubated at 37^0C for 1 h without shaking. This incubation allowed the transformed cells to begin expressing β-lactamase. After 1 h the cells were plated by spreading (using a glass rod) onto B broth plates containing ampicillin (50 µg/ml), IPTG (0.1 mM) and X-Gal (0.02%). Different volumes were plated on separate plates (10 µl, 25µl, 50 µl, 100 µl, 200 µl) that were incubated 24 h at 37^0C until small colonies appeared. (Note: Plating at too high a density produces "satellite " ampicillin sensitive colonies in dense portions of the plate, especially with prolonged incubation at 37^0C). After 24 h incubation, the plates were screened for white colonies.

Well separated white colonies were picked by loop for plasmid purification: Five milliliters of NZY medium was added to a 15 ml polypropylene tube and inoculated with a single white colony. This was incubated overnight at 37^0C with shaking. Two milliliters of the bacterial suspension were centrifuged at 4^0C and 9000 rpm for 5 min, suspended in 200 µl of solution A and 200 µl of Tris-HCl.buffered phenol/chloroform (1:1), vortexed and microfuged for 5 min. Supernatants were removed, the pellets were washed with 70% ethanol and vacuum dried. To complete the plasmid preparation, the dry pellets were suspended in 10 µl of ultrapure water and 1 µl of 1 mg/ml RNase A and incubated at room temperature for 5 min.

Each plasmid DNA isolate (2 µl) was treated with *Eco*RI, which was expected to linearize the plasmid and the digested samples were then compared to 2 µl aliquots of undigested plasmid DNA on 0.7% agarose gels. A 2 µl aliquot of each plasmid DNA

showing the expected insert size was digested with *Hin*dIII and analyzed on a 0.7% agarose gel. Each plasmid DNA containing *Hin*dIII fragments of pHPM179 was transformed into competent DH5αF' cells, as described above. These derivatives served as a source of either single or double stranded DNA for sequence determination. The plasmids containing *Hin*dIII fragments of pHPM179 were given the prefix pHPJQ.

PCR amplification and cloning

PCR amplification was done as described by Saiki et al. (53), using an automatic thermal cycler (Techne model PHC-2). The mixture contained 750 fmol template DNA (pHPM179), QAS-40 and QAS-41 oligonucleotide primers (flanking the 1.1 kb *Hin*dIII fragment) at 100 pmol each, four dNTP mixture (200 mM each), 1X reaction buffer (Bio-Synthesis), and 2.5 units Pyrostase (thermostable DNA polymerase from Bio-Synthesis), and H_2O to a final volume of 100 µl. The reaction mixture was overlaid with 50 µl of mineral oil to prevent evaporation. The amplification cycle consisted of an initial denaturation of target DNA at 94^0C for 45 sec, primer annealing at 55^0C for 60 sec, and the final polymerization cycle at 70^0C for 90 sec. The sample was amplified through 25 consecutive cycles. The PCR reaction product was analyzed by electrophoresis on a 0.7% agarose gel containing 0.5 mg/ml ethidium bromide and DNA bands were visualized by excitation under UV light. Fragments of the expected 1.1 kb size were extracted from the agarose gel and purified using the Elu-Quik DNA purification system (Schleicher & Schuell). The PCR dsDNA product was digested with *Hin*dIII and purified with Elu-Quik. Equimolar amounts of this DNA and 100 ng of *Hin*dIII digested pTZ19R vector DNA were ligated and transformed into DH5αF'. To eliminate the possibility that misincorporation(s) occurred during the early rounds of PCR amplification, three

independent recombinant clones were isolated and the plasmid DNAs were purified and sequenced.

Southern hybidization

Hybridization was performed by using general procedures as recommended by Sambrook et al. (54) and commercial kits. First, 500 ng of plasmid DNA (pHPJQ38, pHPJQ7, pHPM376) and pTZ19R vector as control were digested with 1µl (~10 units) *Hin*dIII and 2 µl 10X restriction buffer in a total volume of 20 µl at 37^0C for overnight. The digest was analyzed on a 0.7% agarose gel at 7-10 volt/cm for 1 h in TBE buffer containing 0.5 mg/ml ethidium bromide. The gel was visualized over UV transilluminator and photographed using Polaroid type 57 film.

Following electrophoresis the gel was exposed to UV in the Stratalinker to nick the ethidium bromide-stained DNA in the agarose gel. Then the DNA was denatured with 0.5 N NaOH and 1.5 M NaCl solution for 15 min at room temperature with gentle shaking, followed by neutralization with 1 M Tris, pH 7.5, containing 1.5 M NaCl for 15 min with gentle shaking.

The membrane was pre-wetted by soaking in H_2O for 20 min and then in 25 mM sodium phosphate, pH 6.5 for 10 min. The DNA was transferred from the agarose gel to the nylon membrane using a Stratagene pressure blotter (PosiBlot), as recommended by the manufacturer .

The membrane was removed from the device and placed on clean Whatman 3mm paper to remove excess buffer. Once the membrane was free of standing liquid, but not dry, the membrane and Whatman paper were placed in the Stratalinker UV crosslinker and crosslinked to the membrane with 1200 µJ of energy. The membrane was placed in a

50 ml conical tube with 5 ml of hybridization solution (QuikHyb by Stratagene) for 3 h at 68^0C.

Labeling the probe: The DNA was labeled with ^{32}P using the Nick Translation procedure (BRL). In this procedure the DNA is incubated with DNase that will nick the dsDNA randomly such that the pieces will act as primers that polymerase can elongate. The reaction mix contained 5 µl of solution A2 containing dATP, dGTP, dTTP, 1 µg of pHPM179 DNA, 8 µl of [α^{32}P]-dCTP (3000 Ci/mmol), 5 µl of solution C (DNA polymerase + DNase) in a final volume of 50 µl. The mixture was incubated at 15^0C for 60 min. The volume was adjusted to 100 µl with H_2O, then passed through the Sephadex G-50 nick columns. In the column, equilibration buffer was allowed first to completely enter the gel bed, then the sample (100 µl volume) was added, followed by 400 µl H_2O. A test tube was placed for sample collection under the column and the purified sample was eluted with 400 µl H_2O. Assuming recovery of 1µg of labeled probe, the specific activity of the sample was calculated as follows: 2 µl of the sample (labeled probe) was added to 8 ml of Beta phase solution in a small vial and counted in a Beckman Liquid Scintillation Counter (model LS9000). With a specific activity of 4.9×10^7 cpm/µg, 30 µl of the labeled probe was heat denatured for 2 min at 100^0C, cooled quickly on ice, added to the hybridization solution and incubated at 68^0C for 15 h.

The membrane was washed 2 times with 25 ml of 2X SSC, 0.1% SDS at room temperature for 15 min and one last time at 68^0C for 30 min. The membrane was dried, exposed to X-ray film for 3 h, and the film was developed.

Preparation of double stranded DNA for sequencing

For double stranded sequence determination, DNA from the pHPJQ recombinant plasmids was prepared using Magic Maxipreps DNA purification system (Promega). For determining the DNA sequence from dsDNA by the dideoxy method the plasmid DNA must first be denatured. The alkaline denaturation method takes advantage of the fact that covalently closed circular DNAs will form special "collapsed" structures when denatured and do not readily re-anneal. Also, denaturation with alkali is favored over heat denaturation, because most circular plasmid DNAs have melting temperatures above 100^0C.

The method recommended by USB Corp. involved the addition of NaOH directly to a mixture of purified plasmid DNA and primer, denaturing the double-stranded DNA (63). After a short incubation at 37^0C, an equimolar amount of HCl was added to the mixture, thereby neutralizing the alkali. Plasmid reaction buffer was added to fix the pH at an appropriate value. The mixture was then incubated briefly at 37^0C to allow the primer to anneal to the appropriate sequence within the template. The NaOH and HCl combined to form NaCl which is a normal component of the DNA sequencing reaction mixture. As long as the concentration of NaCl is kept below approximately 0.2 M, the polymerase works well and high-quality sequence is obtained. Specifically the DNA was prepared by mixing template DNA (1 pMol), 1.0 M NaOH (2 µl), primer (2 pMol), and H_2O to 11 µl and incubating for 10 min at 37^0C. The mixture was placed on ice and 1.0 M HCl (2 µl) and plasmid reaction buffer (2 µl) were added. The DNA was now ready for the annealing step in USB Sequenase DNA sequencing protocol.

Preparation of single-stranded plasmid DNA

The DNA was prepared by inoculating 5 ml of 2X YT medium containing 50 μg/ml ampicillin with a fresh single colony of DH5αF' containing the desired plasmid and incubated overnight at 37^{0}C with shaking. One ml of this overnight culture was used to inoculate 50 ml of 2X YT broth containing 50 μg/ml ampicillin in a 250 ml flask. After incubation for 30 min at 37^{0}C with vigorous agitation, 133 μl of M13K07 phage stock was added. The multiplicity of infection (moi) was about 20 pfu per cell. Shaking was continued at 37^{0}C for 30 min and then 70 μl of 50 mg/ml kanamycin was added (final concentration 14 μg/ml) followed by continued shaking for 14 to 18 h. The culture was centrifuged in a JA-14 rotor for 15 min at 17,000 rpm, 4^{0}C, and the supernatant, which contained the phage-like single-strand DNA particles, was retained. The supernatant was centrifuged again to remove any residual cells.

One quarter volume of a solution containing 20% polyethyleneglycol (PEG 8,000) and 3.5 M ammonium acetate was added. The solution was mixed gently, stored on ice 30 min, and centrifuged 15 min,17,000 rpm, 4^{0}C, to collect the precipitated phage particles. The supernatant was removed and the pellet was suspended in 0.2 ml TE buffer and then transferred to a 1.5 ml microcentrifuge tube. The DNA was extracted by adding 0.1 ml of buffer-washed phenol and 0.1 ml of chloroform; the solution was mixed on a vortex mixer 1 full min and then centrifuged 5 min to separate the phases. The top (aqueous) phase containing the DNA was removed to a fresh tube. This extraction was repeated five to six times and the amount of interface material was diminished until there was almost none seen in the final extraction. The final aqueous phase was extracted 4 times with 0.2 ml chloroform to remove the last traces of phenol. The DNA was precipitated by adding 0.1 ml of 7.5 M ammonium acetate and 0.6 ml of 100% ethanol,

chilled to -70^{0}C (dry ice / ethanol bath) for 30 min, and centrifuged for 15 min at 4^{0}C. The supernatant was drained and the pellet was washed with 70% ethanol, vacuum dried, and dissolved in 20 µl of distilled water. The DNA was stored in a fresh tube at 4^{0}C. This DNA was ready for sequence determination using Sequenase V.2.0 (USB).

Sequencing reactions

DNA sequence determination of both strands of the recombinant clones was accomplished using the Sequenase Version 2.0 kit and [α^{35}S]-dATP. Initial priming reactions utilized the universal primer provided in the kit. This primer corresponds to a sequence upstream of the multicloning site of pTZ19R and provided sequence information into the cloned fragments.

Subsequent primers were 20 to 22 base oligonucleotides with sequences complementary to the end portions of the sequence previously determined. Thus, the base sequence of each strand was determined in an overlapping fashion (since only approximately 300 bases were sequenced with each primer). Reactions were analyzed on 6% polyacrylamide gels [15 ml Acryl/Bis (19:1) solution, 20 ml 5X TBE, 48 g urea, ultrapure H_2O to a final volume of 100 ml. TEMED (30 µl) and 1 ml 10% ammonium persulfate were added and the mixture was poured into CBS glass plates]. Following polymerization, samples were heated at 70^{0}C for 5 min and applied to the gel. Electrophoresis was at 2000 V in 1X TBE buffer in a Model DASG-500-33 nucleic acid sequencing apparatus (CBS Scientific Co.). Multiple loadings of the samples were done at intervals of 2, 4, and 6 h. Gels were vacuum dried and autoradiographed using Kodak XAR X-ray film for 24 h. Films were developed according to manufacturer's specifications.

Analysis of DNA sequence

The DNA sequence of pHPM179 was analyzed for restriction sites, open reading frames, promoter and ribosomal binding consensus sequences, and transcription termination sequences using the DNAsis (Hitachi, Brisbane, CA) and DNA Inspector *IIe* (Textco, W. Lebanon, N. H.) computer programs. Any open reading frames obtained in this manner were compared to protein databases using BLASTP (1). DNA sequence was compared to DNA databases using BLASTN (1).

CHAPTER III

RESULTS

Preparation of *H. pylori* plasmid DNA (pHPM179) and restriction enzyme analysis

Plasmid DNA from *H. pylori* strain HPM179 was purified using a modification of the alkaline/SDS extraction procedure. Several bands were evident (Fig 1, lane 2) in the plasmid preparation, representing different conformations of the plasmid DNA. Purified plasmid DNA (pHPM179) was digested with restriction endonucleases whose sites occur in the cloning vector pTZ19R multicloning site and *Hin*dIII was found to produce fragments of useful sizes. Fig 1, lane 3 shows four fragments of 2.2, 1.6, 1.1, 0.68 kb. Therefore *Hin*dIII was chosen as a suitable enzyme for digestion of both the plasmid and vector DNA in preparation for cloning experiments.

Restriction endonuclease analysis of pHPM179 DNA on agarose gels

It was important for this study to determine the order of the four *Hin*dIII fragments and their orientation in pHPM179. Therefore several restriction enzymes were used: *Eco*RI; *Hin*dIII; *Hin*fI; *Hpa*I; *Eco*RV; and *Pvu*I. The analysis utilized both single and double digestions of pHPM179 DNA. The results are shown in Fig 2 and the map generated from the digestions is shown in Fig 3.

As shown in Fig 2, *Hin*fI digestion (lane 2) resulted in two fragments of 0.8 and 4.8 kb, and *Eco*RI digestion (lane 4) produced only one fragment, indicating one restriction site. When *Hin*fI and *Eco*RI were used together (lane 3), three fragments were generated (0.6, 0.8, and 4.2 kb). This result showed that the single *Eco*RI site was within the 4.8 kb *Hin*fI fragment, digesting it to 0.6 and 4.2 kb.

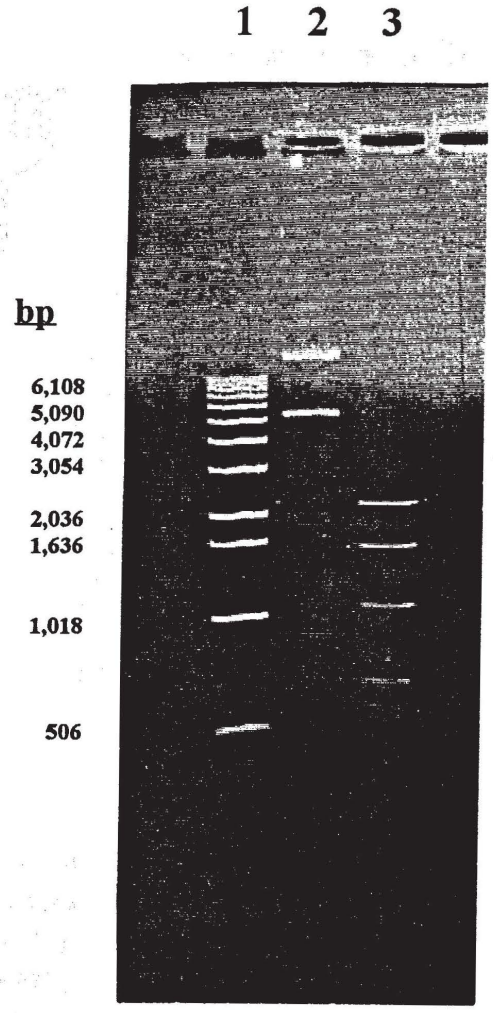

Fig. 1. Restriction digestion of pHPM179 DNA with endonuclease *Hin*dIII. 1, standard DNA (kb ladder); 2, undigested pHPM179 DNA; 3, pHPM179 + *Hin*dIII. Numbers on left represent sizes in bp of the kb ladder standard.

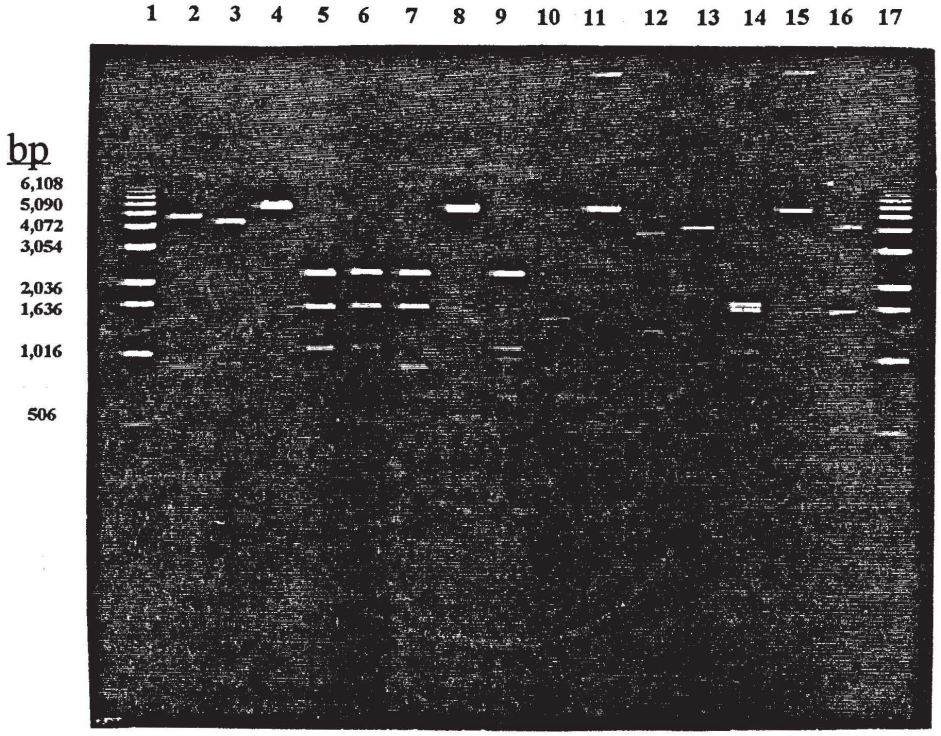

Fig. 2. Single and double digestions of pHPM179 DNA. The enzymes are as follows: lane 1, standard (kb ladder); lane 2, *Hin*fI; lane 3, *Hin*fI + *Eco*RI; lane 4, *Eco*RI; lane 5, *Hin*dIII; lane 6, *Hin*dIII + *Eco*RI; lane 7, *Hin*dIII + *Hin*fI; lane 8, *Hpa*I ; lane 9, *Hpa*I + *Hin*dIII; lane 10, *Hpa*I + *Hin*fI; lane 11, *Hpa*I; lane 12 *Hpa*I + *Eco*RV; lane 13, *Eco*RI + *Hpa*I; lane 14, *Hin*dIII + *Pvu*I; lane 15, *Pvu*I; lane 16, *Hpa*I + *Pvu*I; lane 17, standard (kb ladder). Numbers on left represent sizes in bp of the kb ladder standard.

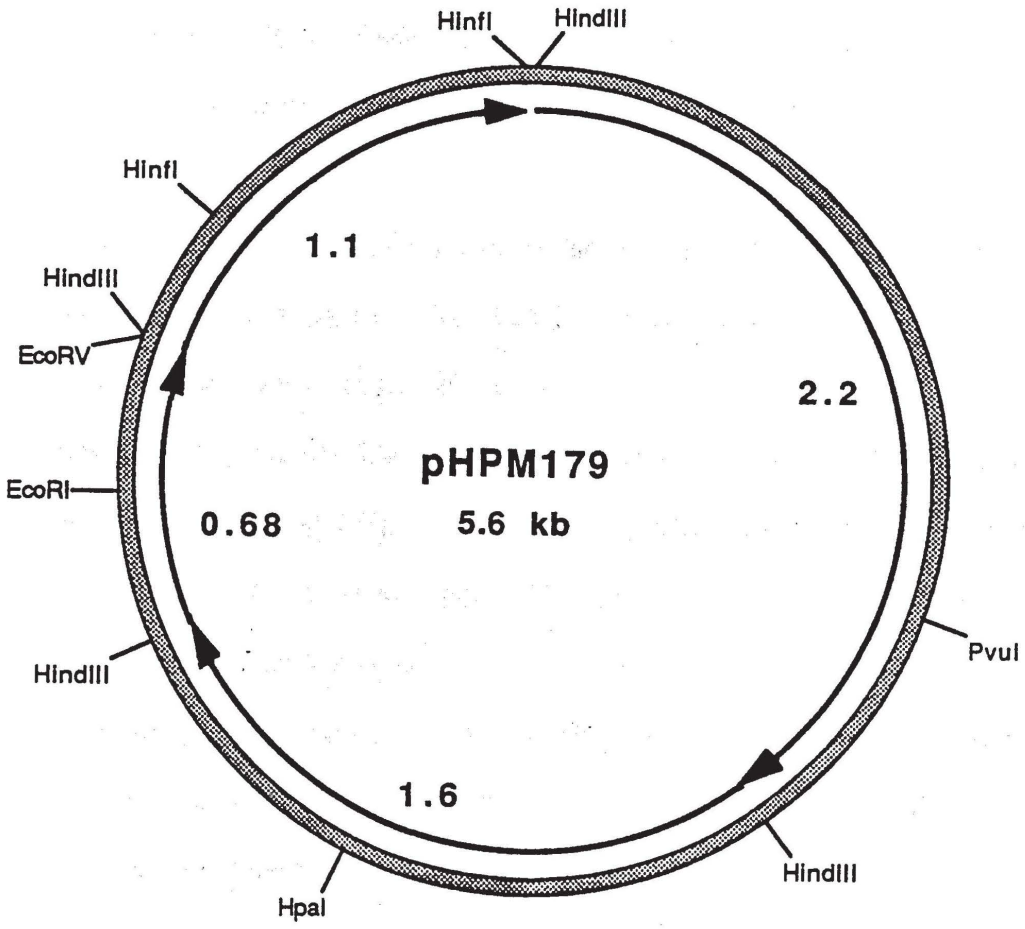

Fig. 3. Restriction map of pHPM179

Digestion with *Hin*dIII only (lane 5) produced four fragments 2.2, 1.6, 1.1 and 0.68 kb. A double digestion with *Hin*dIII and *Eco*RI (lane 6) indicated that the single *Eco*RI site was located in the 0.68 kb *Hin*dIII fragment. Similarly, double digestion with *Hin*dIII and *Hin*fI (lane 7) showed that the two *Hin*fI sites were located within the 1.1 kb *Hin*dIII fragment.

Digestion with *Hpa*I enzyme (lane 8) produced one fragment, again indicating one restriction site. A double digestion with *Hpa*I and *Hin*dIII (lane 9) showed that this *Hpa*I site was located in the 1.6 kb *Hin*dIII fragment.

To determine the relative locations of the *Hin*dIII 1.6 and 0.68 kb fragments, a double digestion with *Hin*fI and *Hpa*I (lane 10) was performed, producing three fragments of 0.8, 1.6, and 3.2 kb. These fragments could only appear if the 1.6 and 0.68 kb fragments were adjacent, as shown in Fig 3.

To determine the orientation of the 0.68 kb fragment, the plasmid was digested with *Eco*RV, which also digested at one site, near one end of the 0.68 kb *Hin*dIII fragment (lane 11). The double digestion with *Hpa*I and *Eco*RV (lane 12) resulted in two fragments of 1.3 and 4.3 kb. To determine the orientation of the 1.6 kb *Hin*dIII fragment, double digestion with *Eco*RI and *Hpa*I was done, producing fragments of 0.97 kb and 4.7 kb (lane 13). Thus, the orientation is as shown in Fig 3.

Digestion with *Pvu*I enzyme (lane 15) produced one fragment indicating one restriction site. A double digestion with *Hin*dIII and *Pvu*I (lane 14) located a *Pvu*I site in the 2.2 kb *Hin*dIII fragment, since the digestion produced fragments of 1.7 and 0.5 from the 2.2 kb fragment. This enzyme also helped determine the orientation of the 2.2 kb fragment. Double digestion with *Hpa*I and *Pvu*I (lane 16) generated two fragments of 1.5 and 4.1 kb, which could occur only if the fragments were as depicted in Fig 3.

Cloning of the *Hin*dIII fragments in pTZ19R vector

Initial attempts to clone and transform pHPM179 fragments into DH5αF' were unsuccessful presumably because *H. pylori* has highly modified DNA (39, 52). Further cloning was attempted using a completely restriction negative transformation recipient, strain DH5αMCR.

Linearized pHPM179 DNA was ligated to phosphatase-treated pTZ19R vector DNA and transformed into competent DH5αMCR as described in Methods. Forty white colonies were picked for screening by restriction enzyme analysis with *Eco*RI to determine clones that contained pHPM179 DNA. One clone (38s) contained an insert of about 3 kb (Fig 4). Clone 38s was digested with *Hin*dIII and two fragments of 2.2 kb and 1.6 kb were detected (Fig 5). Therefore, clone 38s, containing the 1.6 and 2.2 kb *Hin*dIII fragments and renamed pHPJQ38, was transformed into *E. coli* DH5αF' for sequencing.

Amplification of 1.1 kb *Hin*dIII fragment by PCR.

Repeated attempts to recover a recombinant containing the 1.1 kb fragment were unsuccessful. Once the sequence of the neighboring 2.2 kb fragment was determined (see below), part of this sequence was used as one of the primers for PCR amplification of the 1.1 kb fragment. Sequence of the 684 bp fragment, the other neighboring fragment, had been determined by Chen (8). Accordingly two 22 mer oligonucleotide primers that bracket the 1.1 kb fragment were designed (sequence included in Table 1 below), and used to amplify the 1.1 kb fragment. The 3' end of the rightward PCR primer (in the 684 fragment) was located 10 bp from the 1.1 kb fragment and the 3' end of the leftward PCR primer (in the 2.2 kb fragment) was 13 bp from the 1.1 kb fragment. Thus, synthesis from both primers passed through the *Hin*dIII sites bracketing the 1.1 kb fragment. The PCR-

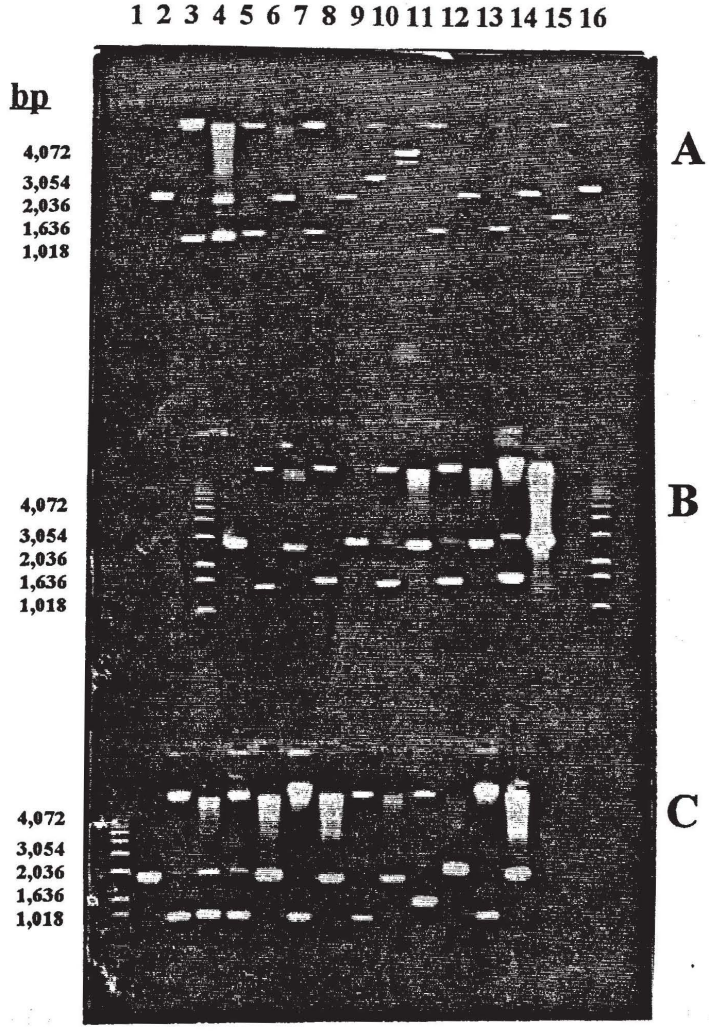

Fig. 4. Screen of recombinant plasmids. Each was examined as undigested DNA and digested with *Eco*RI which should linearize the recombinant. In sections A, B, and C :1, standard; 2, vector + *Eco*RI. The remaining lanes are pairs of each recombinant, undigested and digested. Lanes 9 and 10 section A are clone 38s. Numbers on left are bp sizes of kb ladder standard.

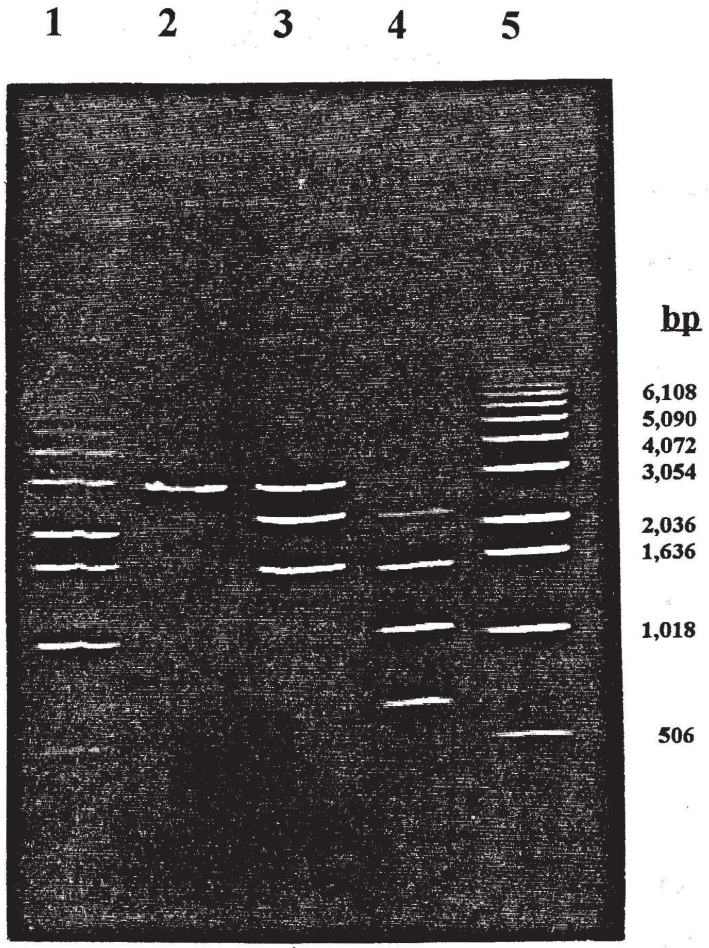

Fig. 5. Restriction digest of clone 38s with *Hin*dIII. 1, standard; 2, vector + *Hin*dIII; 3, 38s + *Hin*dIII; 4, pHPM179 +*Hin*dIII; 5, standard. Numbers on right are bp sizes of kb ladder standard.

dsDNA product was digested with *Hin*dIII to release the 1.1 kb fragment (Fig 6), which was purified, ligated into pTZ19R and transformed into DH5αMCR as described in Methods.

Screening for recombinant colonies with PCR-dsDNA

Seven white colonies were screened for a recombinant plasmid using *Eco*RI, as described previously. One clone, 7s, had an increased size after linearizing with *Eco*RI (Fig 7). This clone (renamed pHPJQ7) was digested with *Hin*dIII and two fragments were seen, one equal to the vector pTZ19R and the other to the 1.1 kb *Hin*dIII fragment of pHPM179 (Fig 8). The sequences of the inserts in two additional recombinant clones, #8 and #9 were analyzed to determine whether misincorporation occurred during amplification. The sequences of these additional recombinants were identical to that determined for pHPJQ7.

Southern blot hybidization

To confirm that pHPJQ38, pHPJQ7, and pHPM376 [recombinant carrying the 684 bp *Hin*dIII fragment (8)] contained the *Hin*dIII fragments of pHPM179, they were digested with *Hin*dIII and hybridized to native pHPM179 DNA. DNA from pTZ19R was linearized with *Hin*dIII and used as a negative control and pHPM179 was linearized with *Eco*RV and used as a positive control. The fragments were separated on a 0.7% agarose gel, shown in Fig 9-A. The native pHPM179 plasmid was nick translated using [α-^{32}P]-dCTP to form a labeled probe. The DNA molecules were transferred from the gel to a nylon membrane and were incubated with denatured probe as described in Methods. Fig 9-B shows that the *Hin*dIII fragments (2.2 kb, 1.6 kb, 1.1 kb, and 0.6 kb) were detected.

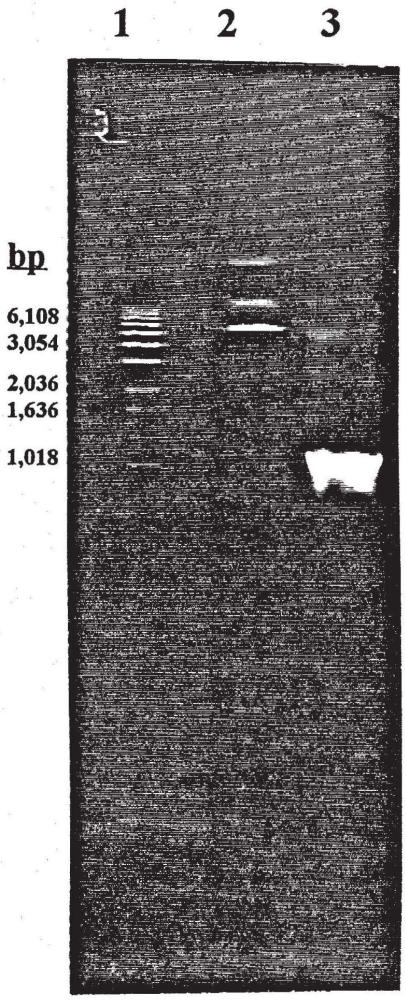

Fig. 6. Amplification of 1.1 kb *Hin*dIII fragment directly from pHPM179 plasmid. Lane 1, standard; 2, pHPM179 undigested; 3, PCR dsDNA product (~1.1 kb). Numbers on left are bp sizes of kb ladder standard.

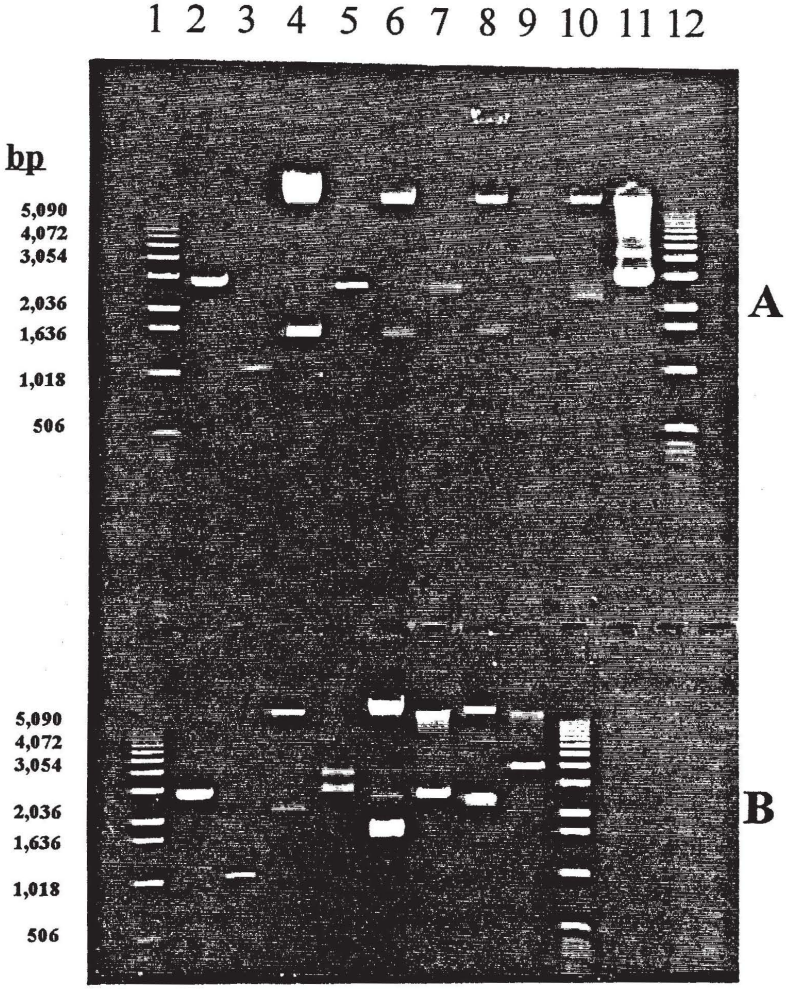

Fig. 7. Screening PCR amplified recombinants. Each clone was examined as undigested DNA and digested with *Eco*RI which linearized the recombinant. In both A and B, 1 is the standard (kb ladder), 2 is pTZ19R vector + *Eco*RI and 3 is PCR-dsDNA. The remaining lanes are pairs of each recombinant, undigested and digested. Lanes 8 and 9 in section B are clone 7s (construct pHPJQ7). Numbers on left are bp sizes of kb ladder standard.

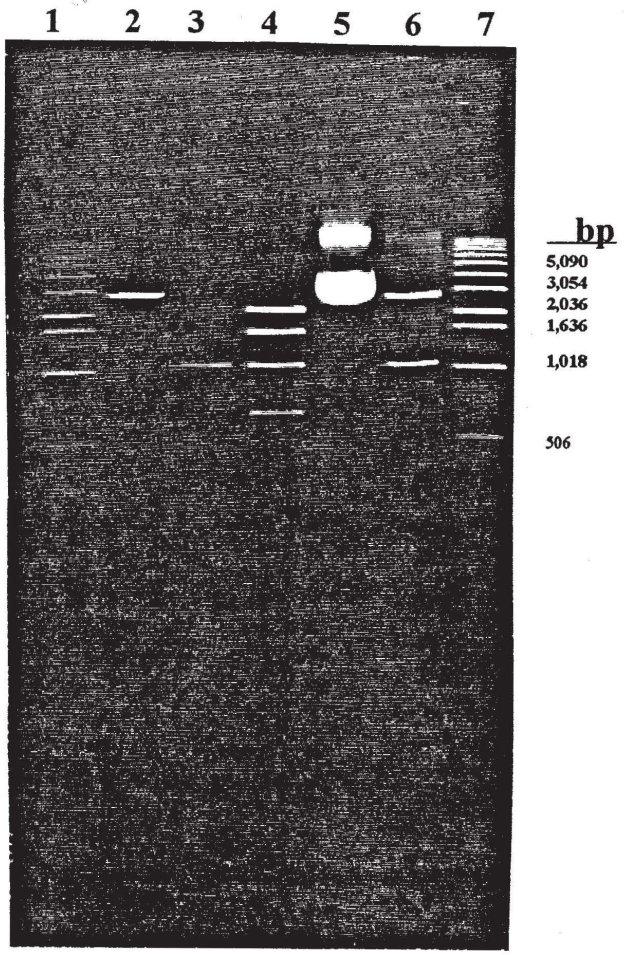

Fig. 8. Restriction digest of clone pHPJQ7 with *Hin*dIII. Lane 1, standard; 2, vector + *Hin*dIII; 3, PCR-dsDNA + *Hin*dIII; 4, pHPM179 + *Hin*dIII; 5, clone pHPJQ7 undigested; 6, clone pHPJQ7 + *Hin*dIII; 7, standard (kb ladder). Numbers on right are bp sizes of kb ladder standard.

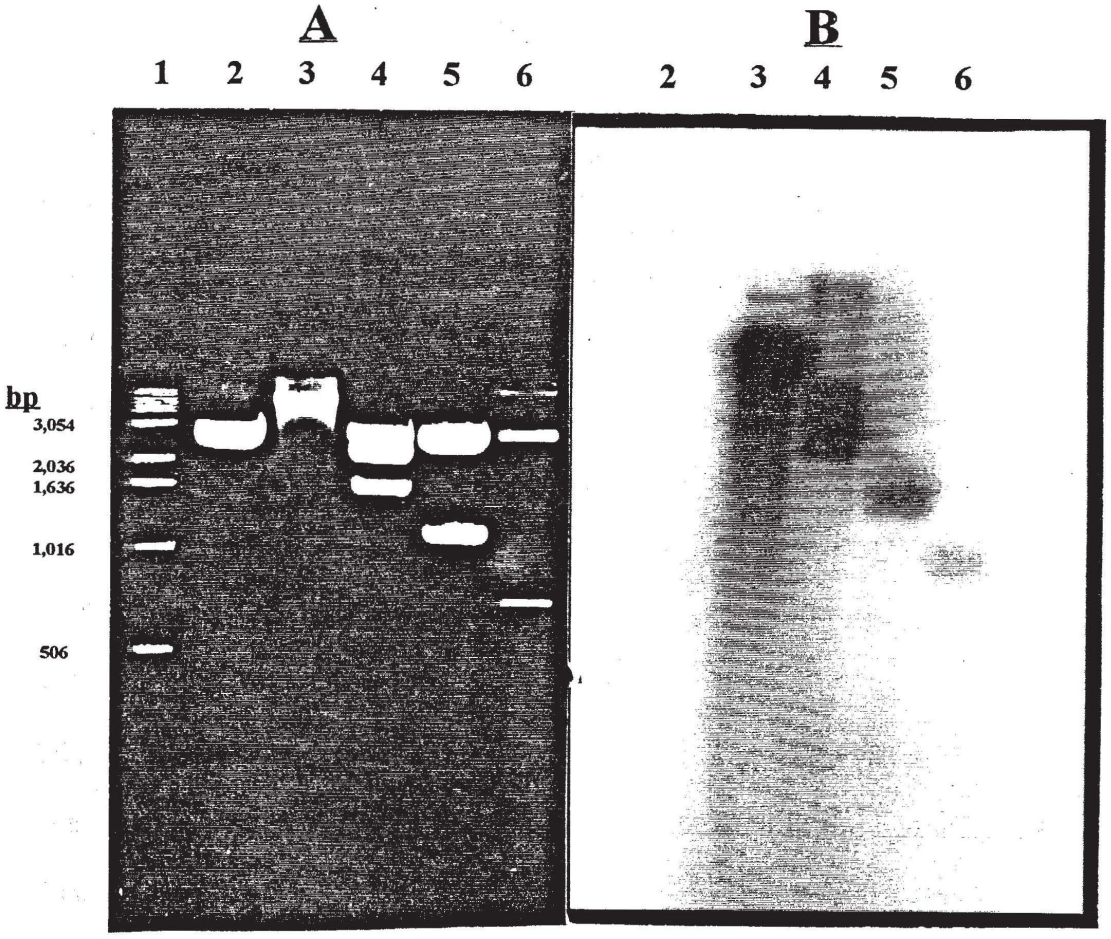

Fig. 9. Hybridization results.
A. Agarose gel. 1, standard (kb ladder); 2, pTZ19R vector + *Hin*dIII; 3, pHPM179 + *Eco*RV; 4, pHPJQ38 + *Hin*dIII; 5, pHPJQ7 + *Hin*dIII; 6, pHPM376 + *Hin*dIII. Numbers on left are bp sizes of kb ladder standard. B. Autoradiogram showing hybridization of pHPM179 probe DNA to DNA transferred from the gel to nylon membrane. 2, pTZ19R; 3, pHPM179; 4, pHPJQ38; 5, pHPJQ7; 6, pHPM376.

This indicated that the fragments cloned in pHPJQ38, pHPJQ7, and pHPM376 were derived from pHPM179.

Determination of nucleotide sequence

Recombinant pHPJQ38 contained both the 2.2 and the 1.6 kb fragments, while pHPJQ7 contained the 1.1 kb fragment alone. Initial nucleotide sequences of these inserts were determined using the universal M13 sequencing primer and reverse primer. Each reaction yielded approximately 250 bases of sequence. Subsequent 19, 20 and 22 mer primers were chosen from a region near the end of each sequenced section of DNA (primer walking). They were chosen based on having a 40-60 % G-C content, and are shown in Table 1. This protocol provided the DNA sequence of both strands as depicted in Fig 10.

All sequences were combined with that determined by Chen (8) producing the 5660 bp total sequence. This is shown in Fig 11, with coordinate 1 at the beginning of the unique *Pvu*I site.

Analysis of DNA sequence

The 5660 bp sequence was analyzed using DNAsis and DNA inspector *IIe* and was compared to GenBank (2) data base sequences using BLAST algorithms (1) as provided by the NCBI (National Center for Biotechnology Information).

DNAsis analysis revealed a G-C content of 36.2 %. Further computer analysis showed three open reading frames, two (ORF1 and ORF2) on one strand and one (ORF3) on the opposite strand. The amino acid sequences of the three ORFs are given below the DNA sequence in Fig. 11. ORF1 begins at nucleotide (n.t.) 1784, and continues through

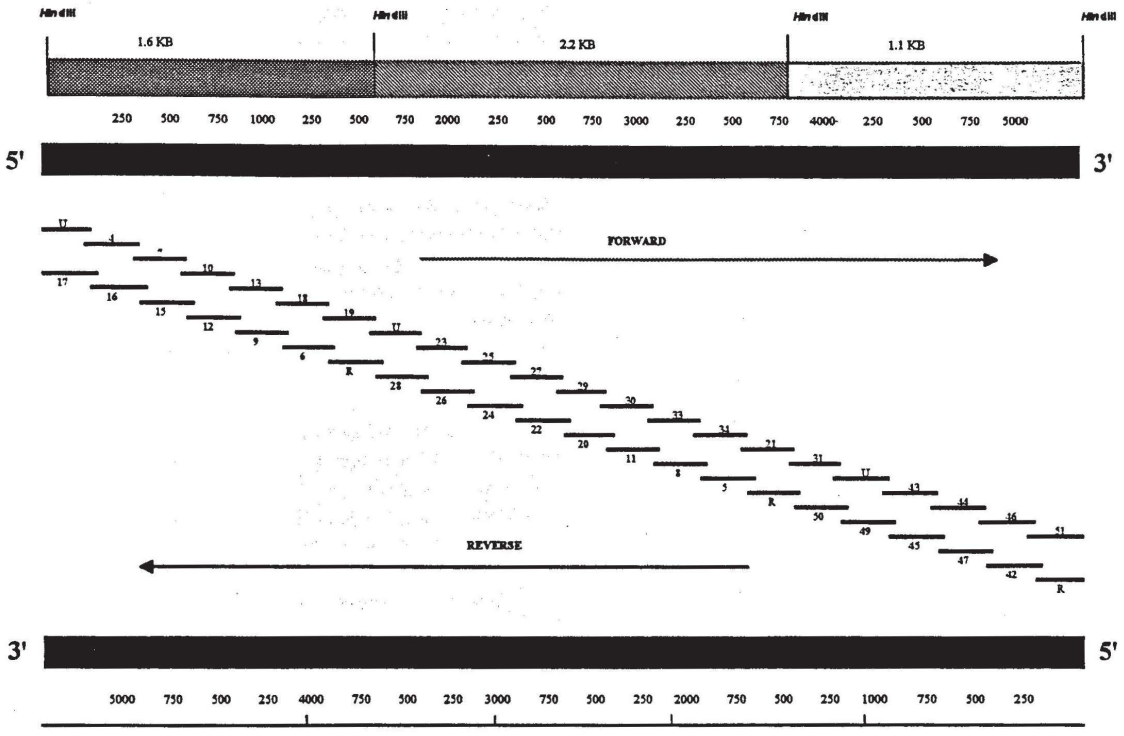

Fig. 10. Strategy used for DNA sequence analysis.
Double stranded DNA of clones pHPJQ38 (1.6+2.2. kb) and pHPJQ7 (1.1kb) are represented. U is the M13/pUC forward sequencing primer and R is the M13 reverse sequencing primer. The numbers associated with the lines designats 19, 20 or 22 mer oligonucleotide primers. The direction of the sequence is indicated by the black arrows.

PRIMER	SEQUENCE	TM0C	G+C%	1st nt pos
Forward (U	GTTTTCCCAGTCACGACGTTGTA	68	47.8	
Reverse (R	TTGTGAGCGGATAACAATTTC	58	38	
QAS-4	GGTGTTATCTTTAGGGGTGCTA	64	45.45	2059
QAS-5	TCGTTTCTGTCGTTCTCTAA	56	40	4100
QAS-6	CACTCCAACCTAATCATCAG	58	45	729
QAS-7	GGGGTTTGTGGGAGCGTATC	64	60	1856
QAS-8	GTAGGGCAGAATGATGAGAA	58	45	4357
QAS-9	CTTCGGCTTACAATCCAACC	60	50	978
QAS-10	GTGGCTGTGTTAAGATTACC	58	45	1651
QAS-11	GGCTGTCTCTAACTCTAGG	58	52.63	4527
QAS-12	CAAGATAGCGCAAACCTGG	58	52.63	1152
QAS-13	GCTCAAAATGGTTAATATGG	54	35	1473
QAS-15	GAAAATGAAAGAATGGTAAC	52	30	1349
QAS-16	GGTTGGGAGTTTGTAGTGGG	62	55	1575
QAS-17	AACAAGAAATACAAGACCGC	56	40	1893
QAS-18	AGTTAGCATTCTTTCATTTT	50	25	1369
QAS-19	GCCTACAACAAACGAAAAAC	56	40	831
QAS-20	GCTTGTATCTTTGTTCTGGC	58	45	4732
QAS-21	ACTAATGCCACTAACTTTCA	54	35	4422
QAS-22	GGTAATTATCTAGCTCTTTG	54	35	4962
QAS-23	GCCCCTAAGAATTAAGAGCG	60	50	602
QAS-24	GTAACGCCCTTGCATTTTTT	56	40	5234
QAS-25	CCCCCATTTCTACATTTTTT	54	35	341
QAS-26	TGATGACTGATTTCTATGAG	54	35	5532
QAS-27	GGCAAACTATCCAAACCCAA	58	45	102
QAS-28	GTTTGTATCCACTTGTTCTC	56	40	162
QAS-29	GGGATAGATTAGGTTTAGAA	54	35	5555
QAS-30	AAAAAATGCAAGGGCGTTAC	56	40	5234
QAS-31	GCAAAGAACAGCAAGATAGC	58	45	4124
QAS-33	CCCACAAGGTTAAGGCTATC	60	50	4993
QAS-34	GCTCCTAACACAACACCAAG	60	50	4682
QAS-40	AACGCACAACACATTATTAGCC	62	40.91	3964
QAS-41	CACAAGCAACGCGCTAAAAGGG	68	54.55	2819
QAS-42	CTATTTTACAGATTGACGAC	54	35	2967
QAS-43	CAAATCAAAGAGCAAGAAAC	54	35	3806
QAS-44	GCGAATTAGAGACTTCTTTC	56	40	3636
QAS-45	ATCCACATCACCAACAACAA	56	40	3213
QAS-46	TTGTATGAAAAATGAAATCT	48	20	3416
QAS-47	CACCAACGAACCCATCAAGG	62	55	3167
QAS-49	GTGAAACTGAATGACAATCG	56	40	3285
QAS-50	GTTCTCTTGTTTCTTGTTCC	56	40	3585
QAS-51	GTTCGTTGGTGATAGGGTGG	62	55	3158

Table 1. Primers used in the DNA sequence analysis of pHPM179. Sequences are shown 5' to 3'. Positions of the first bases of each primer with respect to the double stranded sequence, starting with the first base of the unique *Pvu*I site are shown in the far right column. The melting temperature of the primers is indicated in the third column and the G+C% in the fourth column. Primers QAS-40 and QAS-41 were used in the PCR amplification of the 1.1 kb *Hin*dIII fragment.

```
   1    G CGA TCG CTT TTT GCT TTT TAG ATT TTA TTG GAG TTT GTG AGT TTT      46

  47  GGT GCA AAA GAG AAT TTA AAA CTA GTT GAG AGA AGT TTA AAT TTT TTT      94

  95  CTT GCA TTT GGG TTT GGA TAG TTT GCC ATT GTT CTT TAG ATA GCC TGA     142

 143  GAC TTT TAA TGA TTT GTT TGT TTG TAT CCA CTT GTT CTC TCA TTT GAT     190

 191  TTT TAG GGG GTC AAG GGG GCT TAC CCC ATT GCA AGC TAT TTT GTA TTA     238

 239  CAA AAT ATA TGC TTG CCC TAT TAT TTT TAT TTT AGC ATG GAT TTG GTT     286

 287  AAT TTA GAT ATA ATG GGC TTA CGA TGA TTG TTT AAT ACA AGG AGA TAA     334

 335  G ATG TTA AAA AAT GTA GAA ATG GGG GCT AAT TTT TAT AAA GAG CTT       380
   1     M   L   K   N   V   E   M   G   A   N   F   Y   K   E   L      15

 381  GGC AAA TTA GAA AAA CAA TTA GCT AAA TAT CAA AGT AAA GTT TTA GAA     428
  16   G   K   L   E   K   Q   L   A   K   Y   Q   S   K   V   L   E      31

 429  ATT AAA ACA CAA ATG AAA GAG ATT AAA AAG CGG TAT TCT CAA GCT AAG     476
  32   I   K   T   Q   M   K   E   I   K   K   R   Y   S   Q   A   K      47

 477  AAA GAT GAG AAT TTA GCT AGT AAA TAT ATT CCT AAT GAT GAG TTA AAA     524
  48   K   D   E   N   L   A   S   K   Y   I   P   N   D   E   L   K      63

 525  AAA GAT TTA TTA GAT ATT GAT AAT CCT AAT ACG ACT GAA AGC TTT AAA     572
  64   K   D   L   L   D   I   D   N   P   N   T   T   E   S   F   K      79

 573  CCT AAA AAT GCT AGT GAA ATT TGG CAA TTC GCT CTT AAT TCT TAG GG G     620
  80   P   K   N   A   S   E   I   W   Q   F   A   L   N   S   *          94

 621  CT TGA ATG CTA GAG ATT GAG TTA AAA AAG AAA TTC ACT AAG GAT TTA A     668

 669  AA AAG CAC ATT TTA AAT CAA AAA ATT GAG TTA GAA GTG TTT GAC TTA G     716

 717  TG ATT GAA AAT TCA CTC CAA CCT AAT CAT CAG TTT GTG AAT GAA AGG A     764

 765  TT TTT TCG CTC ATA TTT TCG TTT TGA GCG TTT TTT AGC TTG TAG GTA G     812

 813  TA CAT TAG TCT GTT TTT TGT TTT TCG TTT GTT GTA GGC GAT TTT AGA T     860

 861  AG CAA TAA ACA GCT AAA AAA TCC AAA CAA CCT AAT TGA CTT CAA AAA A     908

 909  CT TCA TTT CCC CAT TAG TTG CTA ACC ATT TAG CCA ATC CCA CTT GTT T     956

 957  AG CAT CTA AGA GCG CAT GTA ACT TCG GCT TAC AAT CCA ACC TAT ACT A    1004

1005  AA ACC GCC TAA GCG AGC GTC TAG TTC AAG CGG GTT TTG CAC CGA TTG T    1052

1053  TT GCT GAC AAG CAA ACA CAG ACA AGC GAA CGA TGG ACA AAG GCG CGT C    1100

1101  GC AGT TTG AAA GCG TAG GCG TTA CGT ACG TGT TTT GCG TCA CGA AAT C    1148

1149  AA ACA AGA TAG CGC AAA CCT GGC GTT AGG CTA AAA ATC CCT AAA ACT A    1196

1197  AA ACC CCA AAT ATG TAG CGC GTC ATG CGT TGT TTT TAA TTA CAT TTT T    1244

1245  AA ACA ACT ATG TTG TTT TTA CAT GTT TTT ACC ATG CGC GCG CGT GAG G    1292
```

Continue sequence next page

```
1293  GA TTG GGG GTT GCA ACC CCC TAA ATA ACG AAG CTG TAG GAT TTC TCA T   1340

1341  TT TTG AGT GAA AAT GAA AGA ATG GTA ACT TCT TGT TAC TGA TAA GGG A   1388

1389  AC TTC TTG TTA CTG ATA AGG GAA CTT CTT GTT ACT GAT AAG GGA ACT T   1436

1437  CT TGT TAC TGA TAA GGG AAC TTC TAT GAA ATA GTT ACC ATA TTA ACC A   1484

1485  TT TTG AGC TAC AAT ACG ACC TAA GGC GGT GTA ACA TGG TTG CTA ATC C   1532

1533  TA GTG TTA ACA AAT TTG GAG CAA TTA GCT TTA AAA GCT AGT GGG TTG G   1580

1581  GA GTT TGT AGT GGG TAG CAC TCC GTT AGG AGG CAC ACC ATG AAA GCA T   1628

1629  TT TTG ATA GTA GTG ATT TTA GTG GTA ATC TTA ACA CAG CCA CTA TAT T   1676

1677  AA AAC CTT AGC GTT TTA ATA ACC CTT ATA AGT CCG CCA AGA CTT TTT A   1724

1725  AG GGT TTC ACT CCT ATC ATT ATA TCG TCT TTT GGA AAA TAA GCA TTA A   1772

1773  AA GGC TCT CAA ATG CCC ATG AAT ACG AAT TTT GAT CAA CTT AGA AAA     1819
   1                  M   P   M   N   T   N   F   D   Q   L   R   K      12

1820  CAA GAA TTG GAA TTA CGA AAA TTA TTA GAA GAA TTG GAT ACG CTC CCA    1867
  13   Q   E   L   E   L   R   K   L   L   E   E   L   D   T   L   P     28

1868  CAA ACC CCA CAA ATT GAG TTA CAA AAA CAA GAA ATA CAA GAC CGC GTC    1915
  29   Q   T   P   Q   I   E   L   Q   K   Q   E   I   Q   D   R   V     44

1916  AAC AAA ATA ACA GAC ACA ATC ATT AAA GAA TTA CTA TCA AAG CAT GAA    1963
  45   N   K   I   T   D   T   I   I   K   E   L   L   S   K   H   E     60

1964  ATC AAA AAA GAA GAA CTA AAA CCC ACT CTA AAA AAA GAA CCC ACA CCA    2011
  61   I   K   K   E   E   L   K   P   T   L   K   K   E   P   T   P     76

2012  CTC AAA GAT CCA CAA ACC ACC CCC ACA CCA TGC AAA AAT TTA GTG GTT    2059
  77   L   K   D   P   Q   T   T   P   T   P   C   K   N   L   V   V     92

2060  AGC ACC CCT AAA GAT AAC ACC TAT ACC ACC TAC CAC AAT AAC GCT AAT    2107
  93   S   T   P   K   D   N   T   Y   T   T   Y   H   N   N   A   N    108

2108  AAG GTT AAT CTA GGG AAA TTG AGC GAA AGG GAA GCC AAT CTT TTA TTC    2155
 109   K   V   N   L   G   K   L   S   E   R   E   A   N   L   L   F    124

2156  GCT ATT TTT CAA AAG CTT AAA GAT CAA GGG AAC ACT CTC ATT CGT TTT    2203
 125   A   I   F   Q   K   L   K   D   Q   G   N   T   L   I   R   F    140

2204  GAA CCG CAA GAT TTA AGG AGA ATG TTG GGC ATT AAA ATA TCC TAT GAT    2251
 141   E   P   Q   D   L   R   R   M   L   G   I   K   I   S   Y   D    156

2252  AAT CTT ACA AGA ACC GCT AGA AGC ATG TGG AAT AAA ATA AAA ACC GCT    2299
 157   N   L   T   R   T   A   R   S   M   W   N   K   I   K   T   A    172

2300  GAT TTT TGG GAA GTT AGA GAC ATT ATA GTG AAT GGT AGA GAA TGC GTT    2347
 173   D   F   W   E   V   R   D   I   I   V   N   G   R   E   C   V    188

2348  TCT GAA AAA AAT TAT ATG CTT TTT CAA GTT TGT GAA ATC GTA AGC GAT    2395
 189   S   E   K   N   Y   M   L   F   Q   V   C   E   I   V   S   D    204
```

Continue sequence next page

```
2396  AAA GAA ACA AGA GAA TTT TTG TAT ATG GAT ATT CAG CTC AAC ACA GGT  2443
 205   K   E   T   R   E   F   L   Y   M   D   I   Q   L   N   T   G   220

2444  TAT AAC TAT CTG CTC AAC AAT CTA GGC ATG GGC GGT CAA TAC ACT TCC  2491
 221   Y   N   Y   L   L   N   N   L   G   M   G   G   Q   Y   T   S   236

2492  TTT AAA CTT TTA GAA TTC CAA AGG GTG AGG GGC AAA TAC GCT AAG ACG  2539
 237   F   K   L   L   E   F   Q   R   V   R   G   K   Y   A   K   T   252

2540  CTT TAT CGC TTG CTC AAG CAA TAC AAA AGC ACA GGG ATT TTG AGC GTG  2587
 253   L   Y   R   L   L   K   Q   Y   K   S   T   G   I   L   S   V   268

2588  GAA TGG TCT CAA TTC AGG GAA CTT TTA GAC ATT CCA AAA GAC TAT GAA  2635
 269   E   W   S   Q   F   R   E   L   L   D   I   P   K   D   Y   E   284

2636  ATG AGA AAC ATC GAT CAA AAA GTT TTA ACC CCA AGC CTC AAA GAA CTC  2683
 285   M   R   N   I   D   Q   K   V   L   T   P   S   L   K   E   L   300

2684  AGA AAA ATC TAC CCC TTT GAA CAC CTG AGC TAC AAA AAG GAA CGC AGA  2731
 301   R   K   I   Y   P   F   E   H   L   S   Y   K   K   E   R   R   316

2732  AAC AAC CAC GAC AGG CGC AAA GTA ACG CAC ATT GAT TTC TAC TTT GAG  2779
 317   N   N   H   D   R   R   K   V   T   H   I   D   F   Y   F   E   332

2780  CAA TTG CCA CAG GGC GAA ACC AAG AAA CAA AAG CAA AAA CAC AAG CAA  2827
 333   Q   L   P   Q   G   E   T   K   K   Q   K   Q   K   H   K   Q   348

2828  CGC GCT AAA AGG GAT ATC AAG CTT GTA GCA TGG GAC ATC AAT AAC CAA  2875
 349   R   A   K   R   D   I   K   L   V   A   W   D   I   N   N   Q   364

2876  ATC GCT AAA AGA AAC GCT AAA GCC ACT ATG GAA GCT AGG TTT CTT GAA  2923
 365   I   A   K   R   N   A   K   A   T   M   E   A   R   F   L   E   380

2924  TTG AAA ACC TTG ATC GGC TAT CAG TTC AAG CAC AAC AAT GGG ACT ATT  2971
 381   L   K   T   L   I   G   Y   Q   F   K   H   N   N   G   T   I   396

2972  TTA CAG ATT GAC GAC GCC ACT TTT GAA AAG AAT CAA ATG TTT TTG CAT  3019
 397   L   Q   I   D   D   A   T   F   E   K   N   Q   M   F   L   H   412

3020  GTT TTG ACT AAC AAA AAT TCT CAA AAA TTC TCA AAA ATT GCC TGT GTC  3067
 413   V   L   T   N   K   N   S   Q   K   F   S   K   I   A   C   V   428

3068  CAA CAA AAC ATT CGC TTT AGA ACT TCT GTT TGT TAA TGG ATA CAG CCT  3115
 429   Q   Q   N   I   R   F   R   T   S   V   C   *                   440

3116  GAA AAA AGA CAG CTT GCT AGA AGA AAT TGA TCC CCC TAA AAT CCA CCC  3163

3164  TAT CAC CAA CGA ACC CAT CAA GGA ATT TGA TGA ATA CAT CGG CAA AAC  3211

3212  GAT CCA CAT CAC CAA CAA CAA TGT GGA TCA ATG CCC TGA TGG GAT AAC  3259

3260  AAG CTA TTT GAA AAT CAC TAG GAT TGT GAA ACT GAA TGA CAA TCG GAT  3307

3308  CTG TGT TTC AGT CCA AGA TGT GGA TAA GCC TGA CAA ACT CCT AAA ACC  3355

3356  TTT CAT CGC TAA AGA TGA AAA ACA CTT GAA AAA TTG GTT CAA GAA ACA  3403

3404  CTA CCA CTA GGC TTG TT TCA TTT TTC ATA CAA AGG GTA GGA AAT ATC  3451
                          *   K   E   Y   L   P   Y   S   I   D
```

Continue sequence next page

```
3452  CAT GCA GAC TTT AAA TAA CAA AGC CTT TCG AGT TAG CAT TAA ACG CCC   3499
        M   C   V   K   F   L   L   A   K   R   T   L   M   L   R   G

3500  ATA CCC ACG ACT ATA ATT TCT CTC GTT TTC TAT TTT TTG ATT GTG TTC   3547
        Y   G   R   S   Y   N   R   E   N   E   I   K   Q   N   H   E

3548  TCT AAT CTT TCT ATC CAG CTG TAT TGT TTT TGT ATG AAG TTC TCT TGT   3595
        R   I   K   R   D   L   Q   I   T   K   T   H   L   E   R   T

3596  TTC TTG TTC CAT TCG CTT TGA AAG TCG CTC ATC TGT TCT TTG AAA GAA   3643
        E   Q   E   M   R   K   S   L   R   E   D   T   R   Q   F   F

3644  GTC TCT AAT TCG CTC TTT TGC GTT TCT GTA GAT GTC TTT AAT TGC GTT   3691
        D   R   I   R   E   K   A   N   R   Y   I   D   K   I   A   N

3692  AAT TGC GTC TCG TGT TCT TGC ATA AAG GTC ATG TAA GAT TGT TCT AGC   3739
        I   A   D   R   T   R   A   Y   L   D   H   L   I   T   R   A

3740  TCT TGT TTG AGC GTT GTT TCT AAA TTC TCT AAT TGT GTT TTC AAT TCT   3787
        R   T   Q   A   N   N   R   F   E   R   I   T   N   E   I   R

3788  TTT GTA TCG AGC TTG TGA AGT TTC TTG CTC TTT GAT TTG AGC GCT CTG   3835
        K   Y   R   A   Q   S   T   E   Q   E   K   I   Q   A   S   Q

3836  TAA TTC TTG TTG TAA AAT CTG TTC GCA CTT GCT CTG TAA ATC GTT TTT   3883
        L   E   Q   Q   L   I   Q   E   C   K   S   Q   L   D   N   K

3884  GAG CGA GAT TTC TTT CTG TTC TAA GGT TTG TAT CAC TTG ATT CAA TTC   3931
        L   S   I   E   K   Q   E   L   T   Q   I   V   Q   N   L   E

3932  GGC TCT CTC TGC TTT CTC TAA AGC TTG AAA TTT GGC TAA TAA TGT GTT   3979
        A   R   E   A   K   E   L   A   Q   F   K   A   L   L   T   N

3980  GTG CGT TTG GGA TAG CGT TTC GTA GCT CTC TTT CAA GTT CAA GGT TTC   4027
        H   T   Q   S   L   T   E   Y   S   E   K   L   N   L   T   E

4028  TGT CTC TAA TTT CAA GTA ATC GTT GCT TAG TTT TTG AAA TTC TGC TGT   4075
        T   E   L   K   L   Y   D   N   S   L   K   Q   F   E   A   T

4076  AAC TTT CTC TAG CTT GGT TAA TAA CTC GTT TCT GTC GTT CTC TAA ACT   4123
        V   K   E   L   K   T   L   L   E   N   R   D   N   E   L   S

4124  CGC TAT CTT GCT GTT CTT TGC TTG TAG TAA GCT CTC TTT CTC GTT CTG   4171
        A   I   K   S   N   K   A   Q   L   L   S   E   K   E   N   Q

4172  TAA GTT CCC TAC TTC GCT GAT TAA AGC ACT CTG CAA CTT GTT TAA GCA   4219
        L   N   G   V   E   S   I   L   A   S   Q   L   K   N   L   C

4220  ATC GTT CTG CAT TTT TTC TCC TTA TTT CTT GGT CGG CAT TTG GTC GTA   4267
        D   N   Q   M

4268  AAT TCT CTG TAG ATT GGC GTT GTT TAG TTG GTC TAT TTC TCG TTT TAA   4315

4316  CTG AAA GAC TTC TTT CTG TGA AAG CTC TAA TTC CTT GTT TTT GTA GGG   4363

4364  CAG AAT GAT GAG AAA GTA TGC CAC TAC TAA TAT TGC CCC CAT GAT GAC   4411

4412  TAA TCC CAT GAT GAA AGT TAG TGG CAT TAG TAA GGA TTT CTT GGC TAA   4459
```

Continue sequence next page

```
4460  GTG TTC TTG TGG ATT TGT TTC TCT AGC TTT TGT AGT TTC GTT TGC ATG     4507

4508  CTT TTG TTG TGT TCT TTG TTG GCT GTC TCT AAC TCT AGG TAG CTG TTC     4555

4556  TTT TAA GCT CTC TTG GGC TTT TGA AGT AGA TTG ATG TAA TTT TTC ATA     4603

4604  GTA GCG TTG TTT GTA TTC AAT GTG TCT GTG TAG TTG TTG TTC CAA GTC     4651

4652  TCT AAG CTC TTG TAT GAG AGA TTG TGT AGC TCT TGG TGT TGT GTT AGG     4699

4700  AGC TGT TGA TAT TTC TCT TGT ATC TCT TTC TCG GCT TGT ATC TTT GTT     4747

4748  CTG GCT ATA TTC TCT AAG CTC TTG TTC AAT AAG TCT TTG TTC AGT AAA     4795

4796  TCG TTC ATC ATA AAT TAT TCC TTT CAA TCG TTT GGC TTT TTG TTG GTT     4843

4844  AGG GAG CTT GAC ACT GAT ATA ATC CTT GCC TTG TCT AGT AAT CTC ATA     4891

4892  ATT ACT CGC TTG TAA AAG CTC TAT TAT TTC TGC CCT TGA ATT TAA TAA     4939

4940  GCC TTG TAT TAC GGC GTT GTA AAG GTA ATT ATC TAG CTC TTT GTA GTT     4987

4988  CTT AAA GAT AGC CTT ACC TTG TGG GGT CTT TGT GTT GTG GTG TTG AAT     5035

5036  GGT GTT TTG TTT TTC TAA GTC TTT AGG GTT AGT AAA GGC GTG TTT TAA     5083

5084  GTT GGT ATA ATC CTT CCA AAT GTC AAT GCG TGT GAT GTC TGC ACT ATG     5131

5132  ATA GTA AGG GGT AAA GGC TTT TTG AGT TAT CAA ATC AAT TCT AGG TAT     5179

5180  AAC AAA ATT CAA CTC CAA ACG CCC CTT GTC AGT GTG TTT CTA CCC ATA     5227

5228  AGA TGT TGT AAC GCC CTT GCA TTT TTT CAG TCA ATA AAG CGG TTT CAA     5275

5276  AAC TTT CTA TCA ATT CGT GTT TGA GAT TTT CAT CAA TGT CAG GCT CTT     5323

5324  CAA AGG ATA AAC ACC CTA CGC ATG CCT TGT GTT TTT GAG TGA GAG AGA     5371

5372  GTA AAA GGC TTT TAG TTA AAT TAG CAT CGC CTT TTA AAA CTC TGG CTG     5419

5420  TGC TTG TTT CAC TCT TTC ATT AAG CAA GTA ATT AAC GCT CCC ATC ACC     5467

5468  ATC GCC ACC CCC TTG ATT AAT GCC CCA AAA TTT AAC CAA CAT ACT TGG     5515

5516  ATC TTA TTT GTT CTA ATT GAT GAC TGA TTT CTA TGA GTT GTT CTA AAC     5563

5564  CTA ATC TAT CCC ATA CCC CTT GAT TAG TGT TGA GAC ATT TAG CGA TTT     5611

5612  GGT TTA AAT TGT TTC CCC ACT TAG CCA ACT CAA TAA CTA ACT CTT TAT     5659

5660  TA                                                                5660
```

Fig. 11. DNA sequence of pHPM179. The sequence starts at the first nucleotide of the unique *Pvu*I site. ORF1, ORF2 and ORF3 amino acids are indicated below the sequence. Sequences that are homologous to ribosome binding and promoter consensus sequences are boxed. Direct repeats are indicated by arrows above the sequence.

the 0.68 kb fragment, ending at n.t. 3103 within the 1.1 kb fragment. The protein encoded by ORF1 would consist of 440 amino acids, with a molecular weight of 52 kDa. ORF2 begins at n.t. 336 and ends at n.t. 617. If translated the protein would consist of 94 amino acids with a molecular weight of 10.8 kDa. ORF3 begins at n.t. 4231 on the bottom strand and ends at n.t. 3421. If translated the protein would consist of 270 amino acids, with a molecular weight of 31.9 kDa.

Sequences that are homologous to the consensus -10 and -35 promoter (TATTAT and TTGACA, respectively) and ribosome binding site (RBS) (AGGAGA) are boxed in Fig. 11. For ORF1, the putative promoter begins at bp 1690 and 1658, with the RBS at bp 1771. The ORF2 putative promoter begins at bp 295 and 251, with the RBS at bp 326. The ORF3 putative promoter would be on the bottom strand beginning at bp 4395 and 4438 with the RBS at bp 4241.

In the non-coding DNA, four short 22 bp directly repeated iterons of sequence TTCTTGTTACTGATAAGGGAAC occur at bp 1369, 1391, 1413 and 1435. This is shown in Fig. 11 with arrows over the nucleotides. Fig. 12 is a schematic map of pHPM179 and shows the relative locations of the ORFs and iterons.

Comparison of the pHPM179 DNA sequence with sequences in the GenBank databank revealed several regions of near identity with pHPM180 (46), as well as regions highly homologous to pHPK255 and to the chromosomal *cagA* region. These are indicated in Fig. 12.

The relationship between the pHPM179 and pHPM180 sequences is shown by the diagonal "dot-matrix plot" generated by DNA Inspector *IIe* (Fig. 13). This method allowed visual comparison of pHPM180 and pHPM179, where the X axis was the pHPM179 sequence and the Y axis was the pHPM180 sequence. The matrix was scored

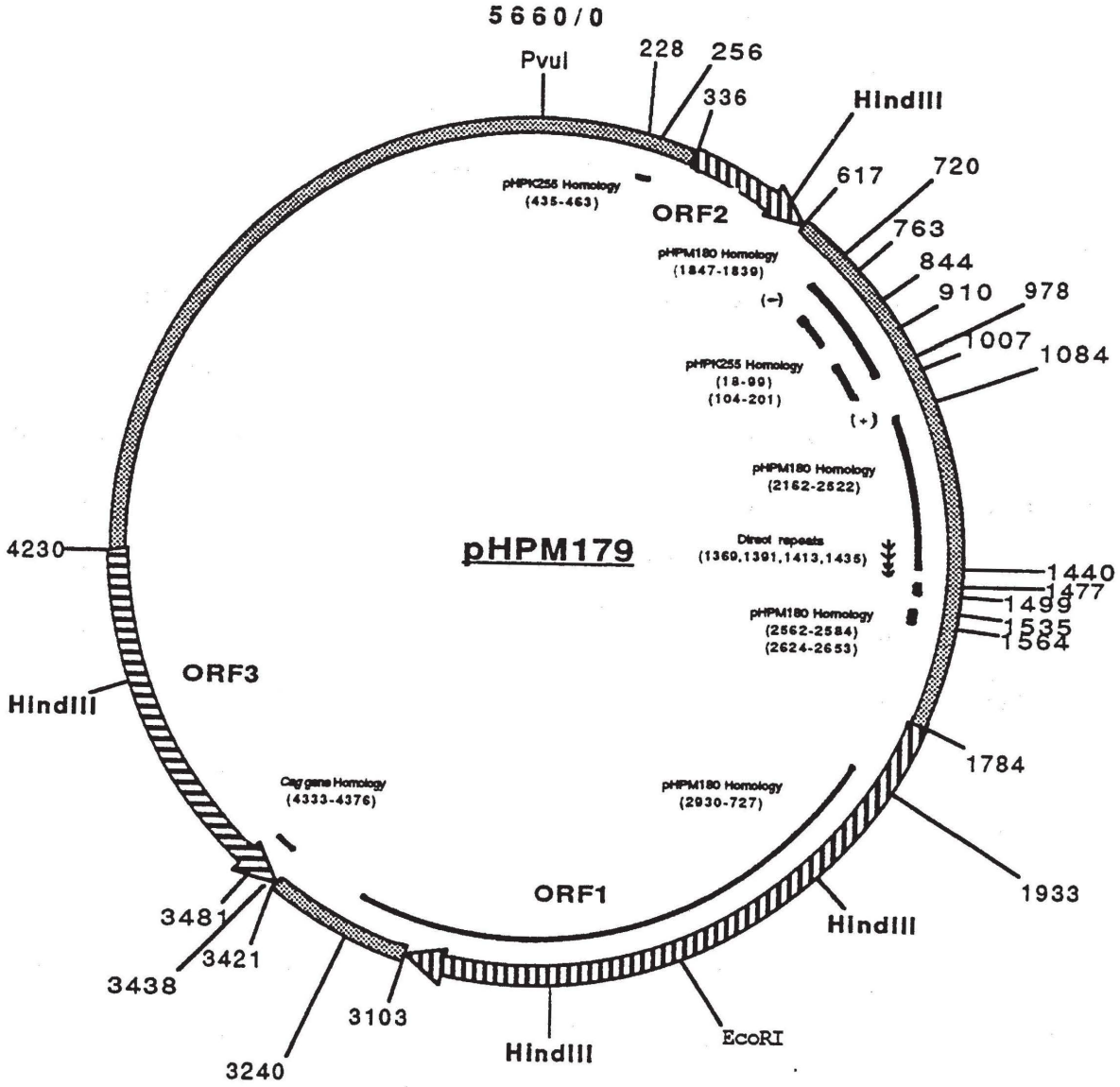

Fig. 12. Structural map of pHPM179.
The relative position of restriction sites, ORFs, direct repeats and locations of nucleotide sequence homologies between pHPM180, pHPK255 and pHPM179 are shown. The ORFs are indicated by the striped arrows, the direct repeats are indicated by short arrows. Outside numbers refer to pHPM179 nucleotide positions starting at the unique *Pvu*I site as nucleotide 1. The regions of homology with pKPK255, pHPM180 and *cagA* gene are marked by lines inside the circle. The plus and minus signs refer to the top and bottom (+/-) strands. Drawing scale is 150 nt/per cm.

by putting a "dot" at each x, y position that corresponded to an identity. In the regions where the two sequences were homologous, a clear line materialized along the major diagonal (23, 38). The diagonal lines in the upper left quadrant in Fig. 13 are in non-coding regions of both pHPM179 and pHPM180, and include the 232 bp direct repeats observed in pHPM180 (46), as well as the 22 bp iterons mentioned above. Fig. 14 is a direct comparison of the DNA sequence in these regions of the two plasmids, showing that the iterons are highly similar (12 of 22 n.t. identical), but not identical.

A comparison of pHPM179 with pHPK255, the small *H. pylori* plasmid sequenced by Kleanthous et al. (31), revealed several regions of homology, most of which also were homologous to pHPM180. Two of these regions were on the same strands (i.e. +/+), and one was on opposite strands (i.e. -/+). Their sequences are given in Fig. 15 and their relative locations are indicated in Fig. 12.

GenBank analysis of pHPM179 also revealed another homology with an *H. pylori* chromosomal gene now termed *cagA* for *c*ytotoxin *a*ssociated *g*ene (64). This short (44 bp) stretch of homology contained 97% identical bases and occurred at the end of *cagA*, just outside the ORF. The region in pHPM179 was located at the C-terminus of the putative ORF3, i.e. within the ORF, and is indicated in Fig. 12.

Analysis of ORF amino acid sequences

Comparison of the pHPM179 ORF1 amino acid sequence with the protein data bank (GenBank) revealed an extensive homology with ORF1 of pHPM180. Among other prokaryotic protein alignments with ORF1 were nine plasmid replication (Rep) proteins: RepA protein from plasmid pCCT1 of *Campylobacter coli* (Stonnet, unpublished); RepA protein from a cryptic plasmid of *Pediococcus halophilus* (Benachour, unpublished);

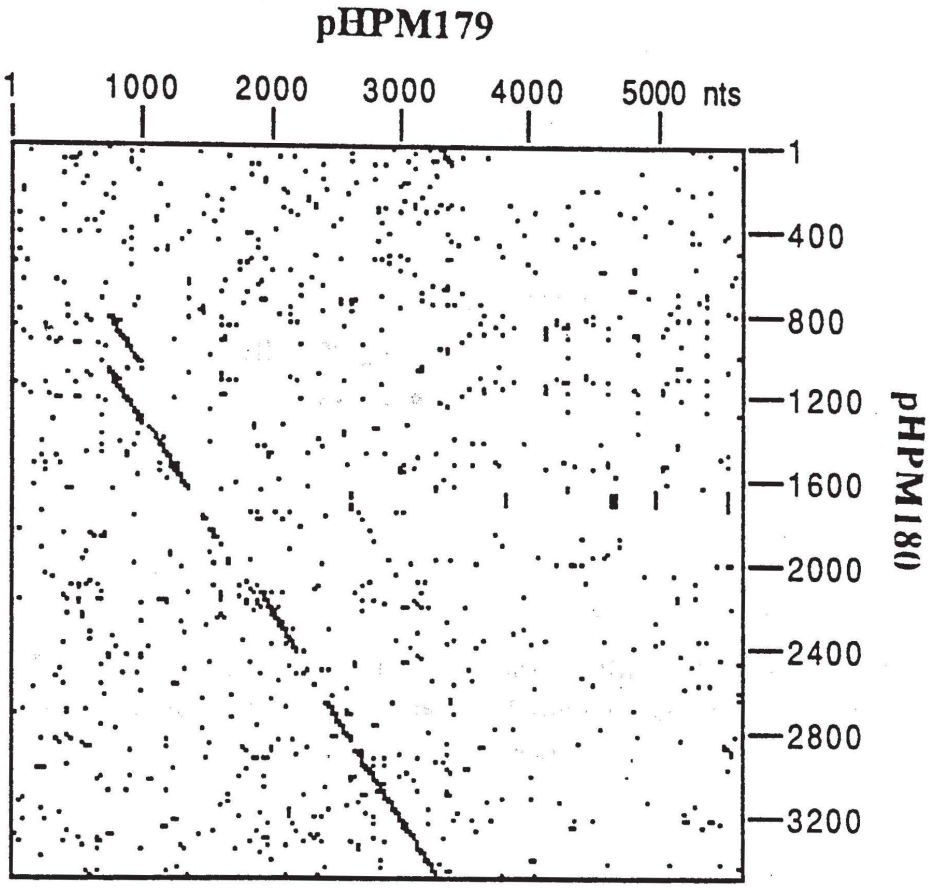

Fig. 13. Homology matrix analysis of the nucleotide sequences pHPM179 and pHPM180. Homology matrix analysis of the sequence was obtained using the following setting: span length, 10; matrix resolution, 200 x 200; maximum number of mismatches, 1. The first nucleotide for pHPM179 was nt#1 at the unique *Pvu*I, and the first nucleotide for pHPM180 was nt#791 at the *Mbo*I site.

```
pHPM179  1369 TTCTTGTTACTGATAAGGGAAC    1391
pHPM180  2450 TTCTTGCGAACACTATGAGTAG    2472
              ******   *     ** * * *
```

Fig. 14. Comparison of iterons from pHPM179 and pHPM180. Asterisks indicate identical bases. Numbers on right and left refer to start and end nucleotides, respectively.

A

```
pHPM179     910  TTCATTTCCCCATTAGTTGCTAACCATTTAGCCAATCCCACTTGTTTAGC  960
pHPK255     104  TTCCACTCACACCTAGAAGCTAATCAGTTAGCAAATCTCACGAGGGTAGC  153
pHPM180     2039 TTCATTTCCCTTTTAGTTGCTAGTCAGTTAGCCAATCCCACTGGTTTAGC 2088
                 ***---**-*..-***--*****--**-*****-****-***-.*--****

pHPM179     961  ATCTAAGAGCGCATGTAACTTCGGCTTACAATCCAACCTATACTAAAA    1007
pHPK255     154  CAATAAAAGCGCATGTAACCTCGGCTATCAATCTCACCAAAAAGCAAA    201
pHPM180     2089 ATCTAAAAGCGCATCTAACTAGGGCTTACAATCCTACC             2126
                 ---***-*******-****---*****--*****-.***
```

B

```
pHPM179     844    AACAAACGAAAAACAAAAAACAGACTAATGTACTACCTACA      805(-)
pHPK255     18     ACCAACGCAAAACAAAAAACCGACTATTGATACTACCTACA      58 (+)
pHPM180     1703   ACCAACGAAAAGCAAAAAACCGACATATGATACTACCTACA      1663(-)
                   *-***--.***---******-*---.-----************

pHPM179     804    AGCTAAAAAACGCTCAAAACGAAAATATGAGCGAAAAAATC      763(-)
pHPK255     59     AGCTAAAAATCGCTTAAAACGAAAATATGAGCGAAAAAAGC      99 (+)
pHPM180     1662   AGCTAAAAAACGCTTAAAACGAAAATATGAGCGAAAAAATC      1622(-)
                   *********-****-*************************-*
```

C

```
pHPM179     228          ATTTTGTATTACAAAATATATGCTTGCCC        256
pHPK255     435          ATCTTGTATTACAAAATATATGCTTGCCC        463
                         **.**************************
```

Fig. 15. Comparison of pHPM179, pHPM180 and pHPK255 using the NCBI BLAST algorithm. Asterisks indicate identical bases among all three *H. pylori* plasmids, a dash indicates identical bases between two of the three plasmids and a period indicates different bases among all three plasmids. A, comparison of all three plasmids on the same strand (+); B, comparison of pHPK255 plus (+) strand to minus (-) strand of pHPM179 and pHPM180; C, comparison of homologous regions on the plus (+) strands of pHPM179 and pHPK255.

RepA protein from a *Pseudomonas aeruginosa* plasmid pR01600 (29); RepA protein from *Pseudomonas syringae* plasmid pPS10 (48); RepA protein from *Lactococcus lactis* plasmid pUCL22 (21); RepA protein from *Neisseria gonorrhoeae* plasmid pFA3 (24); RepB protein from *Lactococcus lactis* plasmid pWV04 (Seegers, unpublished); RepB protein from *Klebsiella pneumoniae* plasmid pGSH500 (13); and RepB protein from *Campylobacter coli* plasmid pCCT2 (Stonnet, unpublished). The regions of greater than 46% identity or neutral replacement among these Rep proteins and pHPM179 ORF1 are shown in Fig. 16. These plasmids also contained a series of iterons, regions that are the possible binding site of REP protein. The plasmid similarities and the GenBank accession numbers are listed in Table 2.

The deduced amino acid sequence of ORF1 showed a comparatively high content of lysine (51 or 11%), leucine (50, 11%), glutamic acid (32, 7%) and asparagine (22, 5%). On the basis of amino acid composition, the net charge ORF1 polypeptide was predicted to be +25 at pH 7. The hydrophilicity profile (32) of ORF1 peptide suggested that the deduced protein is predominantly hydrophilic.

Comparison of ORF2 and ORF3 to both DNA and protein databases did not reveal any significant homologies with known prokaryotic proteins.

```
pHPM179   MPMNTNFDQLRKQELELRKLLEELDTLPQTPQIELQKQEIQDRVNKITDTIIKELLSKHEIKKEELKPTLKKEPTPLKDPQTTPTPCKNLVVSTPKDNTY  100
pHPM180   --------------DLQK-LK-LD-LPQT-K--L-K-EIK----K-T---IKELLSKHEIKKEELKPTLKEEP-P-KEPQTTPTPCKDLVVSTPKDNTY
CRYPTIC   ------------------------------------------------------------------------------------------------N--
pR01600   VPI-SSYNQLKE--L-LKR-------------------------------------------------------------------------------
pFA3      -------------------------------------------------------------------------------------------------
pPS10     -------------------------------------------------------------------------------------------------
pUCL22    -------------------------------------------------------------------------------------------------
pWV04     -------------------------------------------------------------------------------------------------
pGSH500   -------------------------------------------------------------------------------------------------
pCCT2     -------------------------------------------------------------------------------------------------
pCCT1     -------------------------------------------------------------------------------------------------

pHPM179   TTYHNNANKVNLGKLSEREANLLFAIFQKLKDQGNTLIRFEPQDLRRMLGIKISYDNLTRTARSMWNKIKTADFWEVRDIIVNGRECVSEKNYMLFQVCE  200
pHPM180   TTYHNNANKVNLGKLSEREANLLFAIFQRLKDQGNTLIRFEPQDLKRMV---LT---L-Q--KNL-D-I--ANFW-IRE-V-NG-------SYMLFK--D
CRYPTIC   --Y----N-I-L-R-T--E-NL-FSV--RMRDKGD--VRF----LK-L---K-T-NN-------------------------------------------
pR01600   -------------------------------------------------------------------------------------------------
pFA3      -------------------------------------------------------------------------------------------------
pPS10     -------------------------------------------------------------------------------------------------
pUCL22    -------------------------------------------------------------------------------------------------
pWV04     -------------------------------------------------------------------------------------------------
pGSH500   -------------------------------------------------------------------------------------------------
pCCT2     --YHND-NKI-L---TEQE-NLLF-IF-A-RIKEKG----------------------------------------------------------------
pCCT1     --YNND-N-L-M-EL---Q-DM--AI----KDK-------------------------------------------------------------------

pHPM179   IVSDKETREFLYMDIQLNTGYNYLLNNLGMGGQYTSFKLLEFQRVRGKYAKTLYRLLKQYKSTGILSVEWSQFRELLDIPKDYEMRNIDQKVLTPSLKEL  300
pHMP180   I---K-TQ---YLEVQLN--YHYLLNNLGMGGQYTSFKLLEFQRVRGKYAKTLYRLLKQYKSTGILSVEW-QFRELLDIPKDYKM-NIDK-VL--ALKEL
CRYPTIC   ----K-S-E--YVDIQI------LLNNL-----W--Y-L-EFR-LK--YAKTMFRLLKQFRTTGY-----S-F-ELLDIPQ-Y---NVE-RVI-P--EEL
pR01600   --------------------------------QFT-Y-L-D--KM---HA--LY-LL-QW-S-G---IE--Q-RK---L---Y----------------
pFA3      --------------------------------GQFT-YQL---------HS--IY-LI-QYRS-G---I---K-KE-L-V---Y---NLNQRVL-PAI-EI
pPS10     --------------------------------EFTSYQL-Q---L---YA--LY-LM-QF----------AQ-REM-DL---YDVKDM-KRVL-PALEEV
pUCL22    ----------------------------I-L--EFT-YKI-E-QKL--KYS--LYR-L-------------------------------------------
pWV04     ----------------------------I-L--EFT-YKI-E-QKL--KYS--LYR-L-------------------------------------------
pGSH500   -------------------------------------------------------------------------------------------------
pCCT2     --ES--F--I-LKIN--F-YLVN-L--------FTAFEL-EF--L-GKYAKTLYRLLKQFRTTG----EW-EF--IMDIPQDY----IDKR------K--
pCCT1     --------------------------------FTAFEL-EF--L-GKY-KTLYRLLKQFRTTG----EW-EF--IMKIPENY---DIDQRIL-PAIKEL

pHPM179   RKIYPFEHLSYKKERRNNHDRRKVTHIDFYFEQLPQGETKKQKQKHKQRAKRDIKLVAWDINNQIAKRNAKATMEARFLELKTLIGYQFKHNNGTILQID  400
pHPM180   RKIYPFEHLSYKKERRNSHDKRKVTHIDFYFEQLPQ-ETK-Q-QK-KQRAKRDIKLIAWDIHNQIAKRNAKATMEARFLELKTLIGYQFRHNNGTILQIN
CRYPTIC   --LF-----------------------------------------------------------------------------------------------
pR01600   ------EH---Q-E-------RKVTHL-LF-F-----------------------------------------------------------------
pFA3      -E--------------------------------------------------------------------------------------------------
pPS10     -K--------------------------------------------------------------------------------------------------
pUCL22    -------DHLSY---KK--Y-KR-V----------------------------------------------------------------------------
pWV04     ------DHLSY---KK--Y-KR-V----------------------------------------------------------------------------
pGSH500   -----------KKQ-Q----R-R----E---EQ----ES-K--RKYERR-K---------------------------------------------------
pCCT2     -K--PFKNLAYEKEK-------------------------------------------------------------------------------------
pCCT1     -K--PFKNLAYEKEK-------------------------------------------------------------------------------------

pHPM179   DATFEKNQMFLHVLTNKNSQKFSKIACVQQNIRFRTSVC  440
pHPM180   NATFEKNQMFLHV-TNKNSQKF-----------------
CRYPTIC   ------------------------------------
pR01600   ------------------------------------
pFA3      ------------------------------------
pPS10     ------------------------------------
pUCL22    ------------------------------------
pWV04     ------------------------------------
pGSH500   ------------------------------------
pCCT2     ------------------------------------
pCCT1     ------------------------------------
```

Fig. 16. Comparison of plasmid Rep proteins. Amino acid sequence segments of plasmid replication (Rep) proteins showing greater than 46% identities or neutral substitutions with pHPM179 ORF1 were determined using the NCBI BLAST algorithm. Identical or neutral substitutions are shown, with (-) indicating non identical. Numbers on the right refer to pHPM179 ORF1 amino acids. Sources of the plasmids are listed in Table 2.

Table 2. Homologous Plasmid Encoded Replication Proteins.

Plasmid	Host	Rep protein	Iterons	GenBank
pHPM179	*Helicobacter pylori*	RepA	4 x 22	
pHPM180	*Helicobacter pylori*	RepA	4 x 22	U12689
pCCT1	*Campylobacter coli*	RepA	4 x 22	X82079
CRYPTIC	*Pediococcus halophilus*	RepA	4 x 22	X75607
pR01600	*Pseudomonas aeruginosa*	RepA	4 x 17	L22691
pFA3	*Neisseria gonorrhoeae*	RepA	4 x 22	P17492
pPS10	*Pseudomonas syringae*	RepA	4 x 11	S20615
pUCL22	*Lactococcus lactis*	RepA	4 x 22	X56550
pWV04	*Lactococcus lactis*	RepB	2 x 22	Z25476
pGSH500	*Klebsiella pneumoniae*	RepB	10 x 22	S33663
pCCT2	*Campylobacter coli*	RepB	5 x 22	X82080

CHAPTER IV

DISCUSSION

Many strains of *H. pylori* contain plasmids of various sizes, but no functions have been assigned to them. To date, two plasmids have been studied in detail. Kleanthous et al. (31) characterized a 1.5 kb plasmid, pHPK255, which has a single ORF, encoding a 25 kDa replication protein (REP) homologous to replication proteins isolated from other plasmids that replicate by a rolling circle type mechanism (RCR). Using Southern hybridization and PCR, they showed that pHPK255 contains other RCR signals and conclude that this is its mechanism of replication. They did describe, however, other *H. pylori* plasmids in their collection that do not show homology with any of the RCR signals and concluded that other replication mechanisms probably exist in some *H. pylori* plasmids.

Minnis (46) characterized a somewhat larger (3.5 kb) plasmid, pHPM180, which contains two ORFs, neither of which shows significant homology to the 25 kDa Rep protein of pHPK255. The larger ORF encodes a protein of 54.5 kDa which shows significant amino acid homology with RepA and RepB replication proteins identified in plasmids that replicate via the theta type mechanism.

The present study was undertaken to analyze a larger *H. pylori* plasmid, pHPM179 (approximately 6 kb), to determine whether additional plasmid encoded proteins might be detected that would provide information as to the function of *H. pylori* plasmids. Since the DNA sequence of one small *Hin*dIII fragment (684 bp) was already known (8), the goal of this project was to clone and determine the DNA sequence of the remaining fragments of pHPM179.

It was important for this study to determine the order and orientation of the cloned *Hin*dIII fragments. A multiple enzyme digestion was done so as to construct a physical map of pHPM179. Here the cleavage products of one endonuclease were characterized with respect to size and then digested with a second endonuclease. Analysis of the resultant double-digestion products established the relationship between the cleavage sites of the two enzymes. This result was confirmed by ordering DNA fragments from partial endonuclease digestion reactions (data not shown). From these digestions the order of the fragments was found to be 2.2, 1.6, 0.68 and 1.1 kb.

Initial attempts to clone these *Hin*dIII fragments with the M13mp18 vector, using DH5αF' as the transformation recipient were unsuccessful. A second set of cloning attempts was made using the vector pTZ19R, and introducing it into the DH5αF' host. These attempts were also unsuccessful. One possible explanation for these failed attempts at cloning *H. pylori* DNA into *E. coli* was the extensive modification that is known to occur in *H. pylori* DNA. Such methylation patterns would cause the DNA to be recognized by *E. coli* as foreign and to be rapidly destroyed by host enzymes (39). Therefore, a strain of DH5 (DH5αMCR) that had been mutated in the methylation genes so as to accept highly methylated plant DNA (4) was used as a transformation recipient for ligations of pHPM179 DNA fragments into the pTZ19R vector

With these experiments, two recombinants from the first twenty examined were found to contain pHPM179 DNA. One recombinant contained the 1.6 kb fragment alone and the other recombinant contained the 1.6 kb and 2.2 kb fragments together. The plasmid containing both the 1.6 and the 2.2 kb fragments was renamed pHPJQ38 and was used for DNA sequence determination.

Further attempts to clone the remaining 1.1 kb fragment were unsuccessful, even by purifying the fragment prior to ligation. The reasons for this failure are unknown, although the possibility remained that some modification of this fragment, synthesized in *H. pylori*, still caused restriction in *E. coli*. Therefore, an additional cloning attempt was made using PCR amplification of this fragment (i.e., synthesized in a test tube) prior to ligation and transformation. The rational for this approach was that the *in vitro* synthesized copies lack any nucleotide modifications. With the known sequence of the 684 bp fragment and, after sequence determination of pHPJQ38 was started, the known sequence of the neighboring 2.2 kb fragment, two primer nucleotides were designed that flanked the 1.1 kb fragment and could be used for PCR amplification. With this technique, the 1.1 kb *Hin*dIII fragment was successfully cloned and the DNA sequence was determined.

To confirm that all of the cloned fragments represented DNA from pHPM179, Southern hybridization analysis was done. As expected, all cloned pieces hybridized with native pHPM179 DNA.

Double-stranded DNA sequence determination of the three insert fragments initially utilized the universal forward sequencing primer and reverse sequencing primer. Subsequently primers designed from each of the previously determined sequence were used. At regions where, on occasion, there was compression or other anomalous spacing, single-strand DNA preparations were used to overcome these problems.

Combining the sequences of the 2.2, the 1.6, and the 1.1 kb fragments with the previously determined 684 bp fragment sequence produced the total DNA sequence of pHPM179. The plasmid was 5660 bp in length, as schematically shown in Fig. 12, using the unique *Pvu*I site as the beginning coordinate on the circular map.

Computer analysis of pHPM179 revealed a G+C content of 36.2%, similar to the 37% G+C of pHPM180 (46), to the 38 % G+C of pHPK255 (31), and to the 35-37% G+C reported for the *H. pylori* chromosome (68). The analysis also revealed three open reading frames greater than 90 amino acids in length: ORF1, starting with the ATG codon at n.t. 1784, would encode a polypeptide of 52 kDa; ORF2, starting with an ATG codon at n.t. 336, would encode a polypeptide of 10.8 kDa; and ORF3, starting with an ATG codon on the bottom strand at n.t. 4231, would encode a polypeptide of 31.9 kDa. All three ORFs contained upstream sequences homologous to the *E. coli* consensus RBS (AGGAG) and -10 (TATTAT) and -35 (TTGACA) promoter sequences.

Located 323 bp. upstream from the beginning of ORF1 were four 22 bp directly repeated sequences (at n.t. 1369, 1391, 1413 and 1435). There are 22 bp repeats 332 bp upstream of ORF1 in pHPM180 as well, although the sequences in pHPM179 and pHPM180 are not identical. Such repeats are called iterons and are presumed to be the site of Rep protein binding to initiate theta type plasmid replication (see below).

The BLASTN algorithm (1) from NCBI, which compares n.t. sequences to databank n.t. sequences, reported matches of pHPM179 DNA sequence with several significant prokaryotic sequences: five short stretches of homology with the *H. pylori* plasmid, pHPM180; three short stretches of homology with the *H. pylori* plasmid, pHPK255; and one short stretch of homology with the *H. pylori* chromosomal gene *cagA*. The matches with pHPM180 DNA included the majority of ORF1 DNA (discussed below), as well as four regions in non-coding DNA (see Fig 12). These non-coding regions of matching in both pHPM179 and pHPM180 included the iterons mentioned above.

The matches with pHPK255 DNA were partly on the same strand and partly on opposite strands, and included some of the sequences that matched with pHPM180 as well. The conservation of these sequences among all three *H. pylori* plasmids may mean that they are involved some way in plasmid maintenance or partitioning in *H. pylori*. Alternatively, since pHPK255 presumably replicates by a different mechanism, the maintenance and partitioning systems may be different as well. These regions may simply indicate that *H. pylori* plasmids undergo major recombinations over many years, and that analysis of additional plasmids will show similar sorts of conserved sequences, all related to an ancestral progenitor plasmid.

The BLASTN algorithm also showed a short (44 bp) region of near identity with the *cagA* chromosomal DNA. The region of homology in *cagA* is actually just outside of the *cagA* ORF at the carboxy terminus of the encoded protein. In pHPM179, the 44 bp are located within ORF3 in the carboxy terminus of the encoded protein. Although only about 60% of *H. pylori* isolates produce the 128 kDa *cagA* protein, almost all *H. pylori*-infected patients with duodenal ulcers develop an antibody response to *cagA*. In addition, the presence of *cagA* protein was reported as strongly associated with cytotoxin (*vac*uolating toxin) production by *H. pylori* cells (64). The two genes, *cagA* and *vac*, however, are about 300 kb apart on the chromosome (7). Tummuru et al. (65) recently constructed *cagA* mutations and showed that the *cagA* gene is not necessary for cytotoxin production. To explain the previous findings of association between *cagA* and cytotoxin production, Tummuru et al. (65) hypothesize that the *cag* gene in *cagA*$^-$ strains may be due to either insertion into or deletion of the *cagA* gene. If true, the DNA in and around the *cagA* gene may contain hot spots for recombination. One of the common observations with *H. pylori* DNA, both chromosomal and plasmid, is that rearrangements seem to have

occurred, leading to variation in the restriction enzyme patterns among different isolates. It is therefore tempting to propose that the observed presence of *cagA* sequences in pHPM179 indicates that such rearrangements occur between plasmid DNA and chromosomal DNA.

Databank analysis of the ORF1 amino acid sequence uncovered a match of a long stretch of the protein sequence to known Rep proteins from a variety of other plasmids. These plasmids, from such diverse genera as *Neisseria*, *Pseudomonas*, *Pediococcus*, *Lactococcus*, *Klebsiella*, and *Campylobacter*, share the characteristic of theta-type replication, not the RCR mechanism proposed by Kleanthous et al. (31) for *H. pylori* plasmid pHPK255. This mode of replication is the one proposed for pHPM180 by Minnis (46), and is based also on the homology with other Rep proteins and the occurrence of iterons. Based on these observations, it would seem likely that pHPM179 also replicates by a theta type mechanism.

Databank analysis of ORF2 and ORF3 amino acid sequences did not provide significant prokaryotic matches with known proteins. Although both ORFs contain the appropriate transcription and translation signals, their role in this plasmid is unknown.

In summary, this project produced the DNA sequence of pHPM179, a 5.6 kb *H. pylori* plasmid, the third *H. pylori* plasmid to be analyzed. The results of sequence analysis revealed several interesting features : 1) There are three presumptive ORFs that could encode proteins of 52 kDa, 10.8 kDa, and 31.9 kDa; 2) ORF1 protein shared striking homology with other plasmid replication proteins, including *H. pylori* pHPM180 and plasmids from *Campylobacter coli*, *Pediococcus halophilus*, *Pseudomonas aeruginosa*, *Pseudomonas syringae*, *Lactococcus lactis*, *Neisseria*

gonorrhoeae and *Klebsiella pneumoniae* (all of these plasmids are thought to replicate via a theta mechanism, rather than by RCR); 3) There are four 22 bp iterons, similar to the number and type of iterons found in the above-mentioned theta-type replicons; 4) Non-coding DNA sequences are similar to those of the other two *H. pylori* plasmids and to the chromosomal *cagA* gene.

The conclusions from this work are: 1) pHPM179 is a medium sized plasmid that probably replicates via a theta type mechanism, not via the RCR mechanism that pHPK255 utilizes; and 2) the presence of non-coding DNA sequences in pHPM179 that are homologous to non-coding sequences in two other *H. pylori* plasmids (one of which replicates by RCR) and to a short sequence near the *cagA* chromosomal gene may mean that plasmids are involved in extraordinary recombination events, both between plasmids and between plasmids and the chromosome.

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. Birnboim, H., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1523.

4. Bishr, M.. 1991. Devlopment of a mutant strain of *E. coli* for molecular cloning of highly methylated DNA. PhD. Dissertation. Texas Woman's University, Denton, Texas.

5. Borady, J. T., P. Andrew, N. Mancuso, D. McCauley, E. Jankiewicz, N. Ferch, N. Shortis, and S. Brandl. 1994. *Helicobacter pylori* reinfection rate in patients with cured duodenal ulcer. Am. J. Gastroenterol. Vol 89,4:529-532.

6. Buck, G. E. 1990. *Campylobacter pylori* and gastroduodenal disease. Clin. Microbiol. Rev. 3:1-12.

7. Bukanov, N. and D. Berg. 1994. Ordered cosmid library and high-resolution physical genetic map of *Helicobacter pylori* strain NCTC11638. J. Mol. Biol. 11:509-523.

8. Chen, J. 1993. DNA sequence analysis of a *Helicobacter pylori* plasmid fragment. MS thesis. Texas Woman's University, Denton, Texas.

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

10. Covacci, A., S. Censini, M. Bugnoli, R. Petracca, D. Burroni, G.Macchia, A. Massone, E. Papini, Z. Xiang, N. Figura, and R. Rappuoli. 1993. Molecular characterization of the 128-KDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc. Natl. Acad. Sci. USA 90:5791-5795.

11. Cover, T. L., and M. J.Blaser. 1992. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. J. Biol. Chem .267:10570-10575.

12. Cover, T. L., S. A. Halter, and M. J. Blaser. 1992. Characterization of Hela cell vacuoles induced by *Helicobacter pylori* broth culture supernatant. Hum. Pathol. 23:1004-1010.

13. DaSilva-Tatly, F. M., and L. M. Steyn. 1993. Characterization of a replicon of the moderately promiscuous plasmid, pGSH5000, with features of both the mini-replicon of pCU1 and the *ori*-2 of F. Mol. Microbiol. 7:805-823

14. Doglioni G., A. C. Wotherspoon, A. Moschini, M. De Boni, and P. G. Isaacson. 1992. High incidence of primary gastric lymphoma in north eastern Italy. Lancet 339:834-835.

15. Drumm, B., G. I. Perez-Perez , M. J. Blaster, P. M. Sherman. 1990. Intrafamilial clustering of *Helicobacter pylori* infection. N. Engl. J. Med. 322:359-63.

16. Dunn, B., C. C. Sung, N. S. Taylor, and J. G. Fox. 1991. Purification and characterization of *Helicobater mustelae* urease. Infect. immun. 59:3343-3345.

17. Eaton, K. A., C. L. Brooks, D. R. Morgan, and S. Krakowka. 1991. Essential role of urease in pathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. Infect. Immun. 59:2470-2475.

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

19. Ferrero, R. L., and A. Lee. 1991. The importance of urease in acid protection for the gastric-colonising bacteria *Helicobacter pylori* and *Helicobacter felis* . Microb. Ecol. Health. Dis. 4:121-134.

20. Figura, N., P. Guglielmette, A. Rossolini, A. Barberi, G. Gusi, 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.

21. Frere, J., M. Novel, and G. Novel. 1993. Molecular analysis of the *Lactococcus lactis* subspecies lactis CNRZ270 bidirectional theta replicating lactose plasmid pUCL22. Mol. Microbiol. 10:1113-1124.

22. George, L. L., T. J. Borody, P. Andrews, M. Devine, J. D. Moore, M. Walton, and S. Brandle. 1990. Cure of duodenal ulcer after eradication of *Helicrobacter pylori*. Med. J. Aust. 1533:145-149.

23. Gibbs, A. J. ,Jr., and G. A. McIntyre. 1970. The diagram, a method for comparing sequences. Eur. J. Biochem. 16:1-11.

24. Gilbride, K. A., and J. L. Brunton. 1990. Identification and characterization of a new replication region in the *Neisseria gonorrhoeae* β-lactamase plasmid pFA3. J. Bacteriol. 172:2439-2446.

25. Hazell, S. L., D. J. Evans, and D. Y. Graham. 1991. *Helicobacter pylori* catalase. J. Gen. Microbiol. 137:57-61.

26. Hazell, S. L., A. Lee, L. Brady, and W. Hennessy. 1986. *Campylobacter pyloridis* and gastritis; association with intracellular spaces and adaptation to an environment of mucus as important factors in colonization of the gastric epithelium. J. Infect. Dis. 153:658-666.

27. Hirai, M., T. Azuma, S. Ito, T. Kato, Y. Kohli, and N. Fujiki. 1994. High prevalence of neutralizing activity to *Helicobacter pylori* cytotoxin in serum of gastric-carcinoma patients. Int. J. Cancer 56:56-60.

28. Hupertz, V., and S. Czinn. 1988. Demonstration of a cytotoxin from *Campylobacter pylori*. Eur. J. Clin. Micrbiol. Infect. Dis. 7:576-578.

29. Jansons, I., G. Touchie, R. Sharp, K. Almquist, M. A. Farinha, J. S. Lam, and A. M. Kropinski. 1994. Deletion and transposon mutagenesis and sequence analysis of the pR01600 OriR region found in the broad-host-range plasmids of the pQF series. Plasmid 31:265-74.

30. Kim, F., H. Mobley, M. Burken , and J. G. Morris . 1989 Molecular epidemiology of *Campylobacter pylori* infection in chronic care facility. Gastroenterology. 96 (Supp A256) abstract.

31. 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.

32. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying hydropathic character of a protein. J. Mol. Biol. 157:105-132.

33. Lambert, J.R., S.K. Lin , L. Nicholson , et al. 1990 High prevalence of *H. pylori* antibodies in institutionalized adults. Gastroenterology, 98 (Suppl A74) abstract.

34. Leunk, R. D. 1991. Production of a cytotoxin by *Helicobacter pylori*. Rev. Inf. Dis. 13:S686-S689

35. Leunk, R., P. Johnson, B. David, W. Kraft, and D. Morgan. 1988. Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. J. Med. Microbiol. 26:93-99.

36. Levine, M. M., J. P. Nataro, H. Karch, M. M. Baldini, J. B. Kaper, R. E. Black, M. L. Clements, and A. O'Brien. 1985. The diarrhea response of humans to some classic serotypes of enteropathogenic *E. coli* is dependent on a plasmid encoding an enteroadhesiveness factor. J. Infect. Dis. 152:550-559.

37. Logan, R. P. H., P. A. Gummett, B. T. Hegarty. 1992. Clarithromycin and omeprazole for *Helicobacter pylori*. Lancet 340:239

38. Maizel, J. V., Jr., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA. 78:7665-7669.

39. Mamelak, L., and H. W. Boyer. 1970. Genetic control of the secondary modification of deoxyribonucleic acid in *Escherichia coli*. J. Bacteriol. 104:57-62.

40. Marshal, B. 1994. Newer approaches to *Helicobacter pylori* infection. Inf. Medicine 11:27-31.

41. Marshall, B. J., L. J. Barrett, C. Prakash, R. W. McCallum, and R. L. Guerrant. 1990. Urea protects *Helicobacter pylori* from the bacterial effect of acid. Gastroenterology 99:697-702.

42. Marshall, B. J., H. Royce, D. J. Annear, C. S. Goodwin, J. W. Pearman, J. R. Warren, and J. A. Armstrong. 1984. Original isolation of *Camplobacter pyloridis* from human gastric mucosa. Microbiol. Lett. 25:83-88.

43. Marshall, B. J., J. R. Warren, and C. S. Goodwin. 1989. Duodenal ulcer relapse after eradication of *Campylobacter pylori*. Lancet i:836-837.

44. McIntire, S. A., J. Knesek, and S. Montanez. 1991. Plasmids of *Helicobacter pylori*. Prog. Abstr., p. 182, 91st Gen. Mtg. of the Am. Soc. Microbiol., Dallas, TX, May, 1991.

45. McIntire, S. A., and W. L. Peterson. 1990. Analysis of plasmid DNA in clinical isolates of *Helicobacter pylori*. Prog. Abstr., p. 36. The Ann. Mtg. of the Am. Gastroenterol. Assoc., San Antonio, Tx. , May, 1990.

46. Minnis, J. 1994. Cloning and DNA sequence analysis of a plasmid from *Helicobacter pylori*. PhD dissertation. Texas Woman's University, Denton, Texas.

47. Nakazawa, T., M. Ishibashi, H. Konishi, T. Takemoto, M. Shigeeda, and T. Kochiyama. 1989. Hemagglutination activity of *Campylobacter pylori*. Infect. Immun. 57:989-991.

48. Nieto, C., R. Giraldo, E. Fernandez-Tresguerres, and R. Diaz. 1992. Genetic and functional analysis of the basic replicon of pPS10, a plasmid specific for *Pseudomonas* isolated from *Pseudomnas syringae* patovar savstanoi. J. Mol. Biol. 223:415-426.

49. Papini, E., M. Bugnoli, M. De Bernard, N. Figura, R. Rappuoli, and C. Montecucco. 1993. Bafilomycin A1 inhibits *Helicobacter pylori* vacuolization of Hela cells. Mol. Microbiol. 7:323-327

50. Parsonnet, J., G. D. Friedman, D. P. Vandersteen, Y. Chang, J. H. Vogelman, N. Orentreich, and R. K. Sibley. 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. New Engl. J. Med. 325:1127-1131.

51. Penfold, S. S., A. J. Lastovica, and B. G. Elisha. 1988. Demonstration of plasmids in *Campylobacter pylori*. J. Infect. Dis. 157:850-851.

52. Razin, A., and A. D. Riggs. 1980. DNA methylation and gene function. Science 210:604-610.

53. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer directed enzymatic amplification of DNA with thermostable DNA-polymerase. Science 239: 487-491.

54. 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.

55. Sarosiek, J., A. Slomiany, and B. L. Slomiany. 1988. Evidence for weakening of gastric mucus integrity by *Campylobacter pylori*. Scand. J. Gastroenterol. 23:585-590.

56. Shames B, S. Krajden, M. Fuksa, C. Babida, J. L. Penner . Evidence for the occurrence of the same strain of *Campylobacter pylori* in the stomach and dental plaque. J. Clin. Microbiol. 27:2849-50.

57. Simor, A. E., B. Shames, B. Drumm, P. Shearman, D. E. Low, and J. L. Penner. 1990. Typing of *Campylobacter pylori* by bacterial DNA restriction endonuclease analysis and determination of plasmid profile. J. Clin. Microbiol. 28: 83-86.

58. Slomiany, B., J. Bilski and J. Sarosiek. 1987. *Campylobacter pylori* degrades mucin and undermines gastric mucosal integrity. Biochem. Biophys. Res. Commun. 144:307-314.

59. Smoot, D. T., H. L. T. Mobley, G. R. Chippendale, J. F. Lewison, and J.H. Resau. 1990. *Helicobacter pylori* urease activity is toxic to human gastric epithelial cells. Infect. Immun. 58:1992-1994.

60. Talley, N. J., A. R. Zinsmeister, A. Weaver, E. P. Dimagno, H. A. Carpenter, G. I. Perez-Perez, and M. J. Blaser. 1991. Gastric adenocarcinoma and *Helicobacter pylori* infection. J. Nat. Cancer Inst. 83:1734-1739.

61. Taylor, D. N., and Blaser, M. J. 1991. The epidmiology of *Helicobacter pylori* infection. Epidemiol. Rev. 13:42-59.

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

63. Tonneguzzo, F., S. Glynn, E. Levi, S. Mjolsness, and A. Hayday. 1988. Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. BioTechniques 6:460-469.

64. 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.

65. Tummuru, M. K. R., T. L. Cover, and M. J. Blaser. 1994. Mutation of the cytotoxin-associated *cagA* gene does not affect the vacuolating cytotoxin activity of *Helicobacter pylori*. Infect. Immun. 62:2609-2613.

66. Vaira, D., C. D'Anastasio, J. Holton, et al. 1988. *Campylobacter pylori* in abattoir workers: is it a zoonosis? Lancet 2:725-6.

67. Turbett, G. R., P. B. Hoj, R. Horne, and B. J. Mee. 1992. Purification and characterization of the urease enzymes of *Helicobacter pylori* species from humans and animals. Infect. Immun. 60:5259-5266.

68. Warren, J., and B. Marshall. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet i:1273-1275.

69. Wetherall, B., and A. Johnson. 1989. Haemolytic activity of *Campylobacter pylori*. Eur. J. Clin. Microbiol. Infect. Dis. 8:706-710.

70. Wolfgang, S., and R. Haas. 1994. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. Mol. Microbiol. 12:307-319.

71. Wotherspoon, A. C., C. Ortiz-Hidalgo, M. R. Falzon, P. G. Isaacson. 1991. *Helicbacter pylori*-associated gastritis and primary B-cell gastric lymphoma. Lancet 338:1175-1176.