Int. J. Cancer: **108,** 456–463 (2004)
© 2003 Wiley-Liss, Inc.
Publication of the International Union Against Cancer

IDENTIFICATION OF CANDIDATE ANTIGENS FOR SEROLOGIC DETECTION OF *HELICOBACTER PYLORI*-INFECTED PATIENTS WITH GASTRIC CARCINOMA

Alexander KRAH[1,2], Stephan MIEHLKE[3]*, Klaus-Peter PLEISSNER[4], Ursula ZIMNY-ARNDT[2], Christian KIRSCH[3], Norbert LEHN[5], Thomas F. MEYER[1], Peter R. JUNGBLUT[2]* and Toni AEBISCHER[1]

[1]*Department of Molecular Biology, Max Planck Institute for Infection Biology, Berlin, Germany*
[2]*Central Unit of Protein Analysis, Max Planck Institute for Infection Biology, Berlin, Germany*
[3]*Medical Department I, Technical University Hospital Carl Gustav Carus, Dresden, Germany*
[4]*Central Unit of Bioinformatics, Max Planck Institute for Infection Biology, Berlin, Germany*
[5]*Institute for Medical Microbiology, University Hospital Regensburg, Germany*

Helicobacter pylori **colonizes the stomach of almost half the world population and is a causative agent of gastric carcinomas and duodenal ulcers. Only a small fraction of infected people will develop these severe illnesses and a predictive test to identify people at high risk would greatly benefit disease management. Our study aimed to identify conserved bacterial antigens that may be useful for the development of such a diagnostic test. High-resolution immunoproteomics by 2-dimensional electrophoresis of *H. pylori* 26695 proteins was carried out with sera from infected patients with either duodenal ulcer (*n*=30) or gastric carcinoma (*n*=30), 2 clinically divergent conditions. According to their antigen recognition patterns clear groups of patients were identified. Although this classification did not correspond to the clinical status, it may be correlated to other bacterial or host factors that influence the outcome of infection. In general antigen recognition patterns were found to be highly variable, however by utilizing powerful image analysis and statistical tests the recognition of 14 antigenic protein species was found to differ significantly ($p<0.01$) between both diseases. Particular protein species of GroEL, HyuA, GroES and AtpA appear to be useful surrogate markers for gastric carcinoma detection and consequently should be considered for further prospective studies to assess their predictive value. For one protein species of AtpA, evidence was found that different post-translational modifications may confer different immunogenicities.**
© 2003 Wiley-Liss, Inc.

Key words: *immunoproteomics; bacterial virulence factors; duodenal ulcer; immunogenicity*

Helicobacter pylori, presumably the most widespread agent causing chronic bacterial infection in humans, causes in some cases duodenal ulcer (DU) or gastric carcinoma (GC).[1] It is estimated that 90–95% of DUs in Europe[2] originate from a *H. pylori* infection. Infection is also associated with a 2.7–12-fold increased risk of developing gastric cancer.[1]

Despite the fact that *H. pylori* is a common pathogenic factor for DU and GC, both diseases appear to be clinically divergent, *i.e.*, DU patients face a lower risk to develop GC than the normal population.[3–8] As shown in Figure 1, host, bacterial or environmental factors might determine the outcome of a long-term infection. It is known that GC mainly develops in individuals with reduced acid secretion. In contrast, acid hypersecretion is a known risk factor for DU.[9] Additionally, the high genetic diversity and differences in expression of virulence factors between different strains of *H. pylori* might contribute to the outcome,[10–12] for which precedence has been found in the Mongolian gerbil infection model.[13]

A diagnostic test, able to distinguish between infected persons under threat to develop one of the above serious clinical symptoms and *harmlessly* infected individuals, would be of great benefit. Because specific serum IgGs are very sensitive markers of infection, a simple serologic test would be favorable.[14] The antibody responses to hitherto known virulence factors are not capable of discriminating between DU and GC diseases and the existence of bacterial factors that correlate with clinical outcome has been questioned.[15]

We hypothesized that if *H. pylori*-specific serologic markers that were indicative for a clinical status existed, then the comparison of the antigen recognition patterns of sera from *H. pylori*-infected patients that developed either one of the clinically divergent conditions of DU or GC should be the most promising approach to identify such markers. Indeed, 2 studies revealed antibody responses against certain antigenic proteins that differed between GC and DU patients,[16,17] suggesting that this approach might be fruitful. Immunoproteomics, *i.e.*, a combination of classic 2-dimensional electrophoresis (2-DE) and immunoblotting[18] is the method of choice to test this hypothesis. The reactivity of sera against the sequenced *H. pylori* strain 26695 was investigated.[10,17,19] The rational for this approach was that a diagnostic tool would depend on antigens conserved between strains and identification of such antigens would be most straight forward using a sequenced strain and reference to the public proteome database (http://www.mpiib-berlin.mpg.de/2D-PAGE), which is based on the high-resolution power of 23 cm × 30 cm 2-DE gels where more than 1,800 protein species of *H. pylori* 26695 are described compared to 1,590 predicted genes.[19]

Here we compared high-resolution immunoblots stained with sera from patients with DU (*n*=30) and GC (*n*=30). Patients could be grouped to global features of the antigen recognition patterns. These patterns did not correlate with disease but may be indicative of so-far unappreciated relations between patients or *H. pylori* strains. Despite strongly varying antigen recognition patterns, we found 14 antigens with significant differences in recognition between sera from GC and DU patients.

MATERIAL AND METHODS

Patients

For our retrospective study, we used serum samples that have been obtained from consecutive *H. pylori*-positive patients with either DU or GC who participated in previous clinical trials.[20–23]

Grant sponsor: Bundesministerium für Bildung und Forschung; Grant numbers: 031U/107, 031U/207; Grant sponsor: European Union; Grant number: QLRT-1999-31536

*Correspondence to: Peter R. Jungblut, Central Unit of Protein Analysis, Max Planck Institute for Infection Biology, Campus Charité Mitte, Schumannstrasse 21/22, 10117 Berlin, Germany. Fax: +49-30-28460507;
E-mail: jungblut@mpiib-berlin.mpg.de; Stephan Miehlke, Medical Department I, Technical University Hospital Carl Gustav Carus, Fetscherstrasse 74, 01307 Dresden, Germany. Fax: +49-351-4584394.
E-mail: miehlke@mk1.med.tu-dresden.de

DOI 10.1002/ijc.11557

All patients were white and recruited from a comparable catchment area at clinical centers throughout the southern part of Germany between 1995 and 2001. The diagnosis of *H. pylori* infection and the confirmation of gastric adenocarcinoma by histology was established by a central study pathologist. Sera samples were obtained at the time of clinical diagnosis and sent within 48 hr to the Central Department of Medical Microbiology (University Hospital Regensburg) where they were stored at −20°C until further usage. General exclusion criteria for patient recruitment to the study were previous attempts to eradicate *H. pylori*, use of antibiotics, proton pump inhibitors or bismuth compounds within the last 2 weeks prior to endoscopy, and previous gastric surgery. Patients with GC underwent complete clinical staging procedures including endoscopic ultrasound and CT scan. Parameters such as age of patients, sex and degree of colonization were recorded (Table I). Apart from the age distribution, both patient groups were highly comparable. On average, GC patients were older than DU patients; however, 2/3 of patients lie within an overlapping age range of 41–71 years.

H. pylori *cell culture and lysis*

After inoculation and 3 days growth on agar plates containing 10 μg/ml Vancomycin, single clones were resuspended in 1 ml BHI medium containing 10% FCS, and 20 μl of this suspension were grown for 2 days on Vancomycin containing agar plates under microaerobic conditions (5% O_2, 10% CO_2, 85% N_2) at 37°C. Bacteria were transferred into 50 ml cold PBS containing 1 tablet of Complete protease inhibitors (Roche, Basle, CH). After centrifugation at 3,000*g* and 4°C for 10 min and one wash step in 10 ml protease inhibitor containing PBS, the supernatant was omitted. The bacteria containing pellet was diluted with half a volume of distilled water and lysed by addition of urea, CHAPS, Servalyte pI 2-4 (Serva, Heidelberg, Germany) and DTT to obtain final concentrations of 9 M, 1.4%, 2% and 70 mM, respectively. For solubilization cells were shaken for 30 min at room temperature and insoluble components were separated by centrifugation at 100,000g for 30 min. The clear supernatants were stored in aliquots at −70°C.

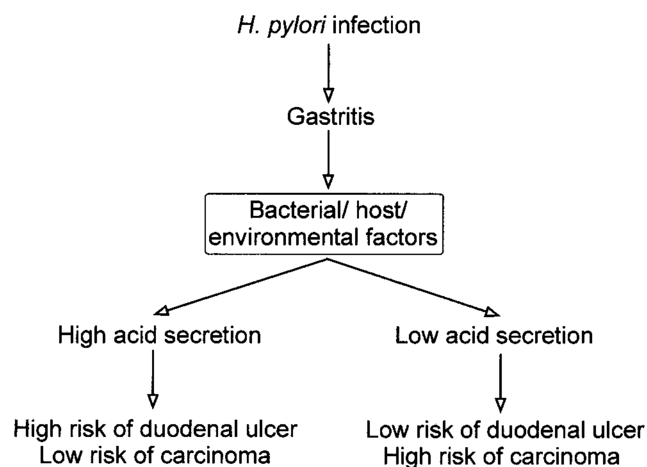

FIGURE 1 – Two clinically divergent outcomes of the *H. pylori* infection. All infected people show signs of gastritis even without any clinical symptoms. The outcome is likely determined by bacterial and/or host and/or environmental factors. Most persons do not develop any disease. However, people with high acid secretion have a higher risk to develop duodenal ulcer rather than gastric carcinoma. People with low acid secretion have a risk to develop cancer but not duodenal ulcer.

Two-dimensional electrophoresis

Proteins were separated using a 23 cm × 30 cm high resolution gel system with a resolution power of up to 5,000 spots as previously described.[10,24] For immunoblotting 150 μg of protein were loaded on the gels. For preparative gels up to 500 μg of protein were applied and the gel was stained with Coomassie Brilliant Blue (CBB) G-250 as described.[25]

Immunoblotting

Gels were cut into 2 pieces [high and low molecular weight (MW)] and blotted onto PVDF membranes (ImmobilonP, Millipore, Eschborn, Germany) using a semi-dry blotting system (Hoefer Large SemiPhor, Amersham Pharmacia Biotech AB, San Francisco, CA).[26] The following blotting buffers were used: for the high MW part of the gel—cathode buffer: 50 mM boric acid, 10% methanol, 5% SDS, pH 9.0; anode buffer: 50 mM boric acid, 20% methanol, pH 9.0; and for the low MW part of gel—cathode and anode buffers: 100 mM boric acid, 20% methanol, pH 9.0. Blotting time was 2 hr at 1 mA/cm^2. After blotting, the 2 membrane halves from 1 gel were sealed into 1 foil bag in order to minimize the volume needed for processing (50 ml). Membranes were blocked for 1 hr in 5% skim milk in PBST (PBS buffer containing 0.05% Tween-20). Serum was added directly to the solution to reach a dilution of 1:250. The bags were covered with glass plates on a shaker so that solutions circulated properly within the packages by shaking overnight at 4°C. Afterwards bags were opened and membranes washed 4 times for 15 min in PBST. Incubation with secondary antibody diluted in 5% milk in PBST buffer followed for 1 hr (1:5,000 Goat anti human polyvalent IgG—Peroxidase Conjugate, Sigma A-8400, Deisenhofen, Germany). Membranes were washed again 4 times for 15 min in PBST buffer. Freshly mixed ECL reagent (Western Lightning Chemoluminescence Reagent NEL-101, NEN, Perkin Elmer, Boston, MA) was applied to the membranes and shaken for 1 min at room temperature. Excess fluid was dripped off the membranes and these were wrapped in plastic foil. Films were exposed to the membranes for 90 sec. To assess the quality of blotting, the PVDF membranes were stained for 5 min in CBB R-250 staining solution (50% methanol, 10% acetic acid, 0.1% CBB R-250 (BioRad, Hercules, CA) followed by destaining in 50% methanol, 10% acetic acid 3 times for 2 min. Dried membranes were stored at room temperature.

Immunoblot analysis

After scanning the films (8 bit gray values, 100 dpi), the 2-DE gel analysis software PDQuest (Version 7.1, BioRad, Hercules, CA) was applied. Spot detection and quantification was done automatically by fitting spot intensities with a 2-dimensional Gaussian model. Corresponding spots in different blots were matched using a distortion model that takes local gel running differences into account. Both, spot detection and matching, were thoroughly checked and corrected manually in an interactive manner. This time-consuming step was essential to achieve reliable results.

TABLE I – CHARACTERISTICS OF PATIENTS WHOSE SERA WERE USED FOR THIS STUDY*

Patient group	Number of patients	Age range	Average age (S.D.)[1]	Number of males	Number of females	Hp status[2]	Degree of colonization[3] (S.D.)
GC	30	41–87	66(±10)	14	16	All pos.	2.3 (±0.8)
DU	30	24–71	47(±13)	15	15	All pos.	2.0 (±0.8)

*Only average ages differ significantly between the 2 groups.–[1]Standard deviation.–[2]*H. pylori* status was examined by histology. All patients were found to be infected by *H. pylori*.–[3]Average degree of colonization of the stomach. Possible degrees were 1, 2 or 3 as assessed by histology.

For the quantitative data analysis of spot intensities, a univariate statistical *t*-test as well as multivariate statistical approaches (correspondence analysis and hierarchical agglomerative cluster analysis) were applied; *t*-tests were performed in PDQuest or in SigmaPlot (SPSS Science, Chicago, IL) where sera were grouped according to the medical status of patients, *i.e.*, a GC and a DU group with 30 members each. Multivariate statistical approaches and their applications for gel data analysis have been described in detail.[27–29] A software package for correspondence analysis written in DELPHI and running on any Windows PC has been developed by R. Wessel and K.-P. Pleissner. For hierarchical agglomerative cluster analysis the statistical programming environment R (http://www.r-project.org), running on a locally installed Rweb-server, was used. In this case, the data matrix comprises the intensities of 611 protein spots that occur in 60 immunoblots. For normalization, all spot intensities greater than 0 were set to 1 and the hierarchical clustering was performed using complete linkage and a binary distance metric.

Antigen identification

Antigenic protein species of interest were identified by 1 of 2 approaches: First, spots and spot patterns on the immunoblots were compared to spots and spot patterns in the 2D-PAGE database (http://www.mpiib-berlin.mpg.de/2D-PAGE). Second, spots of interest were excised from a preparative gel, digested and identified by peptide mass fingerprinting (PMF) using a MALDI-TOF mass spectrometer (Voyager Elite DE, Applied Biosystems, Framingham, MA).[30] For the digest, spots were first destained in 200 mM NH_4HCO_3, 50% acetonitrile and then in 50 mM NH_4HCO_3, 5% acetonitrile (digest buffer) for 30 min at 37°C. After this, spots were dried in a SpeedVac and 25 μl of Trypsin (Sequencing grade modified Trypsin, Promega, Madison, WI, diluted in digest buffer to achieve 0.2 ng/μl) solution was applied directly to the dried spot. Spots were digested overnight at 37°C. Supernatants were transferred to fresh tubes. In order to recover all peptides the remaining spots were shrunk in 25 μl 60% acetonitrile, 0.1% trifluoroacetic acid for 10 min and the supernatants were combined and dried in a SpeedVac at 45°C. The peptide pellet was dissolved in 1.3 μl 33% acetonitrile, 0.1% trifluoroacetic acid and 0.25 μl were applied to the MALDI template. Finally, 0.25 μl matrix solution consisting of 50 mg/ml 2,5-dehydroxybenzoic acid dissolved in 33% acetonitrile and 0.33% trifluoroacetic acid were added. The PMFs were obtained using the following parameters: 20 kV accelerating voltage, 70% grid voltage, 0.08% guide wire voltage, 200 ns delay time and a low mass gate of 500 Da. The database searches were performed with Mascot (http://www.matrixscience.com) and ProFound (http://129.85.19.192/profound_bin/WebPro Found.exe). Search parameters were 100 ppm peptide mass tolerance, 2 missed cleavages and possible oxidation of methionine. Criteria for a reliable identification were significant score values when performing "all species" search and a minimum of 30% sequence coverage.

RESULTS

Two-dimensional gels and immunoblots reveal different subproteomes from H. pylori

In order to find candidate antigens for a diagnostic test, antigens recognized by sera from *H. pylori*-infected patients suffering from either DU or GC were compared. *H. pylori* (strain 26695) proteins resolved by high-resolution 2-DE immunoblot analysis (23 cm × 30 cm) were probed with individual patient sera. The 2-DE technique resolved about 1,800 *H. pylori* protein species (see silver stained gel in Figure 2*a*). About 200 of these spots have already been identified and published in our 2D-PAGE database. The spot patterns of immunoblots (Fig. 2*b*) differ significantly from that of a silver-stained gel (Fig. 2*a*). Immunoblot analysis is highly sensitive and specific; it only reveals spots that are recognized by antibodies contained in the patient serum and, therefore it discloses a different subproteome of *H. pylori*. Only a small proportion of the Helicobacter proteins are antigenic – the average number of spots recognized by the sera was 142 (median 141). Both, highly abundant proteins such as the Urease beta subunit or GroEL (shown by arrowheads in Fig. 2) and antigens that are not detectable by silver staining were visualized. One example for the latter class of antigens was a series of spots diagonally distributed in the upper-left part of Figure 2*b* (see arrow). These antigenic protein species were not detected by silver or CBB staining in the gel and therefore presumably contained less than 1 ng protein per spot which is the detection limit of silver staining.

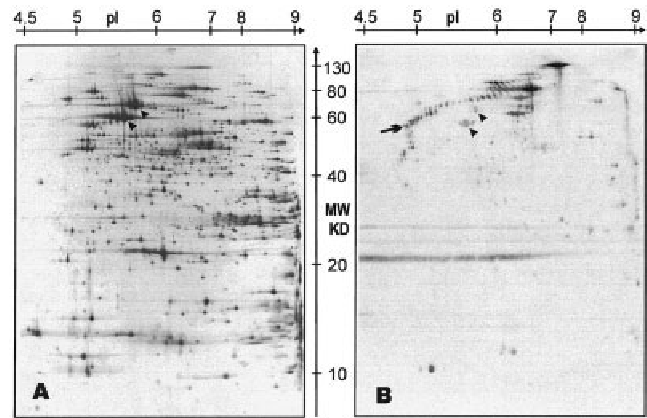

FIGURE 2 – Comparison of a large 2-DE gel (*a*) and a large 2-DE immunoblot (*b*). (*a*) Silver stained 23 cm × 30 cm gel from *H. pylori* 26695 lysate. It contains about 1,800 spots and identified spots can be explored interactively in our public 2D-PAGE database (http://www.mpiib-berlin.mpg.de/2D-PAGE/). (*b*) 23 cm × 30 cm Immunoblot probed with a GC serum (GC02). In this blot, 177 spots are recognized. Arrowheads show the proteins Urease beta (upper arrowhead) and GroEL (lower arrowhead) for orientation. Owing to high sensitivity and selectivity, spots can be seen that are not present in the silver gels, *e.g.*, the diagonal spot series in the upper left part of the blot (see arrow). In contrast to this, there is only a weak recognition of the 2 strongest spots in the silver gel (see arrowheads). pI: isoelectric point; MW: molecular weight in kDa.

Antigen recognition patterns of individual patient sera differ enormously

In Figure 3, 6 immunoblots have been chosen to illustrate the variation of recognition patterns. The upper (GC02, 09 and 22) and the lower (DU14, 15 and 29) 3 blots were incubated with sera from patients with GC and DU, respectively. The figure also shows examples of patterns that we found repeatedly and that were highly reproducible. For example, the aforementioned diagonal series of spots in the upper-left part of blots GC02 and DU29 was found in 37 of 60 blots and may therefore define a subgroup of patients or reflect infection with particular *H. pylori* strains. These subgroups did, however, not correspond to disease states. Overall 611 different spots were recognized by all of the sera and the number recognized by a given individual serum ranged from 24 to 391. On average sera from GC patients recognized a 49% higher number of spots with a concomitant increase in intensities of recognition (75% higher sum of spot intensities). One single spot was recognized by all of the 60 sera, the main species of GroEL, while for example CagA was detected by 58 and urease beta by 57 sera. The great majority of antigenic protein species (531 spots) were detected by less than half of the sera and 32 spots were found only once. The analysis of sera from 30 patients of both well-defined clinical manifestations allowed us to estimate the variance of recognition for particular proteins and analyze the data using statistical tools.

Comparison of entire recognition patterns

Multivariate statistical analysis methods were applied in order to cluster sera according to their recognition patterns without any

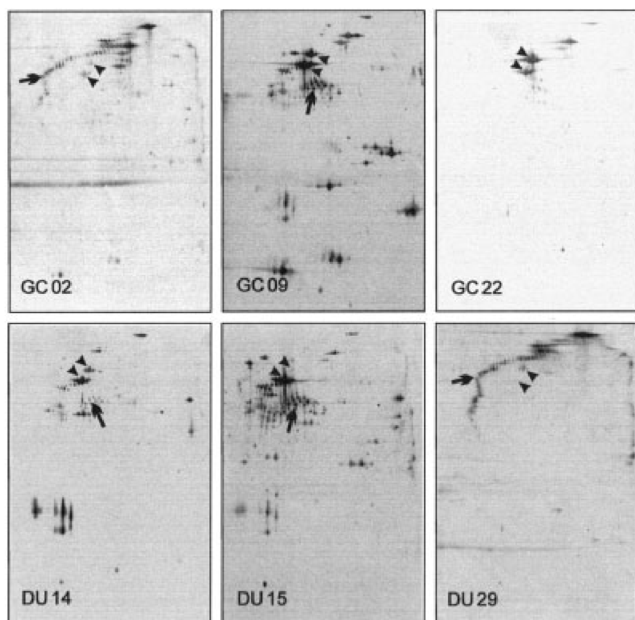

FIGURE 3 – Large 2-DE blots from 6 arbitrarily chosen patient sera from gastric carcinoma (GC02, 09 and 22) and duodenal ulcer patients (DU14, 15 and 29). These 6 immunoblots are not representative for all of the 60 sera since recognition patterns differ enormously. The sera in our study recognized between 29 and 383 spots. GC02 is also shown in Figure 2. Note the different numbers and intensities of spots and differences of background recognition. Arrowheads show the proteins Urease beta and GroEL for orientation (see also Fig. 2). Interestingly, there are recognition patterns to be seen that are recognized repeatedly, *e.g.*, a series of spots diagonally distributed in the upper left parts of GC02 and DU29 or a "spot-cohort" (below GroEL) seen in GC09, DU14 and DU15 (shown by arrows).

input of data concerning the patients clinical status. This analysis aimed for groups of sera with potentially related recognition patterns. Two methods were used to investigate global similarities of blots: correspondence analysis and hierarchical agglomerative cluster analysis. The correspondence analysis, where similarities of objects can be detected by plotting typical characteristics into a 2-dimensional factorial space, revealed one cluster that contained only a subgroup of GC patients (data not shown). Other GC and DU sera were equally distributed within the factorial space. The hierarchical agglomerative cluster analysis using complete linkage and binary distance metric produced a cluster dendrogram where 5 closely related groups of sera could be assigned (numbers 1–5 in Fig. 4). This analysis revealed clusters of patients with similar recognition patterns (Fig. 3). The diagonally distributed spot series is present in immunoblots containing clusters 2 and 3 whereas the "spot-cohort" is found in clusters 2 and 5. In contrast, clusters 1 and 4 contain very few spots with low intensities, exhibit neither of these recognition patterns but differ in recognition of several other spots. The immunoblots shown in Figure 3, for instance, fit into these clusters. Blots GC02 and DU29, which contain the spots diagonally distributed, are to be found in cluster 3, whereas blots GC09, DU14 and DU15 that contain the "spot-cohort" belong to cluster 5. Immunoblot GC22, which contains neither of both patterns, falls into cluster 4.

Although these global statistical analyses did not reveal a patient's clinical status, they point to the existence of patient groups with similar recognition patterns, a fact that has not been described before. This result does not exclude, however, the existence of individual spots that cluster with disease.

Fourteen spots were differentially recognized with statistical significance by sera from GC and DU patients

The knowledge of the variance of antigen recognition intensities gave us the opportunity to perform univariate statistical *t*-tests on all spots individually, which is a more powerful tool to evaluate differential recognition of individual protein species. To this end, the 60 sera were divided into 2 groups according to the conditions of the patients (GC or DU); *t*-tests were performed at different significance levels and the respective results are shown in Table II. In order to select the spots that were truly recognized differently by sera from the 2 groups of patients a significance level of $>99\%$ ($p<0.01$) was used. Only 14 spots were differentially recognized and fulfill this criterion. These are marked by arrows in Figure 5. The master blot shown there contains all 611 spots of all immunoblots. It is a virtual image of Gaussian fitted spots, which was produced by the 2-DE gel analysis software PDQuest. Intensity data of this selection of spots for individual sera are shown by bar charts. The inner boxes contain intensity columns of each spot in all sera: the 30 columns on the left represent GC sera (C) and the 30 on the right represent DU sera (U). All of the 14 significantly different spots were more strongly recognized by sera from GC patients. This was corroborated by calculating the average spot intensities in the 2 groups (columns in the narrow boxes on the outer edges of Fig. 5). The black marks in the middle of these columns represent the average intensity values and the zone filled with white represents the ranges of standard deviations. The spot numbers (SSP) are automatically produced by PDQuest and given below the inner boxes. The differentially recognized spots fall within a wide range of molecular weights (12–80 kDa) but a rather narrow pI range (4.9–6.5).

Sex and degree of colonization had no influence on the analysis

Highly variant recognition patterns of individual sera could also result from other patient parameters than disease. Although both patient groups did not differ in sex ratios and degrees of *H. pylori* colonization *t*-tests were carried out to rule out that these parameters would cluster with spot intensities. Therefore, patients were grouped according to their sexes and degree of *H. pylori* colonization, respectively. Using a *t*-test with the significance level of $p<0.01$ as above, no spots at all were found to be differently recognized. Even when a less stringent significance level of $p<0.05$ was applied none of the 14 between disease groups differentially recognized spots were found. This test shows that these parameters have no influence on our data analysis.

The age of patients had limited influence on the data

The average ages of GC and DU groups were significantly different. Therefore, we examined the influence of this patient parameter closely. First sera from the 30 older and from the 30 younger patients were grouped independent of diseases and *t*-tests were applied. No spot was found to be differently recognized at a significance level of $p<0.01$. Next, age matched subgroups of 18 GC and 19 DU sera were analyzed (average ages of these subgroups: 60 (S.D. $\pm$7) and 55 (S.D: $\pm$9), respectively; the difference was not significant: $p=0.07$); *t*-tests validated 9 (5 with $p<0.01$ and further 4 with $p<0.05$) of the 14 spots that clustered with disease in the analysis using all 60 sera. The remaining 5 spots of the 14 were not found to cluster with disease using the age matched subset of 37 sera (spots 1511, 2505, 4404, 4706 and 5708; marked in Table III). Therefore, the patient age had an influence on the recognition at least when the latter 5 spots were analyzed but could not account for all differences observed.

Ten antigenic protein species of interest were identified

The antigenic spots that were found to be significantly stronger recognized by sera from GC patients (shown in Fig. 5) were identified by pattern comparison or mass spectrometry. The pattern comparison was done by comparing the relative distance between the spot of interest and proximate spots in the immunoblot with the standard gel in the 2D-PAGE database. In cases where no corre-

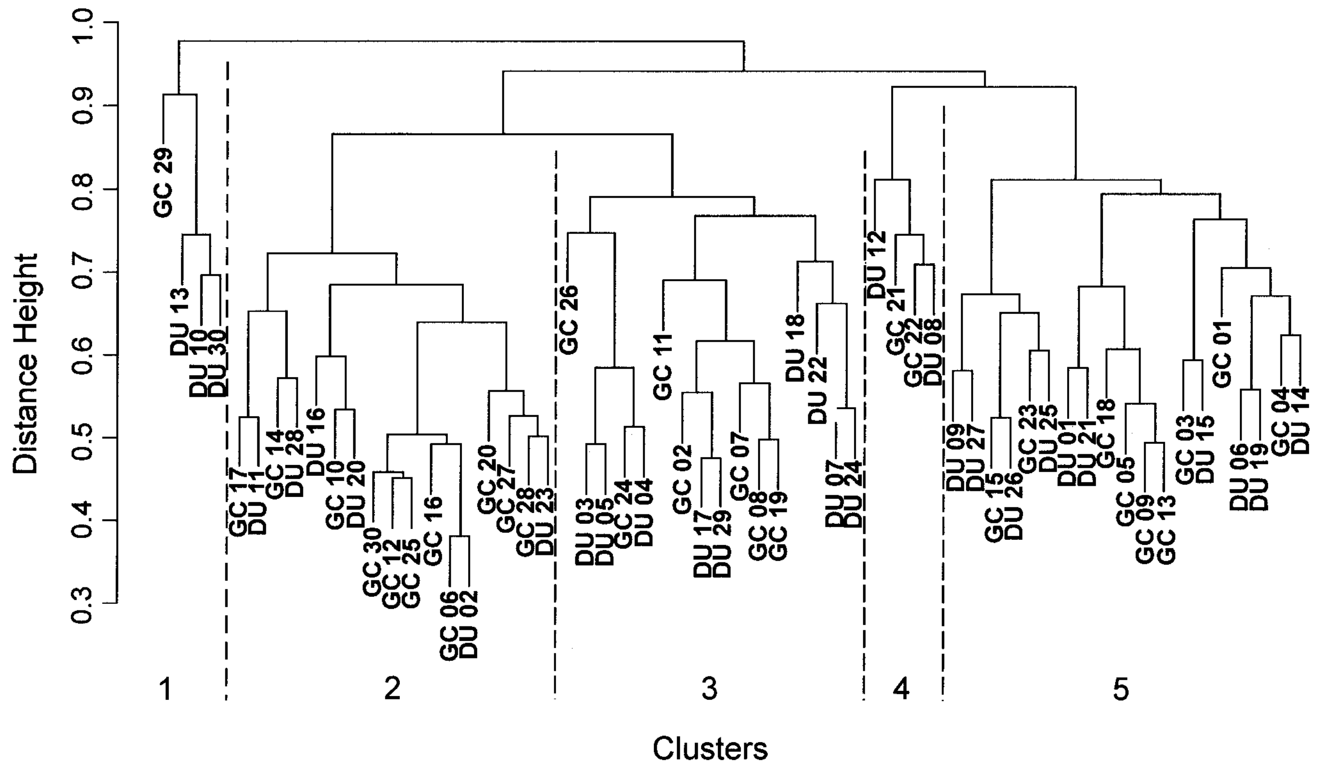

FIGURE 4 – Hierarchical agglomerative cluster analysis of immunoblots. The dendrogram shows relations between the global recognition patterns of the 60 immunoblots (labeled GC01-30 and DU01-30). Five groups of immunoblots that are closely related were assigned as shown by numbers 1–5. Although these groups do not represent clinical parameters of patients, they represent recognition patterns as seen in Figure 3. The series of spots diagonally distributed in the upper left part of some blots is present in groups 2 and 3, whereas the "spot-cohort" is found in 2 and 5. Groups 1 and 4 contain only a few spots, show neither of these patterns but differ in recognition of several other spots.

TABLE II – ANALYSIS OF IMMUNOBLOTS*

Recognition strength in GC sera compared to DU sera	Number of spots by factor		Number of spots by t-test		
	2-fold	3-fold	$p < 0.1$	$p < 0.05$	$p < 0.01$
Stronger	315	192	144	78	14
Weaker	49	35	2	0	0

*The table shows the numbers of differentially recognized antigens (spots) according to different methods of analysis. Note that there are major differences between analyses simply by a factor and statistical analyses.

sponding spot was found in the database or the spot's identity was not yet known, the antigen was identified from a preparative gel by peptide mass fingerprint (PMF) using a MALDI-TOF mass spectrometer. Table III gives the list of the protein species identified, which comprise the following open reading frames (ORFs): AtpA, GroEL, GroES and HyuA. Two of the 14 antigenic species could not be assigned to a spot or spot pattern in the preparative gel and 2 spots could not be identified unambiguously by PMF. Therefore a total of 10 antigenic protein species that were differentially recognized were identified which can be explored interactively in our 2D-PAGE database.

Not all protein species of the identified ORFs were differentially recognized between GC and DU sera

Separation of cell lysates by 2-DE results in a pattern where spots ideally contain individual proteins. However, some ORFs produce several spots with differences in pI and/or MW and cohorts or chains of spots appear. This may be the result of amino acid exchanges or posttranslational modifications such as phosphorylations, glycosylations, partial degradations or others. Individual protein species of a single ORF may have different immunogenicities or antigenicities that will result in different spot intensities on the immunoblots. Two of the 4 identified ORFs exhibit such an effect (Table IV). It must be taken into account, however, that protein species strongly differ in their relative concentration and low abundant spots may be recognized just above the detection limit in 1 immunoblot but might fall below this limit in another, and differential recognition may just reflect the relative concentrations of antibodies. One example for this class might be GroEL where the 3 species of interest account for only 10–25% of the protein content of the main spot (protein content assessed in the silver-stained reference gel of our database). The AtpA species of interest, however, has 43% of the protein content compared to the highest intensity species. Four species of AtpA with lower protein content were also recognized by sera but showed no difference between disease groups. In this case, we conclude that the different protein species have differing immunogenicities in the 2 disease groups.

Search for a set of antigens of potential diagnostic value

The set of 14 antigenic protein species identified in the statistical analysis was the starting point to evaluate combinations of these antigens in a potential diagnostic test kit. We performed a hypothetical serologic test and asked whether recognition of a given subset of antigens would diagnose most cases of GC in our test set of 60 sera. Manual testing of different subsets identified the following 5 spots as potential candidates: 2509 (1 species of GroEL), 4003 (1 species of GroES), 4005 (another species of GroES), 4706 (not id.) and 5708 (one species of HyuA). Based on the criterion that at least 3 out of these 5 spots had to be recognized by the serum under investigation, we detected cases of GC with a sensitivity of 77% and a specificity of 83%.

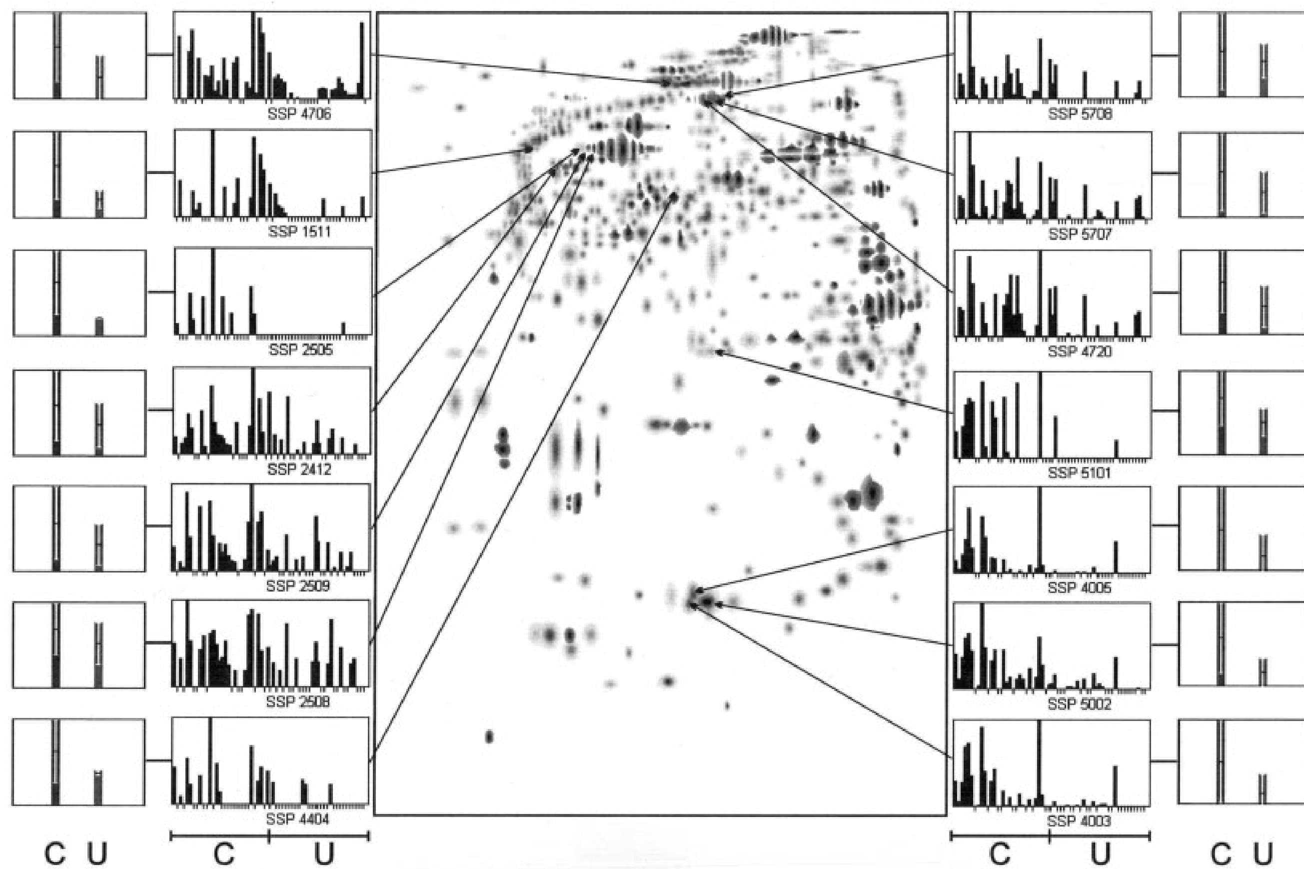

FIGURE 5 – Immunoblot master image with spot intensities of the 14 differentially recognized spots. The master blot is a virtual image of Gaussian fitted spots and contains all of the 611 different spots of the 60 immunoblots in the analysis set. Arrows point towards the 14 between gastric carcinoma and duodenal ulcer sera differentially recognized spots (*t*-test, $p<0.01$). Wide boxes show the spot intensities of each of the 60 sera and the narrow boxes show average spot intensities in both groups. The standard deviation range is given in white within the columns in the narrow boxes. The SSP numbers below the boxes are the spot numbers that were automatically generated by PDQuest. Note that column heights are normalized to the maximum intensity in each box. This figure was produced using graphs exported from the PDQuest software. C: gastric carcinoma sera; U: duodenal ulcer sera.

TABLE III – LIST OF IDENTIFIED PROTEINS FROM THE 14 SIGNIFICANTLY DIFFERENT RECOGNIZED SPOTS (*t*-TEST, $p < 0.01$)

Spot number (SSP)[1]	TIGR locus[4]	Protein name	Short name	Method of identification
1511[2]	—	Not identified[5]	—	—
2412	Hp1134	ATP synthase F1, subunit alpha	AtpA	PMF[7]
2505[2]	Hp0010	Heat shock protein 60	GroEL	database[8]
2508	Hp0010	Heat shock protein 60	GroEL	database
2509[3]	Hp0010	Heat shock protein 60	GroEL	database
4003[3]	Hp0011	Heat shock protein 10	GroES	database
4005[3]	Hp0011	Heat shock protein 10	GroES	database
4404[2]	—	Not assigned[6]	—	—
4706[2,3]	—	Not identified	—	—
4720	Hp0695	Hydantoin utilization protein A	HyuA	database
5002	Hp0011	Heat shock protein 10	GroES	database
5101	—	Not assigned	—	—
5707	Hp0695	Hydantoin utilization protein A	HyuA	database
5708[2,3]	Hp0695	Hydantoin utilization protein A	HyuA	database

[1]Spot number automatically assigned by the software PDQuest.–[2]The older age of GC patients might have contributed to the significance of these spots.–[3]This subset of 5 spots was used as hypothetical serologic test.–[4]The gene locus is given according to the TIGR database of *H. pylori* 26695 (http://www.tigr.org/tigr-scripts/CMR2/GenomePage 3.spl?database=ghp).–[5]Criteria for unambiguous identification with PMF were not reached.–[6]The spot on the immunoblot could not unambiguously assigned to a spot or spot region in the preparative gel.–[7]Peptide mass fingerprinting by MALDI-TOF mass spectrometry.–[8]Identification was done by finding the corresponding spots by comparing the spot patterns on blots to the standard gel in our 2D-PAGE database.

TABLE IV – FOUR ANTIGENS WHICH ARE STRONGER RECOGNIZED BY GC SERA COMPARED TO DU SERA (t-TEST $p < 0.01$)

Short name	TIGR cellular roles[1]	Protein family according to Swiss-Prot. database[2]	Number of protein species[3]	Known marker for GC or DU[4]
AtpA	Energy metabolism, ATP-proton motive force interconversion	ATPase alpha/beta chains family	1/9	Not mentioned
GroEL	Protein fate: protein folding and stabilization	Chaperonin (hsp60) family	3/10	+[16,17] −[31]
GroES	Protein fate: protein folding and stabilization	GroES chaperonin family	3/3	−[31]
HyuA	Amino acid biosynthesis: other	Oxoprolinase family	3/3	+[17]

[1]See Table III for reference.–[2]http://www.expasy.org/sprot.–[3]number of protein species that are differentially recognized compared to the approximate number of all protein species seen in the standard gel.–[4]all proteins are known antigens of *H. pylori.* References are given for the two proteins whose immune recognition was found to correlate with carcinoma (+). Two proteins were not found to be markers in another study (−).

DISCUSSION

Although *H. pylori* is known to be a causative agent of both duodenal ulcers and gastric carcinomas, it is not clear what factors determine these divergent outcomes of infection. Here, we tested in a retrospective study the hypothesis that serologic markers exist which could be of diagnostic value and reflect clinical status or outcome. Such diagnostic markers would be welcome candidates for future prospective studies. Two-DE high-resolution immunoblots containing proteins of *H. pylori* 26695 were probed with sera from *H. pylori* positive patients with DU or GC. We found that the recognition patterns of individual sera differed enormously and confirmed that sera from GC patients recognized more antigens and that the recognition signal was commonly stronger than seen for sera from DU patients. This lead on average to 49% higher numbers of spots recognized with a 75% higher sum of recognition signal intensities. Normalizing these data would have facilitated analysis but due to the high variation between individual sera no set of antigens lent itself to be used as a normalizing parameter. Multivariate statistical analysis (hierarchical agglomerative cluster analysis), however, enabled us to identify clear groups of sera with characteristic recognition patterns (*e.g.*, diagonally distributed spots and the "spot-cohort," Figs. 3 and 4). Although these clusters did not distinguish the 2 clinical groups, they may reflect characteristics of strains the patients were infected with, *e.g.*, expression of genes from the variable *H. pylori* gene pool and could be useful in epidemiological studies complementing genotyping data.

Recognition of 14 protein species corresponding to at least 4 ORFs did however differ significantly between GC and DU sera (Fig. 5 and Table III). They were found using univariate t-tests with $p<0.01$, a stricter significance level than used in our previous study. This criterion was considered to be appropriate for our goal to find antigens that correlate with disease for a number of reasons. First, recognition patterns showed high variability between individual sera. The fact that out of 611 antigenic protein species only a single one (GroEL, main spot) was recognized by all sera illustrates this well. Second, no other patient parameter (sex, age or degree of colonization) showed correlated spots at this level of significance. Age was of particular relevance since at the time of diagnosis GC patients were on average older than DU patients and hence may have carried *H. pylori* for longer periods of their life. Indeed, in a comparison of the recognition patterns of sera from age-matched subsets of patients a correlation with age could not be ruled out for 5 of the 14 spots (one species of HyuA and GroEL and 3 unknowns). However, in the same analysis, the recognition of the other 9 spots was correlated with disease ($p<0.05$) and not with age. One study[16] reported that also GroES recognition correlated with age but our results did not confirm this finding. It is clear that the differential recognition of the proteins described here could be a consequence of the diseases rather than being correlated with their potential cause. Our study was not designed to discriminate between these possibilities but with the candidates identified this will be possible in the future.

It was surprising that the 10 identified proteins were derived from only 4 ORFs, namely, AtpA, GroEL, GroES and HyuA, that encode relatively conserved proteins of the central metabolism or of protein fate (Table IV). Products of these ORFs are known from previous studies to induce antibodies in infected patients. In one study, GroEL and GroES were shown to be similarly recognized by sera from GC patients and patients with either gastric or duodenal ulcer. This result is probably due to the low number of sera analyzed (4 and 5, respectively).[31] Interestingly, GroEL was found to be significantly stronger recognized by sera from GC patients in a study where sera from 28 GC patients were compared to samples from 30 DU patients in a 1-dimensional Western blot analysis[16] and our group has reported that recognition of one species of GroEL correlated with disease when comparing sera from 9 ulcer and 6 cancer patients.[17] This latter result suggests different antigenicities for different GroEL species. In the same study, our group also reported that HyuA was differentially recognized by sera from GC or DU patients, which could be confirmed here.

It is of interest that in some cases only particular species of a given protein, *e.g.*, GroEL and AtpA were found to be differentially recognized by DU and GC sera. For GroEL, 3 of 10 species were recognized in this way and these belong to the spots with the lowest protein content based on their staining with silver. Therefore, differential recognition of these protein species may reflect a quantitative rather than a qualitative difference of the relevant antibodies in sera from GC *vs*. DU patients. However, this does not apply to AtpA which was represented by at least 9 antigenic species. The spot that was differentially detected showed medium spot intensities in the silver-stained gel. This finding is consistent with the idea that certain species of one protein are useful diagnostic markers whereas other species are not. Differential recognition may be the result of differences between protein species in certain epitopes due to amino acid exchanges, posttranslational modifications such as phosphorylations, glycosylations or partial degradations. The specific recognition of this particular AtpA species may therefore reflect a differential immunogenicity of such a modification. The same may hold true for the differential recognition of a particular region of GroEL by sera from *H. pylori* infected patients which was not reported for sera from noninfected patients.[32]

The approach used in our study allowed us to define a group of 5 protein species that when applied retrospectively as markers to our set of 60 patients allowed discrimination of GC and DU patients with 77% sensitivity and 83% specificity. Although this confidence level is far from being satisfactory, the result shows that a diagnostic test able to distinguish between *H. pylori* infected GC and DU patients may be achievable.

Expression of virulence factors, age at infection and their duration, immune response, level of acid secretion, and environmental factors have all been proposed to influence the outcome of *H. pylori* infection,[4,6,9] but their relative contributions are not yet clear. We found no association between outcome and recognition of known virulence factors, which agrees with the current view that the presence of the respective genes in *H. pylori* does not correlate with the clinical status.[15] Instead we detected a remark-

able increase in total seroreactivity. Proinflammatory alleles of the IL-1β locus were most consistently associated with increased GC-risk[33] and increased inflammation is likely to boost the antibody production and provides a rational for our immunoproteome results. The confirmed 2 diagnostic marker candidates for GC, GroEL, HyuA, and the new candidates, AtpA and GroES, are all highly conserved antigens with no obvious role in virulence and may therefore be regarded as surrogate markers for the clinical condition. Further studies will have to show if the antigens described here are of predictive value.

ACKNOWLEDGEMENTS

The authors thank A. Walduck for critical reading of the article.

REFERENCES

1. Ernst PB, Gold BD. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. Annu Rev Microbiol 2000;54:615–40.
2. Labenz J, Meining A, Tillenburg B, Stolte M. Helicobacteriose: update 1999. Leber Magen Darm 1999;29:80–92.
3. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK. *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med 1991;325:1127–31.
4. Graham DY. *Helicobacter pylori* infection in the pathogenesis of duodenal ulcer and gastric cancer: a model. Gastroenterology 1997; 113:1983–91.
5. McColl KE and el Omar E. *Helicobacter pylori* and disturbance of gastric function associated with duodenal ulcer disease and gastric cancer. Scand J Gastroenterol Suppl 1996;215:32–7.
6. Farthing MJ, Fitzgerald R, Zhang ZW. Acid, helicobacter and immunity: a new paradigm for oesophagogastric cancer. J Physiol Paris 2001;95:423–7.
7. Hansson LE, Nyren O, Hsing AW, Bergstrom R, Josefsson S, Chow WH, Fraumeni JF, Jr, Adami HO. The risk of stomach cancer in patients with gastric or duodenal ulcer disease. N Engl J Med 1996; 335:242–9.
8. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, Schlemper RJ. *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med 2001;345:784–9.
9. Calam J and Baron JH. ABC of the upper gastrointestinal tract: pathophysiology of duodenal and gastric ulcer and gastric cancer. BMJ 2001;323:980–2.
10. Jungblut PR, Bumann D, Haas G, Zimny-Arndt U, Holland P, Lamer S, Siejak F, Aebischer A, Meyer TF. Comparative proteome analysis of *Helicobacter pylori*. Mol Microbiol 2000;36:710–25.
11. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 1999;397:176–80.
12. Bumann D, Meyer TF, Jungblut PR. Proteome analysis of the common human pathogen *Helicobacter pylori*. Proteomics 2001;1:473–9.
13. Ogura K, Maeda S, Nakao M, Watanabe T, Tada M, Kyutoku T, Yoshida H, Shiratori Y, Omata M. Virulence factors of *Helicobacter pylori* responsible for gastric diseases in Mongolian gerbil. J Exp Med 2000;192:1601–10.
14. Zevering Y, Jacob L, Meyer TF. Naturally acquired human immune responses against *Helicobacter pylori* and implications for vaccine development. Gut 1999;45:465–74.
15. Graham DY and Yamaoka Y. Disease-specific *Helicobacter pylori* virulence factors: the unfulfilled promise. Helicobacter 2000;5:S3–9.
16. Iaquinto G, Todisco A, Giardullo N, D'Onofrio V, Pasquale L, De Luca A, Andriulli A, Perri F, Rega C, De Chiara G, Landi M, Taccone W, et al. Antibody response to *Helicobacter pylori* CagA and heat-shock proteins in determining the risk of gastric cancer development. Dig Liver Dis 2000;32:378–83.
17. Haas G, Karaali G, Ebermayer K, Metzger WG, Lamer S, Zimny-Arndt U, Diescher S, Goebel UB, Vogt K, Roznowski AB, Wiedenmann BJ, Meyer TF, et al. Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease. Proteomics 2002;2:313–24.
18. Jungblut PR. Proteome analysis of bacterial pathogens. Microbes Infect 2001;3:831–40.
19. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 1997;388:539–47.
20. Miehlke S, Kirsch C, Agha-Amiri K, Gunther T, Lehn N, Malfertheiner P, Stolte M, Ehninger G, Bayerdorffer E. The *Helicobacter pylori* vacA s1, m1 genotype and cagA is associated with gastric carcinoma in Germany. Int J Cancer 2000;87:322–7.
21. Miehlke S, Hackelsberger A, Meining A, Hatz R, Lehn N, Malfertheiner P, Stolte M, Bayerdorffer E. Severe expression of corpus gastritis is characteristic in gastric cancer patients infected with *Helicobacter pylori*. Br J Cancer 1998;78:263–6.
22. Miehlke S, Meining A, Lehn N, Hochter W, Weingart J, Simon T, Kramer W, Klann H, Bolle K, Sommer A, Stolte M, Bayerdorffer E. Comparison of omeprazole, metronidazole and clarithromycin with omeprazole/amoxicillin dual-therapy for the cure of *Helicobacter pylori* infection. Digestion 1998;59:646–50.
23. Bayerdorffer E, Miehlke S, Mannes GA, Sommer A, Hochter W, Weingart J, Heldwein W, Klann H, Simon T, Schmitt W, Bastlein E, Eimiller A, et al. Double-blind trial of omeprazole and amoxicillin to cure *Helicobacter pylori* infection in patients with duodenal ulcers. Gastroenterology 1995;108:1412–7.
24. Klose J, Kobalz U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. Electrophoresis 1995;16:1034–59.
25. Doherty NS, Littman BH, Reilly K, Swindell AC, Buss JM, Anderson NL. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis 1998;19:355–63.
26. Jungblut P, Eckerskorn C, Lottspeich F, Klose J. Blotting efficiency investigated by using two-dimensional electrophoresis, hydrophobic membranes and proteins from different sources. Electrophoresis 1990; 11:581–8.
27. Pun T, Hochstrasser DF, Appel RD, Funk M, Villars-Augsburger V, Pellegrini C. Computerized classification of two-dimensional gel electrophoretograms by correspondence analysis and ascendant hierarchical clustering. Appl Theor Electrophor 1988;1:3–9.
28. Schmid HR, Schmitter D, Blum P, Miller M, Vonderschmitt D. Lung tumor cells: a multivariate approach to cell classification using two-dimensional protein pattern. Electrophoresis 1995;16:1961–8.
29. Pleissner KP, Regitz-Zagrosek V, Krudewagen B, Trenkner J, Hocher B, Fleck E. Effects of renovascular hypertension on myocardial protein patterns: analysis by computer-assisted two-dimensional gel electrophoresis. Electrophoresis 1998;19:2043–50.
30. Schmidt F, Schmid M, Mattow J, Facius A, Pleissner KP, Jungblut PR. Iterative data analysis is the key for exhaustive analysis of peptide mass fingerprints from proteins separated by two-dimensional electrophoresis. J Am Soc Mass Spectrom 2003;14:943–56.
31. Kimmel B, Bosserhoff A, Frank R, Gross R, Goebel W, Beier D. Identification of immunodominant antigens from *Helicobacter pylori* and evaluation of their reactivities with sera from patients with different gastroduodenal pathologies. Infect Immun 2000;68:915–20.
32. Yamaguchi H, Osaki T, Kai M, Taguchi H, Kamiya S. Immune response against a cross-reactive epitope on the heat shock protein 60 homologue of *Helicobacter pylori*. Infect Immun 2000;68:3448–54.
33. Gonzalez CA, Sala N, Capella G. Genetic susceptibility and gastric cancer risk. Int J Cancer 2002;100:249–60.