PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/205, 16/12, G01N 33/53, A61K 31/70, 39/106, 39/395

A2

(11) International Publication Number: **WO 98/04702**

(43) International Publication Date: 5 February 1998 (05.02.98)

(21) International Application Number: PCT/IB97/00981

(22) International Filing Date: 25 July 1997 (25.07.97)

(30) Priority Data:
196 30 390.7 26 July 1996 (26.07.96) DE

(71) Applicant *(for all designated States except US)*: CHIRON BEHRING GMBH & CO. [DE/DE]; P.O. Box 16 30, D-35006 Marburg (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNAPP,. Bernhard [DE/DE]; In den Guchsgärten 1, D-35083 Wetter (DE). HUNDT, Erika [DE/DE]; Zum Hirtzborn 8, D-35041 Wehrshausen (DE). SCHMIDT, Karl-Heinz [DE/DE]; Zur Mühle 5, D-35096 Allna (DE).

(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PROTEINS, IN PARTICULAR MEMBRANE PROTEINS, OF *HELICOBACTER PYLORI*, THEIR PREPARATION AND USE

(57) Abstract

The present invention relates to novel proteins, in particular membrane proteins or proteins which are firmly associated with the membrane, which are derived from *Helicobacter pylori* (*H. pylori*) and which contain one of the peptide sequences selected from SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 or 19 according to Tables 1a-1c, or to parts or homologues thereof having a minimum length of five amino acids, and to their preparation and use as pharmaceutical compositions, in particular as vaccines, or as a diagnostic agent. Based on these data, genes coding for these and related proteins were also isolated as shown in SEQ ID NOS: 20, 21, 22, 23, 24, 25, 26 and 27.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Proteins, in particular membrane proteins, of *Helicobacter pylori*, their preparation and use

TECHNICAL FIELD OF THE INVENTION

The present invention relates to novel proteins, in particular membrane proteins or proteins which are firmly associated with the membrane, which are derived from *Helicobacter pylori* (*H. pylori*) and which contain one of the peptide sequences selected from SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 or 19 according to Tables 1a-1c, or to parts or homologues thereof having a minimum length of five amino acids, and to their preparation and use as pharmaceutical compositions, in particular as vaccines, or as a diagnostic agent. Based on these data, genes coding for these and related proteins were also isolated as shown in SEQ ID NOS: 20, 21, 22, 23, 24, 25, 26 and 27.

BACKGROUND OF THE INVENTION

Helicobacter pylori is a Gram-negative, microaerophilic, spiral bacterium which colonizes the mucosa of the human stomach. The bacterium is the cause of chronic active gastritis and of peptic ulcer, in particular duodenal ulcer, and plays a role in the development of carcinomas of the stomach; consequently, *Helicobacter pylori* is an important human pathogen.

Its helical shape and motility, due to from four to six flagellae, enables the bacterium to migrate through the gastric mucus in order to reach the boundary layer, which is virtually at neutral pH, between the mucus and the mucosa. Ammonium ions, which are produced

during the enzymic cleavage of urea by bacterial urease, protect the pathogen from the aggressive gastric acid. The bacterium adheres to the endothelial cells of the stomach using specific adhesins.

A consequence of chronic colonization of the mucosa can be an inflammatory granulocytic, and subsequently monocytic, infiltration of the epithelium which in turn, by way of inflammation mediators, contributes to the tissue destruction. Infection stimulates both a local and a systemic humoral immune response, without these responses being able to eliminate the pathogen effectively. Immunization is the conventional way of preventing infectious diseases. It is therefore important to examine this option with regard to controlling an *H. pylori* infection.

The development of a vaccine involves identifying factors which are crucial for virulence or structures which are accessible to the human immune system for the purpose of eliminating a pathogen. It is to be assumed that antigens of this nature are present in the outer membrane of the bacterium. Thus, adhesins of 19,600 Da (P. Doig et al., 1992, J. of Bacteriology 174, 2539-2547), 20,000 Da (D.G. Evans et al., 1993, J. of Bacteriology 175, 674-683) and 63,000 Da (C. Lingwood et al., 1993, Infection and Immunity 61, 2474-2478) are located in the outer membrane, which adhesins are candidates for an experimental vaccine which has the aim of inducing antibodies which prevent adhesion of the bacterium to the mucosal surface.

In addition, the outer membrane possesses porins of 30,000 Da (M.A. Tufano et al., 1994, Infection and Immunity 62, 1392-1399), 48,000 Da, 49,000 Da, 50,000 Da, 67,000 Da (M.M. Exner et al., 1995, Infection and Immunity 63, 1567-1572) and 31,000 Da (P. Doig et al., 1995, J. of Bacteriology 177, 5447-5452) molecular weight, and also iron-regulated outer membrane proteins of 77,000 Da, 50,000 Da and 48,000 Da (D.J. Worst et al.,

1995, Infection and Immunity 63, 4161-4165) molecular weight, erythrocyte-binding antigens of 59,000 Da and 25,000 Da (J. Huang et al., 1992, J. of Gen. Microbiol. 138, 1503-1513) molecular weight and proteins for binding laminin, collagen I and IV, fibronectin and vitronectin (I. Kondo et al., 1993, European J. Gastroenterol. Hepatol. 5, 63-67). In addition, proteins of 19,000 Da (E.B. Drouet et al., 1991, J. of Clinical Microbiology 29, 1620-1624), 50,000 Da (M.M. Exner et al., 1995, Infection and Immunity 63, 1567-1572) and 30,000 Da (J. Bölin et al., 1995, J. of Clinical Microbiology 33, 381-384) molecular weight, and also a 20,000 Da lipoprotein (M. Kostrzynska et al., 1994, J. of Bacteriology 176, 5938-5948) and strain-specific, surface-located antigens of 51,000 Da, 60,000 Da and 80,000 Da (P. Doig and T.J. Trust, 1994, Infection and Immunity 62, 4526-4533) have been described. The genes for the proteins of 20,000 Da (HpaA) (Evans et al.) and 20,000 Da (lpp20) (M. Kostrzynska et al.) molecular weight have now been isolated. N-terminal protein sequence data have been disclosed for the adhesins of 19,600 Da (P. Doig et al., 1992) and 63,000 Da (C. Lingwood et al.) molecular weight, for the porins of 48,000 Da, 49,000 Da, 50,000 Da, 67,000 Da (M.M. Exner et al.), 30,000 Da (M.A. Tufano, 1994) and 31,000 Da (P. Doig et al., 1995) molecular weight and for the 50,000 Da protein (M.M. Exner et al., 1995).

SUMMARY OF THE INVENTION

According to a first aspect of the present invention there is provided a protein from *Helicobacter pylori* (*H. pylori*) containing one of the peptide sequences selected from SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 and 19 according to Tables 1a-1c, or parts or homologues thereof having a minimum length of

five amino acids. Preferably the peptide sequences of the protein are N-terminal sequences.

The protein according to the first aspect of the present invention preferably contains a peptide sequence having the SEQ ID NO: 1 according to Table 1a and has a molecular weight of approx. 250 kD, or preferably contains a peptide sequence having the SEQ ID NO: 2 according to Table 1a and has a molecular weight of approx. 110 kD, or preferably contains a peptide sequence having the SEQ ID NO: 3 according to Table 1a and has a molecular weight of approx. 100 kD, or preferably contains a peptide sequence having the SEQ ID NO: 6 according to Table 1a and has a molecular weight of approx. 60 kD, or preferably contains a peptide sequence having the SEQ ID NO: 10 according to Table 1b and has a molecular weight of approx. 42 kD, or preferably contains a peptide sequence having the SEQ ID NO: 11 according to Table 1b and has a molecular weight of approx. 42 kD, or preferably contains a peptide sequence having the SEQ ID NO: 12 according to Table 1b and has a molecular weight of from approx. 32 to approx. 36 kD, or preferably contains a peptide sequence having the SEQ ID NO: 14 according to Table 1c and has a molecular weight of approx. 30 kD, or preferably contains a peptide sequence having the SEQ ID NO: 15 according to Table 1c and has a molecular weight of approx. 28 kD, or preferably contains a peptide sequence having the SEQ ID NO: 16 according to Table 1c and has a molecular weight of approx. 28 kD, or preferably contains a peptide sequence having the SEQ ID NO: 17 according to Table 1c and has a molecular weight of approx. 25 kD, or preferably contains a peptide sequence having the SEQ ID NO: 18 according to Table 1c and has a molecular weight of approx. 25 kD, or preferably contains a peptide sequence having the SEQ ID NO: 19 according to Table 1c and has a molecular weight of approx. 17 kD.

The protein according to the first aspect of the present invention is preferably a membrane protein or a protein which is firmly associated with the membrane. More preferably said protein is an integral membrane protein, in particular a Sarkosyl®-insoluble integral membrane protein.

In a second aspect of the invention there are provided proteins according to the first aspect of the present invention, which can be obtained in accordance with the following procedural steps:
(a) isolating the proteins by means of differential solubilization;
(b) separating the proteins, which have been isolated in accordance with step (a), by means of gel electrophoretic methods; and
(c) isolating the proteins, which have been separated in accordance with step (b).

Preferably the proteins according to the second aspect of the present invention can be obtained by means of differential solubilization using Sarkosyl®. The proteins can also be obtained by means of separation by one or more SDS polyacrylamide gel electrophoreses, preferably by means of several SDS polyacrylamide gel electrophoreses having different polyacrylamide contents, more preferably wherein the polyacrylamide content of said gel electrophoreses is approximately 8%, 10% or 16%.

In a third aspect of the present invention there is provided a peptide having the amino acid sequence according to SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 or 19 according to Tables 1a-1c, or parts or homologues thereof having a minimum length of five amino acids.

In a fourth aspect of the present invention there

is provided an antibody against one or more proteins according to the first or second aspects of the present invention and/or against one or more peptides according to the third aspect of the present invention.

In a fifth aspect of the present invention there is provided a polynucleotide encoding one or more proteins according to the first or second aspects of the present invention or one or more peptides according to the third aspect of the present invention.

In a sixth aspect of the present invention there is provided a process for preparing the proteins according to the first or second aspects of the present invention, characterized in that the following procedural steps are carried out:

(a) isolating the proteins, by means of differential solubilization;

(b) separating the proteins, which have been isolated in accordance with step (a), by means of gel electrophoretic methods; and

(c) isolating the proteins, which have been separated in accordance with step (b).

Preferably the process is characterized in that the proteins are isolated in accordance with step (a) using Sarkosyl®.

In a seventh aspect of the present invention there is provided a process for preparing the peptides according to the third aspect of the present invention, characterized in that a chemical peptide synthesis is carried out.

In an eighth aspect of the present invention there is provided a process for preparing the proteins according to the first or second aspects of the present

invention or the peptides according to the third aspect of the present invention, characterized in that a polynucleotide according to the fifth aspect of the present invention is expressed.

In a ninth aspect of the present invention there is provided the use of one or more proteins according to the first or second aspects of the present invention, one or more peptides according to the third aspect of the present invention, one or more antibodies according to the fourth aspect of the present invention or one or more polynucleotides according to the fifth aspect of the present invention for preparing a pharmaceutical composition or a diagnostic agent.

In a tenth aspect of the present invention there is provided a pharmaceutical composition comprising one or more proteins according to the first or second aspects of the present invention and/or one or more peptides according to the third aspect of the present invention or one or more antibodies according to the fourth aspect of the present invention or one or more polynucleotides according to the fifth aspect of the present invention or their expression products. Preferably said pharmaceutical composition is used as a vaccine.

In an eleventh aspect of the present invention there is provided a diagnostic agent comprising one or more proteins according to the first or second aspects of the present invention and/or one or more peptides according to the third aspect of the present invention or one or more antibodies according to the fourth aspect of the present invention or one or more polynucleotides according to the fifth aspect of the present invention or their expression products.

In a twelfth aspect of the present invention

there is provided a protein from *H. pylori* containing one of the peptide sequences deduced from SEQ ID NO: 21, 22, 23, 24, 25, 26 and 27, or parts or homologues thereof having a minimum length of five amino acids.

In a thirteenth aspect of the present invention there is provided a peptide having the amino acid sequence deduced from SEQ ID NO: 21, 22, 23, 24, 25, 26 or 27, or parts or homologues thereof having a minimum length of five amino acids.

In a fourteenth aspect of the present invention there is provided a peptide selected from the C-terminal region of the peptide sequence of SEQ ID NO: 20 or homologue thereof. Preferably said peptide is selected from RDPKFNLAHIEKEFEVWNWDYRA and EKHQKMMKDMHGKDMHHTKKKK, or parts or homologues thereof.

In a fifteenth aspect of the present invention there is provided an antibody against one or more proteins according to the twelfth aspect of the present invention and/or against one or more peptides according to the thirteenth or fourteenth aspects of the present invention.

In a sixteenth aspect of the present invention there is provided a polynucleotide encoding one or more proteins according to the twelfth aspect of the present invention or one or more peptides according to the thirteenth or fourteenth aspects of the present invention.

In a seventeenth aspect of the present invention there is provided a host cell transformed with the polynucleotide according to the fifth or sixteenth aspects of the present invention.

In an eighteenth aspect of the present invention there is provided an expression product expressed from the host cell according to the seventeenth aspect of the present invention.

In a nineteenth aspect of the present invention there is provided a pharmaceutical composition comprising one or more proteins according to the twelfth aspect of the present invention and/or one or more peptides according to the thirteenth or fourteenth aspects of the present invention or one or more antibodies according to the fifteenth aspect of the present invention or one or more polynucleotides according to the sixteenth aspect of the present invention or their expression products. Preferably said pharmaceutical composition is used as a vaccine. More preferably, when the pharmaceutical composition comprises a nucleotide sequence, said pharmaceutical composition is used as a DNA vaccine.

In a twentieth aspect of the present invention there is provided a diagnostic agent comprising one or more proteins according to the twelfth aspect of the present invention and/or one or more peptides according to the thirteenth or fourteenth aspects of the present invention or one or more antibodies according to the fifteenth aspect of the present invention or one or more polynucleotides according to the sixteenth aspect of the present invention or their expression products.

In a twenty-first aspect of the present invention there is provided the use of one or more proteins according to the twelfth aspect of the present invention or one or more peptides according to the thirteenth or fourteenth aspects of the present invention or one or more antibodies according to the fifteenth aspect of the present invention or one or more polynucleotides

according to the sixteenth aspect of the present invention or their expression products for preparing a pharmaceutical composition or a diagnostic agent.

DETAILED DESCRIPTION OF THE INVENTION AND BEST MODE

The present application describes the isolation and determination of, in all, 19 proteins, in particular membrane proteins or proteins which are firmly associated with the membrane, especially integral membrane proteins, which proteins are in a molecular weight range of from 17 kD to approx. 250 kD (Tables 1a-1c). The term membrane protein is generally understood to mean integral and peripheral membrane proteins and transmembrane proteins. Integral membrane proteins are proteins which are partially or entirely inserted into the cytoplasmic membrane. By contrast, peripheral membrane proteins only adhere to the surface of the membrane. Transmembrane proteins pass completely through the membrane (see, for example, B. Alberts et al. (eds), Membrane Proteins in "Molecular Biology of the Cell", 2nd ed., Garland Publishing, Inc., New York & London, 284-287, 1989). Two sequences were identified in one band in seven cases (SEQ ID NO: 2 and 3, 5 and 6, 7 and 8, 10 and 11, 13 and 14, 15 and 16, and 17 and 18), while it was only possible to identify one sequence in one band in a further five cases (SEQ ID NO: 1, 4, 9, 12 and 19). Six N-terminal sequences from the 19 peptide sequences identified had already been described in earlier studies; these were the sequences for urease A and urease B (B.E. Dunn et al., 1990, J. Biolog. Chem. 265, 9464-9469), for the exoenzyme S-like protein (C. Lingwood et al.), for the 50 kD membrane protein and for the porins hop B and hop C (M.M. Exner et al.). The only genes for these antigens which have so far been isolated are those for urease A and urease B (A. Labigne et al., 1991, J. Bacteriol. 173, 1920-1931). It

was not possible to find the N-terminal sequences, which have already been described, of the membrane proteins of 19,600 Da (P. Doig et al., 1992), 48,000 Da, 67,000 Da (M.M. Exner et al., 1995) and 31,000 Da (P. Doig et al., 1995) molecular weight among the 19 sequences which are described in accordance with the invention. Thus, the protein which is described by SEQ ID NO: 14 cannot be attributed, either, to the protein having the molecular weight of 31,000 Da (P. Doig et al., 1995). The remaining 13 amino terminal protein sequences of the 19 amino terminal protein sequences according to Tables 1a-1c have not been described. It is to be assumed that these sequences can be attributed to *Helicobacter pylori* proteins which have not previously been identified.

It was surprising, therefore, that it was possible to demonstrate a large number of additional, novel *H. pylori* proteins in a Sarkosyl®-insoluble fraction. The proteins are very probably integral proteins of the outer membrane or proteins which are firmly associated with the membrane. They are therefore particularly suitable for use as candidates for developing a vaccine or a diagnostic agent.

The invention describes proteins, in particular membrane proteins or proteins which are firmly associated with the membrane, especially integral membrane proteins, in particular Sarkosyl®-insoluble integral membrane proteins of *H. pylori*, which contain one of the peptide sequences selected from SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 17, 18 or 19 according to Tables 1a-1c, or to parts or homologues thereof having a minimum length of five, preferably six amino acids, with these peptide sequences preferably constituting N-terminal sequences of the said proteins. The novel peptides are particularly preferred which exhibit at least ten consecutive amino acids selected from the sequences having the SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16 and 19. In addition,

those said parts are in particular preferred which contain an uninterrupted sequence of unambiguously specified amino acids.

The term "part" in the context of "part(s) of a sequence" in the present invention is defined herein as meaning a sequence of amino acids which can form a T-cell or B-cell epitope. Such an amino acid sequence is usually of a minimum of approximately four to eight amino acids.

The term "homologue(s)" in the context of the present invention is defined herein as meaning the same protein or peptide of a different strain of *H. pylori* but exhibiting the same function. Thus, although the actual amino acid sequences may not be identical between homologous proteins or peptides from different strains of *H. pylori*, the differences between the amino acid sequences merely represent strain-specific differences; the function of the homologues is identical.

In a particular embodiment, the protein containing a peptide sequence having the SEQ ID NO: 1 according to Table 1a has a molecular weight of approx. 250 kD, the protein containing a peptide sequence having the SEQ ID NO: 2 according to Table 1a has a molecular weight of approx. 110 kD, the protein containing a peptide sequence having the SEQ ID NO: 3 according to Table 1a has a molecular weight of approx. 100 kD, the protein containing a peptide sequence having the SEQ ID NO: 6 according to Table 1a has a molecular weight of approx. 60 kD, the protein containing a peptide sequence having the SEQ ID NO: 10 according to Table 1b has a molecular weight of approx. 42 kD, the protein containing a peptide sequence having the SEQ ID NO: 11 according to Table 1b has a molecular weight of approx. 42 kD, the protein containing a peptide sequence having the SEQ ID NO: 12 according to Table 1b has a molecular weight of from approx. 32 to approx. 36 kD, the protein containing a peptide sequence having the SEQ ID NO: 14 according to Table 1c has a molecular weight of approx. 30 kD, the

protein containing a peptide sequence having the SEQ ID NO: 15 according to Table 1c has a molecular weight of approx. 28 kD, the protein containing a peptide sequence having the SEQ ID NO: 16 according to Table 1c has a molecular weight of approx. 28 kD, the protein containing a peptide sequence having the SEQ ID NO: 17 according to Table 1c has a molecular weight of approx. 25 kD, the protein containing a peptide sequence having the SEQ ID NO: 18 according to Table 1c has a molecular weight of approx. 25 kD, and the protein containing a peptide sequence having the SEQ ID NO: 19 according to Table 1c has a molecular weight of approx. 17 kD.

The generally available *H. pylori* strain No. ATCC 43504 is used, for example, as the starting material when isolating the proteins, with it being possible, in particular, to carry out the following procedural steps:

(a) isolating the proteins by means of differential solubilization, in particular using Sarkosyl® (an N-lauroylsarcosine) in accordance with the method of Blaser et al. (1983, Infect. Immun. 42, 276-284),

(b) separating the proteins, which have been isolated in accordance with step (a), by means of gel electrophoretic methods, preferably by means of SDS polyacrylamide gel electrophoresis, with use being made, in particular, of polyacrylamide gels having differing polyacrylamide contents, in particular containing approx. 8, 10 or 16% polyacrylamide, and

(c) isolating the proteins, which have been separated in accordance with step (b), by means of known methods, for example by elution or by isolation on a membrane.

For the purpose of isolating and characterizing the proteins according to the present invention, the proteins were first of all obtained using the method of Blaser et al. (see above). The bacteria, which had been disrupted in a glass bead homogenizer, were freed of intact bacteria by centrifugation at 5000 g; the

supernatant was then centrifuged at 100,000 g. The pellet was dissolved in Sarkosyl®, and the Sarkosyl®-insoluble fraction, which contains the integral membrane proteins in particular, was centrifuged off. The pellet was resuspended in distilled water and fractionated by SDS polyacrylamide gel electrophoresis (PAGE). In this connection, it was found that SDS-PAGE, in contrast to HPLC, was a very effective method for separating Sarkosyl®-insoluble proteins. For this, the gels were pretreated with methionine in order to prevent oxidation of the methionine residues. After the run, the proteins were transferred from the SDS gel to a PVDF membrane (Immobilon P®, from Millipore), with 0.005% SDS being added to the cathode buffer in order to complete the transfer of the very insoluble proteins. For sequence analysis, the protein bands from four tracks, in each case, were cut out of the PVDF membrane and Edman amino acid degradation was carried out in a 477A fluid-phase sequencer (Applied Biosystems, Inc. (ABI)) to determine the amino acid sequence. While it is possible further to fractionate the proteins which run in one band, for example by means of isoelectric focusing or two-dimensional gel electrophoresis, this is not necessary for an unambiguous sequence analysis since the sequences can be assigned unambiguously on the basis of the different protein contents of the proteins which run in one band.

The amino acids which are labelled Xaa in the sequence listing can be explained as follows:

The non-identifiable amino acids can be caused by interference due to impurities in the first sequencing step, a non-analysable amino acid, such as Cys or Trp, a modifiable amino acid which is missing in the elution programme, or an amino acid, such as Ser or Thr, which is difficult to determine, basically due to low sequence yields. Different bands can also contain two proteins of very similar molecular weights in different quantities.

This then results in two sequences which then also have to be assigned unambiguously on account of the different frequency of the individual amino acids.

The present invention also describes the peptides which are designated by the sequences according to SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 16, 17, 18 or 19 according to Tables 1a-1c, or to parts or homologues thereof having a minimum length of five amino acids, in particular of six amino acids, which can be prepared, for example, by well-known chemical peptide synthesis (Barani, G. & Merrifield, R. B. in "The Peptides: Analysis, Synthesis and Biology" (Gross E., ed.), Vol. 2, Academic Press, 1980, Johannes Meyenhofer Verlag; Bodanszky, M. & Bodanszky, A. "The practice of peptide synthesis", Springer Verlag, 1984). The novel peptides are particularly preferred which possess at least ten consecutive amino acids selected from the sequences having the SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16 and 19. Furthermore, those said peptides are, in particular, preferred which contain an uninterrupted sequence of unambiguously determined amino acids, as is the case with the sequences from SEQ ID NO: 12, 14 and 15.

The present application also describes antibodies which can also be prepared by methods which are well known to the skilled person (see, for example, B.A. Diamond et al. (1981), The New England Journal of Medicine, 1344-1349) and which are directed against one or more of the novel proteins or peptides.

The skilled person is also familiar, from J. Sambrook et al. (1989, "Molecular Cloning, A Laboratory Manual", 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), with methods for preparing polynucleotides which encode the novel proteins or peptides. In particular, the skilled person knows, on the basis of the genetic code, the nucleotide sequences which encode the peptides according to the sequence

listing. In particular, the nucleotide sequences are preferred which occur most frequently in accordance with the rules for the frequency of use of the different codons in *Helicobacter pylori*. These nucleotide sequences can be prepared, for example, by means of chemical polynucleotide synthesis (see, for example, E. Uhlmann & A. Peyman (1990), Chemical Reviews, 543-584, Vol. 90, No. 4).

For example, oligodeoxynucleotides which have been prepared in accordance with these rules can be employed for screening *Helicobacter pylori* gene libraries using known methods (J. Sambrook et al., 1989, "Molecular Cloning, A Laboratory Manual", 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). Furthermore, taking the sequence data as a basis, peptides can be synthesized which are employed for obtaining antisera. Gene expression libraries can then be screened using these antisera. The clones resulting from these different screening methods can then be employed, by isolating and sequencing the inserted DNA fragments, for identifying DNA sequence segments which encode the N-terminally sequenced protein segments of the proteins. If the inserted DNA fragments do not contain the complete gene encoding any particular protein, these DNA fragments can be used to isolate the complete genes by screening other gene libraries. The genes which have been completely isolated in this manner can then be expressed, in accordance with the state of the art, in various well-known systems in order to obtain the corresponding protein.

Using oligonucleotides deduced from the N-terminal sequences of SEQ ID NOS: 5, 7, 8, 10, 12 and 15, the genes corresponding to the SEQ ID NOS: 5, 8, 10, 12 and 15 were isolated and are specified as SEQ ID NOS: 20 (catalase), 24 (50 kD membrane protein), 25 (42 kD protein), 26 (36/35/32 kD protein) and 23 (28 kD protein). The gene coding for Hop C could not be isolated

using oligonucleotide 7. However, oligonucleotide 7 hybridizes with an homologous gene specified as SEQ ID NO: 21 (Hop X). Two additional genes which belong to this family were able to be isolated and are specified as SEQ ID NO: 21 (Hop Y) and SEQ ID NO: 22 (Hop Z).

Another approach is given by the recent access to the complete genomic sequence of *H. pylori* on the internet which allowed, for example, the identification of SEQ ID NO: 27.

The novel proteins, peptides, antibodies and polynucleotides, and their expression products, can now be used, in accordance with methods known to the skilled person, for preparing a pharmaceutical composition , in particular a vaccine, or a diagnostic agent.

Those regions of the proteins which, on the one hand, occur, if possible, in all *H. pylori* strains, and, on the other hand, bring about the formation of protective antibodies, are particularly suitable for preparing vaccines. A special preference is given to the regions which project from the surface of the bacteria.

Such vaccines may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). These vaccines comprise antigen or antigens, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as

a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) those formulations described in PCT Publ. No. WO 90/14837, including but not limited to MF59 (containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA)), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, such as StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freunds Adjuvant (CFA) and Incomplete Freunds Adjuvant (IFA); (5) cytokines, such as interleukins (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g., gamma interferon), macrophage colony stimulating factor (M-CSF), tumour necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59 are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-l-alanyl-d-isoglutamine (nor-MDP), N-acetylmuramyl-l-alanyl-d-isoglutaminyl-l-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-huydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (e.g., the antigen, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g., nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant

factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

The present invention describes, therefore, pharmaceutical compositions, in particular vaccines, and diagnostic agents which comprise one or more of the novel proteins and/or one or more of the novel peptides or one or more of the novel antibodies or one or more of the novel polynucleotides or one or more expression products of the novel polynucleotides.

For example, according to the present invention, a DNA vaccine can be prepared on the basis of the polynucleotides, or a diagnostic agent can be prepared on the basis of the polymerase chain reaction (PCR diagnosis), or an immunotest, for example a Western blot test or an enzyme immunotest (ELISA) can be prepared on the basis of the antibodies. Furthermore, the novel proteins or peptides, or their immunogenic moieties, in particular when they contain an uninterrupted sequence of unambiguously determined amino acids, having a minimum length of five amino acids, preferably six amino acids and, in particular, in the case of the novel peptides having the SEQ ID NOS: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16 and 19 and peptides or proteins encoded by the DNA sequences of SEQ ID NOS: 20, 21, 22, 23, 24, 25, 26 and 27, at least ten consecutive amino acids, can be used as antigens for immunizing mammals. In this context, the two C-terminal regions C1 and C2 specific for *H. pylori*

catalase (c.f. Example 6) can also be used as immunogens. The antibodies which are formed by the immunization, or antibodies which are prepared by means of recombinant DNA methods (see, for example, Winter G. & Milstein C. (1991) Nature, 293-299, Vol. 349), can, inter alia, prevent adhesion of the bacteria to the mucosal surface, attract macrophages for the purpose of eliminating bacteria, and activate the complement system for the purpose of lysing the bacteria.

The following examples are intended to clarify the invention.

EXAMPLES

Example 1:

Culture of *Helicobacter pylori*

The *H. pylori* stain ATCC 43504 was passaged under microaerophilic conditions (BBL Jar/Campy Pak Plus, from Becton & Dickinson) on Columbia Agar plates containing 5% horse blood (incubation 48 h, 37°C). Three plates were rinsed off when inoculating a 500 ml flow-spoiler flask (100 ml of Columbia broth, 7% FCS); during the incubation (BBL Jar/Campy Pak Plus; 48 h, 37°C, 90 rpm), the OD_{590} rose from 0.3 to 2.0. The bacteria were harvested by centrifugation at 10,000 rpm and washed twice with physiological sodium chloride solution.

Example 2:

Isolation of *Helicobacter pylori* outer membrane proteins

The preparation of the outer membrane protein fraction, with the inner and outer membrane proteins being separated by means of differential solubilization with Sarkosyl® (Ciba-Geigy AG), was carried out using the method of Blaser et al. In this method, the bacterial cultures are harvested in the phase of late logarithmic growth, washed in 10 mM Tris buffer (pH 7.4) and disrupted with glass beads in a homogenizer (Institut für Molekularbiologie und Analytik (IMA), Germany) at 4°C and 4000 rpm for 15 min. After that, the glass beads are removed by filtration and the bacterial suspension is centrifuged at 5000 g for 20 min in order to remove intact cells. The cell walls are pelleted out of the supernatant by centrifuging at 100,000 g for 60 minutes and at 4°C. The resulting pellet is resuspended with a 1% solution of Sarkosyl® in 7 mM EDTA, and the suspension is incubated at 37°C for 20 min. The Sarkosyl®-insoluble fraction, which contains the integral membrane proteins, is pelleted by centrifugation at 50,000 g for 60 minutes and at 4°C and the pellet is resuspended in sterile distilled water; the suspension is then stored at -20°C.

Example 3:

SDS polyacrylamide gel electrophoresis and blotting

Gel preparation, and the electrophoresis, were carried out in a BioRad (Munich) Protean II xi slab cell apparatus. The chemicals employed, and the polyacrylamide monomer (as a 30% solution containing 0.8% bisacrylamide), were obtained from Oxford GlycoSystems

(Oxford, UK). In addition to a 10% standard gel, gels containing polyacrylamide contents of 8% and 16% were also especially employed for carrying out separations in the high-molecular weight and low-molecular weight ranges, respectively. The thickness of the gel was 1 mm.

In order to eliminate undesirable oxidizing properties of the ammonium persulphate used for preparing the gel, all the wells of the gel were filled with a solution containing 50 pM of L-methionine/microlitre and left to stand overnight. After the solution has been sucked off on the following day, and after each of the wells has once again been filled with 10 microlitres of this solution in each case, a preliminary electrophoresis takes place. This preliminary treatment prevents the methionine residues of the protein from being oxidized and thereby enables a protein cleavage with BrCN (Met cleavage site) to be carried out if required. The membrane protein fraction starting material is dissolved in 1.5% SDS, 2.5% mercaptoethanol, 5% glycerol and bromophenol blue in 63 mmol/l Tris buffer, pH 6.8, and fractionated by SDS polyacrylamide gel electrophoresis.

Protein transfer from the SDS gel to the PVDF membrane (Immobilon P®, from Millipore) is carried out in a BioRad (Munich) Trans Blot SD apparatus, under modified conditions.

For the purposes of completing the protein transfer, 0.005% SDS is added to the cathode buffer, thereby counteracting too rapid an impoverishment of SDS in the gel. The use of six filter papers, which are soaked with this buffer, on the cathode side is found to give optimum results in this connection.

The blot was then stained with amidoblack using the protocol of R. Westermeier (Elektrophorese Praktikum (Electrophoresis Laboratory Manual) VCH Verlag Weinheim, 1990, ISBN 3-527-28172-X).

Example 4:

N-terminal Edman degradation

The Edman amino acid degradation, and the determination of the PTH amino acids, were carried out in a 477 A liquid phase sequencer having an on-line 120A HPLC analyser (ABI).

For the analyses, the corresponding bands from, in each case, four tracks were cut out of the PVDF blot membrane and sequenced after a washing step, as recommended by ABI.

The number of sequencing steps was 5 to 25 (depending on the quantity of substance available for sequencing).

The Cys and Trp PTH amino acids cannot be detected under the conditions which were chosen.

Example 5:

Deduction of oligonucleotides for screening gene libraries and for identifying DNA fragments via Southern Blot analysis

The following oligonucleotides were deduced from the resulting N-terminal sequences of SEQ ID NOS: 5, 7, 8, 10, 12 and 15:

SEQ ID NO:	Oligonucleotide	Amino acid sequence and predicted nucleotide									
5	1	Val GTI Ala GCI	Asn AAT C Phe TTT	Lys AAA Gly GGC	Asp GAT Ala GCI	Val GTI Pro CCT	Lys AAA	Gln CAA	Thr ACT C	Xaa TGT	
7	2	Gly GGC Gly GGC	Gly GGC Gln CAA	Phe TTT Val GTG	Phe TTT Met ATG	Thr ACT C Gln CAA	Val GTG	Gly GGC	Tyr TAT	Gln CAA	Leu TTA G
8	3	(Val) GTG Asn AAT C	(Thr) ACT C Phe TTT	Tyr TAT (Ser) AGC	Glu GAA Lys AAA	Val GTG Val GT	His CAT	(Gly) GGC	Asp GAT	Phe TTT	Ile ATC T
10	4	Lys AAA	Glu GAA	Lys AAA	Phe TTT	Asn AAC	Arg AGA	Thr ACC T	Lys AAA	Pro CCT	
12	5	Glu GAA Leu TTI	Lys AAA Glu GAG	Asn AAT Val GTT	Gly GGI Gly GGI	Ala GCI Arg AGA	Phe TTT Ala GCT	Val GTG Asp GAT	Gly GGC Gln CAA	Ile ATT C Lys AAA	Ser AGC
15	6	Trp TGG Tyr TAT Lys AAA	Ser AGC Gln CAA Met ATG	Ala GCT Val GTG Val GTG	Ala GCT Ser AGC Asn AAT	Phe TTT Met ATG Asp GAT	Val GTG Ile ATT C	Gly GGC Gln CAA	Val GTG Asn AAT	Asn AAT Gln CAA	Thr ACT C

The oligonucleotides were deduced using the species-specific codon usage of *Helicobacter pylori*, which had been determined from 19 known *H. pylori* genes, and using the base inosine (I), which is capable of undergoing stable base pairing with the bases adenine (A), cytosine (C) and thymine (T) with, in each case, two hydrogen bridges. When carrying out the deduction, the degeneracy of the codon was kept as low as possible.

Example 6:

Isolation and characterization of the genes using the oligonucleotides deduced from the peptide sequences of SEQ ID NOS: 5, 7, 8, 10, 12 and 15

The oligonucleotides which had been deduced from the peptide sequences of SEQ ID NOS: 5, 7, 8, 10, 12 and 15 were labelled with digoxigenin (DIG) using a kit manufactured by Boehringer Mannheim (DIG Oligonucleotide 3'-End Labelling Kit) and employed for screening a *H. pylori* gene library which had been prepared using a kit manufactured by Stratagene (Predigested ZAP Express™ *Bam*HI/CIAP Vector Cloning Kit) at 32°C under standard conditions. Using oligonucleotides 1, 3 and 6, it was possible to identify clones which carry DNA fragments containing sequences which encode the peptide sequences of SEQ ID NOS: 5, 8 and 15. Oligonucleotide 2 hybridized with a DNA fragment which encodes an homologous sequence of SEQ ID NO: 7.

Using oligonucleotides 4 and 5, it was only possible to isolate clones whose DNA fragments did not encode SEQ ID NOS: 10 and 12. This is why these oligonucleotides and the clones which had been isolated from the λZAP Express gene library were employed in a Southern Blot analysis, which permitted the unequivocal

identification of DNA fragments which hybridized with the oligonucleotides, but not with the DNA fragments resulting from the screening. With these DNA fragments, in each case one sub-gene library was prepared in the λZAP Express vector, and each sub-gene library was screened with oligonucleotides 4 and 5. This allowed the identification of clones which carry DNA fragments encoding the sequences of SEQ ID NOS: 10 and 12.

Partial digestion of *H. pylori* DNA using the restriction enzymes *Sau*3AI, *Alu*I and *Hae*III gave a DNA which was used for establishing gene libraries in the vector λTriplex (Clontech). These gene libraries were used as starting material for isolating the complete genes of the above-described DNA fragments using standard methods.

SEQ ID NO: 20 describes the DNA sequence which encodes the catalase of *H. pylori*. The nucleotide region 337 to 378 describes the hybridization site with oligonucleotide 1. The catalase gene of *H. pylori* has been described in 1996 by Stefan Odenbreit, Björn Wieland and Rainer Haas (J. Bacteriol. 178, 6960-6967) and is therefore not new. However, when comparing the amino acid sequences of the catalases of *Escherichia coli, Bacillus firmus, B. subtilis* A, *B. subtilis* B, rats, mice, cattle, humans, *Staphylococcus violaceus, Haemophilus influenzae, B. fragilis, Pseudomonas mirabilis, B. pertussis* and *P. syringae* with the amino acid sequence of *H. pylori*, it is possible to identify two C-terminal regions C1 (RDPKFNLAHIEKEFEVWNWDYRA) and C2 (EKHQKMMKDMHGKDMHHTKKKK), which are specific to *H. pylori* catalase. These two peptides were synthesized using standard techniques, coupled to KLH and used for immunizing rabbits. These rabbits developed antibodies against the two peptides, which reacted in the Western Blot analysis with *H. pylori* catalase which had been produced by recombinant technique. These *H. pylori*-

catalase-specific regions may conceivably be used for developing a vaccine which avoids the problem complex of autoimmune reactions or for the development of a diagnostic which reacts specifically with *H. pylori* catalase.

SEQ ID NO: 21 describes a nucleotide sequence which was identified by hybridization with the oligonucleotide 2. The oligonucleotide hybridized with the sequence of nucleotide 1240 to 1284. This encodes a sequence which is homologous to the porin Hop C (Exner et al., 1995) and is identical with the published amino-terminal sequence EDDGGFFTVGYQLGQVMQDVQNPG in positions 1, 2, 3, 4, 9, 10, 11, 12, 14, 18 and 22.

The porins Hop A, Hop B, Hop C and Hop D have identical amino acids in 9 positions of the 20 N-terminal amino acids (Exner et al., 1995). In 8 of these positions, there are identical positions also in the sequence described in the present publication; in the 9th position, a conserved amino acid exchange is present (Val → Ile). It can thus be assumed that the protein described in the present publication is equally part of this group of the porins; it was therefore termed Hop X.

On the basis of the homology data and on the basis of the N-terminal sequence determined and on the basis of the hydrophobicity of the N-terminal protein sequence deduced from the nucleic acid sequence, it can be concluded that the protein deduced has a signal sequence. The mature protein with 428 amino acids has a molecular weight of 47.3 kD and an isoelectric point of 10.0.

A further open reading frame was found upstream of the gene which encodes Hop X. This further open reading frame encodes a protein which is homologous to Hop X (34% identity) and which was therefore termed Hop Y. The gene region found to date encodes the 361 C-terminal amino acids of the protein. The gene region as yet outstanding is currently being isolated using stan-

dard techniques.

We have thus identified a gene region of *H. pylori* which encodes at least two porins which are connected in series.

SEQ ID NO: 22 describes a nucleotide sequence which was concomitantly isolated and sequenced during the screening process. The amino acid sequence deduced encodes the 392 C-terminal residues of a protein which shows a high homology with Hop X (33% identity) and Hop Y (28% identity) and which was therefore termed Hop Z. The gene region which encodes the N-terminal portion of the protein is currently being isolated.

SEQ ID NO: 23 describes a DNA sequence which encodes a hitherto undescribed protein. The nucleotide region 696 to 767 describes the hybridization site with the oligonucleotide 6. On the basis of the N-terminal protein sequence which has been determined, in which it was not possible unequivocally to determine the amino acids in the first two positions, and on the basis of the hydrophobicity of the N-terminal protein sequence deduced from the nucleic acid sequence, it can be concluded that the protein deduced has a signal sequence of 17 amino acids. The mature protein of 231 amino acids has a molecular weight of 26.4 kD and an isoelectric point of 10.3. Thus, the molecular weight is quite close to the molecular weight of 28 kD which had been determined by SDS gel electrophoresis. The amino acid sequence deduced is homologous with the sequences of the proteins Hop X, Hop Y and Hop Z, for which the GCG Bestfit Programme determined identity values of 41%, 38% and 41%, respectively. The 28 kD protein thus also seems to be part of the family of the porins or porin-like proteins.

SEQ ID NO: 24 describes a DNA sequence which encodes the non-heat-modifiable 50 kD membrane protein. This protein was first described by Exner et al., 1995, and an N-terminal sequence of the protein was determined. Using the approach described by us, we were then able to

describe, with SEQ ID NO: 8, an N-terminal sequence which is identical to the sequence described by Exner et al. (1995). With the aid of the oligonucleotide 3, which had been deduced using the method illustrated in Example 5 and had been used for screening a *H. pylori* gene library using the above-described methods, it was then possible to identify a DNA fragment which encodes the 50 kD membrane protein. Using other standard methods, it was then possible to determine the nucleic acid sequence described in SEQ ID NO: 24, which encodes a mature protein of 499 amino acids which has a molecular weight of 56.3 kD and an isoelectric point of 9.75. Due to the data of the N-terminal sequencing procedures and the hydrophobicity of the N-terminal sequence, a signal sequence of 29 amino acids is assumed. The amino acid residues 236 to 254 contain a hydrophobic region which is large enough to act as a transmembrane region. Based on such data and using standard methods for epitope analysis, it is possible to identify regions which might be presented on the surface of bacteria. Such regions might be used for developing a vaccine or a diagnostic.

SEQ ID NO: 25 describes a DNA sequence 2825 bp in size which was identified by means of hybridization with oligonucleotide 4, which was deduced from SEQ ID NO: 10. Oligonucleotide 4 hybridized with the nucleotide region 897 to 923 of the described sequence of SEQ ID NO: 25. The protein has no signal sequence. The encoding region of SEQ ID NO: 25 codes for a protein of 399 amino acids with a molecular weight of 43.6 kD and an isoelectric point of 5.0. A search for homologous sequences using the BLASTP program (S. F. Altschul et al., 1990, J. Mol. Biol. 215, 403-410) identified the 42 kD antigen of *H. pylori* as the elongation factor TU. The maximum percentage of identity (89%) was found with the elongation factor TU from *Wolinella succinogenes* (W. Ludwig et al., 1993, Antonie van Leeuwenhoek 64, 285-

305).

SEQ ID NO: 26 describes a DNA sequence 2182 bp in size which hybridizes with oligonucleotide 5, which had been deduced from SEQ ID NO: 12. Oligonucleotide 5 hybridized with a *Sau*3AI fragment (position 1 to 575) of the gene library starting from position 524. The screening of different DNA libraries with specific oligonucleotides allowed the isolation of the complete gene described in SEQ ID NO: 26. An amino acid sequence which is identical to the one from SEQ ID NO: 12 can be deduced from SEQ ID NO: 26. Both protein sequencing and the hydrophobicity of the N-terminal sequence deduced allow the conclusion that the antigen has a signal sequence. The mature protein consists of 328 amino acid residues with a molecular weight of 36.1 kD and an isoelectric point of 9.95. No homologous proteins were identified using the BLASTP program (S. F. Altschul et al., 1990).

The sequences described in SEQ ID NOS: 20 to 26 indicate nucleotide sequences which encode antigens of the *H. pylori* strain ATCC 43504. However, it is known for *H. pylori* that heterogeneity between identical antigens may exist amongst various strains. We therefore claim not only the sequences described in SEQ ID NOS: 21 to 26, but in addition also the sequences of other *H. pylori* strains which are homologous with the sequences described herein.

Example 7

Identification and isolation of genes from *H. pylori* corresponding to the peptide sequences listed in Tables 1a-1c using the access to the genomic sequence

The Institute for Genomic Research (TIGR) released the DNA sequence from *H. pylori* on 24th June

1997. This new information can be accessed on the internet at "www.tigr.org". Using the TBLASTN program (Altschul et al., 1997, Nucleic Acids Research 25, in press) the peptide sequences listed in Tables 1a-1c can be aligned to amino acid sequence data deduced from all six reading frames of the *H. pylori* strain 26695. Having access to the genomic DNA sequence, DNA sequences corresponding to the aligned amino acid sequences can be identified using GCG (Genetic Computer Group) programs. This approach is shown for SEQ ID NO: 19, for example. The sequence of SEQ ID NO: 19 aligned with a very similar sequence using the TBLASTN program. SEQ ID NO: 27 describes the nucleic acid sequence and deduced amino acid sequence from the coding region of a *H. pylori* gene (strain 26695) localised between position 843212 and 843691 of the genomic sequence. The protein has no signal sequence. The N-terminal sequence of SEQ ID NO: 19 is highly homologous to the N-terminal region of the deduced amino acid sequence from amino acid residue 1 to 15. Only one different amino acid residue is present at position 4: the nucleotide sequence found by the alignment encodes a Ser residue in this position instead of an Asn residue determined by N-terminal sequencing. This can be explained by strain specific differences. The identified nucleic acid sequence in SEQ ID NO: 27 codes for a protein of 159 amino acid residues with a molecular weight of 18.2 kD and an isoelectric point of 7.2. The molecular weight is very close to that of 17 kD determined from SDS polyacrylamide gel electrophoresis. A search for homologous sequences using the BLASTP program (S. F. Altschul et al., 1990) shows that the 17 kD antigen is very homologous to "hydroxymyristol-[acyl carrier protein] dehydratase" from different bacteria.

Table 1a

N-terminal sequences of *Helicobacter pylori* membrane proteins

SEQ ID NO:	Molecular weight (kD)	Sequence	Features	Identification
1	~250	Xaa Pro Asn Gly Xaa Tyr Met Xaa Arg Xaa 5 10 Xaa Xaa Ile Xaa Xaa Xaa Gln 15	Xaa at positions 1, 5, 12, 14 and 16 are unknown amino acids. At position 8, Xaa is probably Gln, while at position 10 it is probably Ser, at position 11 it is probably Tyr and at position 15 it is probably Thr.	unknown
2	~110	Xaa Lys Leu Xaa Pro Gln Xaa Gly Tyr Val 5 10 Leu Met Tyr	At position 1, Xaa is an unknown amino acid. At position 4, Xaa is Ile or Thr and at position 7 it is Ala or Lys.	unknown
3	~100	Xaa Gln Asp Xaa Phe Leu Xaa Glu Gly Xaa 5 10 Ser	Xaa at positions 1 and 10 are unknown amino acids, and at position 4, Xaa is Ile or Thr and at position 7 it is Ala or Lys.	unknown
4	62	Xaa Lys Lys Ile Ser Arg Lys Glu Tyr Val 5 10 Ser Met Tyr Gly Pro 15	At position 1, Xaa is probably Met.	urease B
5	60	Xaa Val Asn Lys Asp Val Lys Gln Thr Xaa 5 10 Ala Phe Gly Ala Pro 15	Xaa at positions 1 and 10 are unknown amino acids.	63 kD exoenzyme-like adhesin
6	60	Xaa Phe Gln Val Xaa Phe Xaa Ile Xaa Ala 5 10 Met Asn	Xaa at positions 1, 5 and 9 are unknown amino acids, and at position 7 Xaa is Ala or Leu.	unknown
7	50	Xaa Xaa Xaa Gly Gly Phe Phe Thr Val Gly 5 10 Tyr Gln Leu Gly Gln Val Met Gln Xaa Val 15 20	At positions 2, 3 and 19, Xaa are unknown amino acids, and at position 1 Xaa is probably Glu.	Hop C

Table 1b

SEQ ID NO:	Molecular weight (kD)	Sequence	Features	Identification
8	50	Xaa Xaa Tyr Glu Val His Xaa Xaa Xaa Ile Asn Phe Xaa Lys Val	Xaa at positions 1, 2, 7 and 13 are unknown, and at position 8 Xaa is probably Asp and at position 9 it is probably Phe.	50 kD membrane protein
9	49	Xaa Xaa Asp Gly Xaa Phe Met Thr Phe Gly Tyr Glu Leu Gly Gln	Xaa at positions 1, 2 and 5 are unknown.	Hop B
10	42	Xaa Lys Glu Lys Phe Xaa Arg Thr Lys Pro Xaa Val Xaa Xaa	Xaa at positions 1 and 11 are unknown, while at position 6, Xaa is probably Asn or Gln, at position 13 it is probably Thr and at position 14 it is probably Ile.	unknown
11	42	Xaa Gly His Xaa Gln Xaa His Xaa Ala Gln	Xaa at positions 1 and 4 are unknown, while at position 6 Xaa is Asn or Gln and at position 8 it is probably Pro.	unknown
12	36/35/32	Xaa Glu Lys Asn Gly Ala Phe Val Gly Ile Ser Leu Glu Val Gly Arg Ala Asp Gln Lys Xaa	Xaa at position 1 is unknown, while at position 21 it is probably Thr.	unknown
13	31	Met Lys Leu Thr Pro Lys Glu Leu Asp Lys Leu Met Leu His Tyr Ala Gly Glu Leu Ala	----------	urease A

Table 1c

SEQ ID NO:	Molecular weight (kD)	Sequence	Features	Identification
14	30	Xaa Glu Phe Ala Gln Phe Val Gly Val Asn 5 10 Tyr Gln Xaa Asn	Xaa at positions 1 and 13 are unknown amino acids.	unknown
15	28	Xaa Xaa Ser Ala Ala Phe Val Gly Val Asn 5 10 Tyr Gln Val Ser Met Ile Gln Asn Gln Thr 15 20 Lys Met Val Asn Asp 25	Xaa at position 1 is an unknown amino acid, while at position 2 it is probably Trp.	unknown
16	28	Xaa Xaa Xaa Ile Xaa Xaa Xaa Leu Tyr Xaa 5 10 Leu Met Leu Xaa Arg 15	Xaa at positions 1, 2, 3, 6, 10 and 14 are unknown amino acids, while at position 5, Xaa is Pro or Val and at position 7 it is probably Lys.	unknown
17	25	Xaa Gln Arg Met Xaa Gln Val Gly 5	Xaa at position 1 is an unknown amino acid, while at position 5 Xaa is Pro or Lys.	unknown
18	25	Xaa Leu Asn Ile Xaa Phe Ala 5	Xaa at position 1 is an unknown amino acid, while at position 5 Xaa is Pro or Lys.	unknown
19	17	Xaa Glu Gln Asn Xaa Gln Asn Leu Gln Xaa 5 10 Xaa Phe Phe Ile Xaa 15	Xaa at positions 1, 5 and 10 are unknown amino acids, while at position 11 Xaa is probably Gln and at position 15 it is probably Lys.	unknown

SEQUENCE LISTING

(1) GENERAL INFORMATION:

APPLICANT:

(A) NAME: Chiron Behring GmbH & Co.

(B) STREET: P.O. Box 16 30

(C) CITY: Marburg

(E) COUNTRY: Germany

(F) POSTAL CODE: D-35006

TITLE OF APPLICATION:

Proteins, in particular membrane proteins, of *Helicobacter pylori*, their preparation and use

NUMBER OF SEQUENCES: 27

COMPUTER READABLE FORM:

(A) MEDIUM TYPE: floppy disk, 3½ inch, 1.44 MB

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PADAT sequence module version 1.0

(2) INFORMATION FOR SEQ ID NO: 1:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 5, 12, 14 and 16 are unknown amino acids. At position 8, Xaa is probably Gln, while at position 10 it is probably Ser, at position 11 it is probably Tyr and at position 15 it is probably Thr.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Xaa Pro Asn Gly Xaa Tyr Met Xaa Arg Xaa Xaa Xaa Ile Xaa
1 5 10

Xaa Xaa Gln
15

(2) INFORMATION FOR SEQ ID NO: 2:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
At position 1, Xaa is an unknown amino acid. At position 4, Xaa is Ile or Thr and at position 7 it is Ala or Lys.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Xaa Lys Leu Xaa Pro Gln Xaa Gly Tyr Val Leu Met Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO: 3:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1 and 10 are unknown amino acids, and at position 4, Xaa is Ile or Thr and at position 7 it is Ala or Lys.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Xaa Gln Asp Xaa Phe Leu Xaa Glu Gly Xaa Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 4:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
At position 1, Xaa is probably Met.
Identification: urease B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Xaa Lys Lys Ile Ser Arg Lys Glu Tyr Val Ser Met Tyr Gly
1 5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO: 5:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1 and 10 are unknown amino acids.
Identification: 63 kD exoenzyme-like adhesin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Xaa Val Asn Lys Asp Val Lys Gln Thr Xaa Ala Phe Gly Ala
1 5 10

Pro
15

(2) INFORMATION FOR SEQ ID NO: 6:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 5 and 9 are unknown amino acids, and at position 7 Xaa is Ala or Leu.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Xaa Phe Gln Val Xaa Phe Xaa Ile Xaa Ala Met Asn
1 5 10

(2) INFORMATION FOR SEQ ID NO: 7:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
At positions 2, 3 and 19, Xaa are unknown amino acids, and at position 1 Xaa is probably Glu.
Identification: Hop C

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Xaa Xaa Xaa Gly Gly Phe Phe Thr Val Gly Tyr Gln Leu Gly
1 5 10

Gln Val Met Gln Xaa Val
15 20

(2) INFORMATION FOR SEQ ID NO: 8:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 2, 7 and 13 are unknown, and at position 8 Xaa is probably Asp and at position 9 it is probably Phe.
Identification: 50 kD membrane protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Xaa Xaa Tyr Glu Val His Xaa Xaa Xaa Ile Asn Phe Xaa Lys
1 5 10

Val
15

(2) INFORMATION FOR SEQ ID NO: 9:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 2 and 5 are unknown.
Identification: Hop B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Xaa Xaa Asp Gly Xaa Phe Met Thr Phe Gly Tyr Glu Leu Gly
1 5 10

Gln
15

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1 and 11 are unknown, while at position 6, Xaa is probably Asn or Gln, at position 13 it is probably Thr and at position 14 it is probably Ile.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Xaa Lys Glu Lys Phe Xaa Arg Thr Lys Pro Xaa Val Xaa Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1 and 4 are unknown, while at position 6 Xaa is Asn or Gln and at position 8 it is probably Pro.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Xaa Gly His Xaa Gln Xaa His Xaa Ala Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at position 1 is unknown, while at position 21 it is probably Thr.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Xaa Glu Lys Asn Gly Ala Phe Val Gly Ile Ser Leu Glu Val
1 5 10

Gly Arg Ala Asp Gln Lys Xaa
15 20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Identification: urease A

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Lys Leu Thr Pro Lys Glu Leu Asp Lys Leu Met Leu His
1 5 10

Tyr Ala Gly Glu Leu Ala
15 20

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1 and 13 are unknown amino acids.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Xaa Glu Phe Ala Gln Phe Val Gly Val Asn Tyr Gln Xaa Asn
1 5 10

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at position 1 is an unknown amino acid and at position 2 it is probably Trp.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Xaa Xaa Ser Ala Ala Phe Val Gly Val Asn Tyr Gln Val Ser
1 5 10

Met Ile Gln Asn Gln Thr Lys Met Val Asn Asp
15 20 25

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 2, 3, 6, 10 and 14 are unknown amino acids, while at position 5, Xaa is Pro or Val and at position 7 it is probably Lys.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Xaa Xaa Xaa Ile Xaa Xaa Xaa Leu Tyr Xaa Leu Met Leu Xaa
1 5 10

Arg
15

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at position 1 is an unknown amino acid, while at position 5 Xaa is Pro or Lys.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Xaa Gln Arg Met Xaa Gln Val Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at position 1 is an unknown amino acid, while at position 5 Xaa is Pro or Lys.
Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Xaa Leu Asn Ile Xaa Phe Ala
1 5

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminus

(vi) FURTHER INFORMATION:
Xaa at positions 1, 5 and 10 are unknown amino acids, while at position 11 Xaa is probably Gln and at position 15 it is probably Lys.

Identification: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Xaa Glu Gln Asn Xaa Gln Asn Leu Gln Xaa Xaa Phe Phe Ile
1 5 10

Xaa
15

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2310 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 334 to 1851 bp protein

(vi) PROPERTIES: catalase from *Helicobacter pylori*

TTAAAAACAT CCAACAACTC TCCTTAAATT TAAAAATTCA AAAAATAAAA ATCAAAAAAA 60

AAAAAAAACA AAATCCGTCA ATGCATTGAT ATAAAATAGT ATAATAATTA TTATTAAAAC 120

CAGATTAAAA ATAAAATTTT GTCCTTAATC TTTCTTATTT TCATTAATTG TTACGAATAG 180

AAATACTTAA GGGGTTTTTT TAATTCTTAA AAAAGGATTT TTTAAGGAAA TTGAATCTTG 240

TTAGTCTTTG TATAACAAAT TATGTAATAA TCACCACAAG TGATCGGCTT AGTGTCAGAT 300

TACGAAGATT TAAGATCAAT TACAGGAAAA AAG ATG GTT AAT AAA GAT GTG AAA 354
Met Val Asn Lys Asp Val Lys
5

```
CAA ACC ACT GCT TTT GGC GCT CCC GTT TGG GAT GAC AAC AAT GTG ATT          402
Gln Thr Thr Ala Phe Gly Ala Pro Val Trp Asp Asp Asn Asn Val Ile
         10                  15                  20

ACG GCC GGC CCT AGA GGT CCT GTT TTA TTA CAA AGC ACT TGG TTT TTG          450
Thr Ala Gly Pro Arg Gly Pro Val Leu Leu Gln Ser Thr Trp Phe Leu
     25                  30                  35

GAA AAG TTA GCG GCG TTT GAC AGA GAA AGA ATC CCT GAA AGG GTG GTG          498
Glu Lys Leu Ala Ala Phe Asp Arg Glu Arg Ile Pro Glu Arg Val Val
 40                  45                  50                  55

CAT GCT AAA GGA AGC GGA GCT TAT GGC ACT TTC ACT GTG ACT AAA GAC          546
His Ala Lys Gly Ser Gly Ala Tyr Gly Thr Phe Thr Val Thr Lys Asp
                 60                  65                  70

ATC ACT AAA TAC ACT AAA GCG AAA ATT TTC TCT AAA GTG GGC AAA AAA          594
Ile Thr Lys Tyr Thr Lys Ala Lys Ile Phe Ser Lys Val Gly Lys Lys
             75                  80                  85

ACC GAA TGC TTC TTC AGA TTT TCC ACT GTG GCT GGT GAA AGA GGC AGT          642
Thr Glu Cys Phe Phe Arg Phe Ser Thr Val Ala Gly Glu Arg Gly Ser
         90                  95                 100

GCG GAT GCG GTA AGA GAC CCT AGA GGT TTT GCG ATG AAG TAT TAC ACT          690
Ala Asp Ala Val Arg Asp Pro Arg Gly Phe Ala Met Lys Tyr Tyr Thr
    105                 110                 115

GAA GAA GGT AAC TGG GAT TTA GTA GGG AAC AAC ACG CCT GTC TTC TTT          738
Glu Glu Gly Asn Trp Asp Leu Val Gly Asn Asn Thr Pro Val Phe Phe
120                 125                 130                 135

ATC CGT GAT GCG ATC AAA TTC CCT GAT TTC ATC CAC ACT CAA AAA CGA          786
Ile Arg Asp Ala Ile Lys Phe Pro Asp Phe Ile His Thr Gln Lys Arg
                140                 145                 150

GAT CCT CAA ACC AAT TTG CCT AAC CAT GAC ATG GTA TGG GAT TTT TGG          834
Asp Pro Gln Thr Asn Leu Pro Asn His Asp Met Val Trp Asp Phe Trp
            155                 160                 165

AGT AAT GTT CCT GAA AGC TTA TAC CAA GTA ACA TGG GTT ATG AGC GAT          882
Ser Asn Val Pro Glu Ser Leu Tyr Gln Val Thr Trp Val Met Ser Asp
        170                 175                 180
```

```
AGA GGG ATT CCT AAA TCT TTC CGC CAC ATG GAT GGT TTT GGC AGT CAC          930
Arg Gly Ile Pro Lys Ser Phe Arg His Met Asp Gly Phe Gly Ser His
    185                 190                 195

ACT TTC AGT CTT ATC AAC GCT AAA GGC GAA CGC TTT TGG GTG AAA TTC          978
Thr Phe Ser Leu Ile Asn Ala Lys Gly Glu Arg Phe Trp Val Lys Phe
200             205                 210                 215

CAC TTT CAC ACC ATG CAA GGC GTT AAG CAC TTG ACT AAC GAA GAA GCC         1026
His Phe His Thr Met Gln Gly Val Lys His Leu Thr Asn Glu Glu Ala
            220                 225                 230

GCA GAA GTC AGA AAA TAT GAT CCT GAT TCC AAT CAA AGG GAT TTA TTC         1074
Ala Glu Val Arg Lys Tyr Asp Pro Asp Ser Asn Gln Arg Asp Leu Phe
        235                 240                 245

AAT GCG ATC GCT AGA GGG GAT TTC CCA AAA TGG AAA TTA AGC GTT CAA         1122
Asn Ala Ile Ala Arg Gly Asp Phe Pro Lys Trp Lys Leu Ser Val Gln
    250                 255                 260

GTG ATG CCA GAA GAA GAT GCT AAG AAG TAT CGA TTC CAT CCG TTT GAT         1170
Val Met Pro Glu Glu Asp Ala Lys Lys Tyr Arg Phe His Pro Phe Asp
265             270                 275                 280

GTG ACT AAA ATT TGG TAC CTC CAA GAT TAT CCG TTG ATG GAA GTG GGC         1218
Val Thr Lys Ile Trp Tyr Leu Gln Asp Tyr Pro Leu Met Glu Val Gly
            285                 290                 295

ATT GTA GAG TTG AAT AAA AAT CCC GAA AAC TAT TTC GCA GAA GTG GAG         1266
Ile Val Glu Leu Asn Lys Asn Pro Glu Asn Tyr Phe Ala Glu Val Glu
        300                 305                 310

CAA GCG GCA TTC AGT CCG GCT AAT GTC GTT CCT GGA ATT GGC TAT AGC         1314
Gln Ala Ala Phe Ser Pro Ala Asn Val Val Pro Gly Ile Gly Tyr Ser
    315                 320                 325

CCT GAT AGG ATG TTA CAA GGG CGC TTG TTC TCT TAT GGG GAT ACA CAC         1362
Pro Asp Arg Met Leu Gln Gly Arg Leu Phe Ser Tyr Gly Asp Thr His
330                 335                 340                 345

CGC TAC CGC TTA GGG GTT AAT TAC CCT CAA ATA CCG GTT AAT AAA CCA         1410
Arg Tyr Arg Leu Gly Val Asn Tyr Pro Gln Ile Pro Val Asn Lys Pro
                350                 355                 360
```

```
AGA TGC CCG TTC CAC TCT TCT AGC AGA GAT GGT TAC ATG CAA AAT GGG        1458
Arg Cys Pro Phe His Ser Ser Ser Arg Asp Gly Tyr Met Gln Asn Gly
            365                 370                 375

TAT TAC GGC TCT TTA CAA AAC TAT ACG CCT AGC TCA TTG CCT GGC TAT        1506
Tyr Tyr Gly Ser Leu Gln Asn Tyr Thr Pro Ser Ser Leu Pro Gly Tyr
        380                 385                 390

AAA GAA GAT AAG AGC GCG AGA GAT CCT AAA TTC AAC TTA GCT CAT ATT        1554
Lys Glu Asp Lys Ser Ala Arg Asp Pro Lys Phe Asn Leu Ala His Ile
    395                 400                 405

GAG AAA GAG TTT GAA GTG TGG AAT TGG GAT TAC AGA GCT GAT GAT AGC        1602
Glu Lys Glu Phe Glu Val Trp Asn Trp Asp Tyr Arg Ala Asp Asp Ser
410                 415                 420                 425

GAT TAC TAC ACC CAA CCA GGT GAT TAC TAC CGC TCA TTG CCA GCT GAT        1650
Asp Tyr Tyr Thr Gln Pro Gly Asp Tyr Tyr Arg Ser Leu Pro Ala Asp
                430                 435                 440

GAA AAA GAA AGG TTG CAT GAC ACT ATT GGA GAG TCT TTG GCT CAT GTT        1698
Glu Lys Glu Arg Leu His Asp Thr Ile Gly Glu Ser Leu Ala His Val
            445                 450                 455

ACC CAT AAG GAA ATT GTG GAT AAA CAA TTG GAG CAT TTC AAG AAA GCT        1746
Thr His Lys Glu Ile Val Asp Lys Gln Leu Glu His Phe Lys Lys Ala
        460                 465                 470

GAC CCC AAA TAC GCT GAG GGG GTT AAA AAA GCT CTT GAA AAA CAC CAA        1794
Asp Pro Lys Tyr Ala Glu Gly Val Lys Lys Ala Leu Glu Lys His Gln
    475                 480                 485

AAA ATG ATG AAA GAC ATG CAT GGA AAA GAC ATG CAC CAC ACA AAA AAG        1842
Lys Met Met Lys Asp Met His Gly Lys Asp Met His His Thr Lys Lys
490                 495                 500                 505

AAA AAG TAA  CCCTTTTCTT TAAGCGTTCT TATTTTTTAG GAACGCTTTG               1891
Lys Lys

TCTTTCAAAA TTTAGGTTTT TGGATACTCA TCAGTCCTTT GGTGGTGTGT CCTATTTTTT      1951

CATTCATTCA ACGAATTTAA AAATTACAAT AAAGAGTTAT AGTTATGAAA CGAAGGGATT      2011
```

TTATTAAAAC GACTGCTTTA GGCGCTACAG GTGCTGTTTT AGGAGCACAG ATTTTGCAGG 2071

CAGAAGAAAG CAAAGGGAGT GTTGCAAAAT ATAAAATAGA AGCTCAATAC AGCATTGATT 2131

TTGATTCTGC AGAACACACT TCGCTTTTCA TTCCCATGCC GAGTGTTGTA GCGAGCAATG 2191

TGCATTTACA AGGCAATCAT GCCAGCTATA AAAGCATGCT CAATTTTGGA GTGCCTTATT 2251

TGCAAGTGGA TTTTTTAAAA AGCGCTCAAA AAAAGCAAGT CCATTTGTCT TATGAGATC 2310

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2639 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 1 to 1086 bp protein HopY
from 1099 to 1230 bp signal peptide of protein HopX
from 1231 to 2517 bp mature protein HopX

(vi) PROPERTIES: coding region coding for related proteins of *Helicobacter pylori* (HopY and HopX)

```
GAT CTA TCC CAA CAA TAC GCT AAT CAG GGT GTC ATT AAG CCT TTG GTG          48
Asp Leu Ser Gln Gln Tyr Ala Asn Gln Gly Val Ile Lys Pro Leu Val
                  5                  10                  15

GTG GAT GTG GGG AAA GAA CAA ATC GGT ATT ACT GAT AGC ATG CTC TTG          96
Val Asp Val Gly Lys Glu Gln Ile Gly Ile Thr Asp Ser Met Leu Leu
             20                  25                  30

GTG GCT CAA AAC ATC GTT TTA GCT TTA GGG CAA GTG GAT TTG AGC AAA         144
Val Ala Gln Asn Ile Val Leu Ala Leu Gly Gln Val Asp Leu Ser Lys
         35                  40                  45

ATC CAA CAA AAT AAA AAT AAT GGT AAC GGA CAG CTA TAC GAA AAC ATC         192
Ile Gln Gln Asn Lys Asn Asn Gly Asn Gly Gln Leu Tyr Glu Asn Ile
     50                  55                  60

ATG AAA GTC ATG CTT TTA GGT GCG GGC GGG ACT AAT GGA GCG TAT AAT         240
Met Lys Val Met Leu Leu Gly Ala Gly Gly Thr Asn Gly Ala Tyr Asn
65                  70                  75                  80
```

```
GGC GTG AGT GTG GGC GAT ATT GCC ACA GGC ATG CAA AAT TTT TCT TCG        288
Gly Val Ser Val Gly Asp Ile Ala Thr Gly Met Gln Asn Phe Ser Ser
                 85                  90                  95

CAA ACG GGC TTG ATA GGG GCT AAT TCT ACG GTT AGC GAG CTC AAC GCT        336
Gln Thr Gly Leu Ile Gly Ala Asn Ser Thr Val Ser Glu Leu Asn Ala
            100                 105                 110

TTG ATT AAG AGC GGG ATT TCT TTA GAT CGT GAG ACT TTG AGG TTA GGG        384
Leu Ile Lys Ser Gly Ile Ser Leu Asp Arg Glu Thr Leu Arg Leu Gly
        115                 120                 125

AGT TTT ATT GAA AAA AAT ATT TGT AGC AGT GCA TCG TCT TGT TTT ACT        432
Ser Phe Ile Glu Lys Asn Ile Cys Ser Ser Ala Ser Ser Cys Phe Thr
    130                 135                 140

GGG AGT CAG CTT ATC TAT AAG AAA GGG CTA GAT AGA ACC ATA AAC ATC        480
Gly Ser Gln Leu Ile Tyr Lys Lys Gly Leu Asp Arg Thr Ile Asn Ile
145                 150                 155                 160

ATT AAT GCG GTA TTA GGT CAG TTT GAA TCT TCG GCT AGT TCT CTT TAT        528
Ile Asn Ala Val Leu Gly Gln Phe Glu Ser Ser Ala Ser Ser Leu Tyr
                165                 170                 175

AAG ATT TCT TAT ATC CCT AAC CTC TTT TCG CTC AAA GAT TAC CAG TCA        576
Lys Ile Ser Tyr Ile Pro Asn Leu Phe Ser Leu Lys Asp Tyr Gln Ser
            180                 185                 190

GCG AGC ATG AAC GGC TTT GGG GCT AAG ATG GGT TAT AAA CAA TTT TTC        624
Ala Ser Met Asn Gly Phe Gly Ala Lys Met Gly Tyr Lys Gln Phe Phe
        195                 200                 205

ACC CAT AAG AAA AAT ATT GGC TTA AGG TAT TAC GGG TTT TTG GAT TAT        672
Thr His Lys Lys Asn Ile Gly Leu Arg Tyr Tyr Gly Phe Leu Asp Tyr
    210                 215                 220

GGC TAT GCG AAT TTT GGC GAT ACG AAT TTA AAA GTG GGA GCG AAT CTT        720
Gly Tyr Ala Asn Phe Gly Asp Thr Asn Leu Lys Val Gly Ala Asn Leu
225                 230                 235                 240

GTT ACT TAT GGG GTA GGA ACG GAT TTT TTA TAC AAC GTG TAT GAA CGC        768
Val Thr Tyr Gly Val Gly Thr Asp Phe Leu Tyr Asn Val Tyr Glu Arg
                245                 250                 255
```

```
TCT AGA AGG AGG GAA AGG ACT ACA ATC GGC CTT TTC TTT GGC GCT CAA          816
Ser Arg Arg Arg Glu Arg Thr Thr Ile Gly Leu Phe Phe Gly Ala Gln
            260                 265                 270

ATT GCA GGG CAA ACT TGG AGC ACT AAT GTA ACG AAC TTA TTG AGC GGG          864
Ile Ala Gly Gln Thr Trp Ser Thr Asn Val Thr Asn Leu Leu Ser Gly
        275                 280                 285

CAA AGG CCT GAT GTC AAG TCT AGT TCG TTC CAA TTC TTG TTT GAT TTG          912
Gln Arg Pro Asp Val Lys Ser Ser Ser Phe Gln Phe Leu Phe Asp Leu
    290                 295                 300

GGC GTG CGC ACC AAC TTT GCA AAA ACC AAT TTC AAC AAG CAC AGG CTA          960
Gly Val Arg Thr Asn Phe Ala Lys Thr Asn Phe Asn Lys His Arg Leu
305                 310                 315                 320

GAC CAA GGG ATA GAA TTT GGG GTG AAA ATC CCT GTT ATC GCT CAT AAA         1008
Asp Gln Gly Ile Glu Phe Gly Val Lys Ile Pro Val Ile Ala His Lys
                325                 330                 335

TAT TTC GCA ACC CAA GGC TCA AGC GCG AGC TAT ATG AGG AAT TTT AGC         1056
Tyr Phe Ala Thr Gln Gly Ser Ser Ala Ser Tyr Met Arg Asn Phe Ser
            340                 345                 350

TTC TAT GTG GGC TAT TCA GTC GGT TTT TAA  GGAAGGCTCT TG  ATG AAA         1104
Phe Tyr Val Gly Tyr Ser Val Gly Phe                     Met Lys
        355                 360

AAT ACC AAT ACA AAA GAG ATA AAG AAT ACA AGA ATG AAA AAA GGT TAT         1152
Asn Thr Asn Thr Lys Glu Ile Lys Asn Thr Arg Met Lys Lys Gly Tyr
        -40                 -35                 -30

AGT CAA TAC CAT GCG CTC AAA AAA GGG CTT TTA AAA ACT GCT CTG CTT         1200
Ser Gln Tyr His Ala Leu Lys Lys Gly Leu Leu Lys Thr Ala Leu Leu
    -25                 -20                 -15

TTT AGC CTT CCC TTA AGC ATG GCG TTA GCT GAA GAC GAT GGC TTT TAT         1248
Phe Ser Leu Pro Leu Ser Met Ala Leu Ala Glu Asp Asp Gly Phe Tyr
-10                 -5                                  5

ATG GGA GTG GGC TAT CAA ATC GGC GGT GCG CAA CAA AAC ATC AAT AAC         1296
Met Gly Val Gly Tyr Gln Ile Gly Gly Ala Gln Gln Asn Ile Asn Asn
```

```
                10                  15                  20

AAA GGC AGC ACC CTA AGG AAT AAT GTC ATT GAT GAT TTC CGC CAA GTG        1344
Lys Gly Ser Thr Leu Arg Asn Asn Val Ile Asp Asp Phe Arg Gln Val
          25                  30                  35

GGC GTG GGT ATG GCA GGG GGT AAC GGG CTT TTA GCT TTA GCG ACA AAC        1392
Gly Val Gly Met Ala Gly Gly Asn Gly Leu Leu Ala Leu Ala Thr Asn
      40                  45                  50

ACG ACC ATG GAC GCT CTT TTA GGG ATA GGC AAC CAA ATT GTC AAT ACC        1440
Thr Thr Met Asp Ala Leu Leu Gly Ile Gly Asn Gln Ile Val Asn Thr
 55                  60                  65                  70

AAT ACA ACT GTT GGC AAC AAC AAC GCA GAG TTA ACC CAG TTT AAA AAA        1488
Asn Thr Thr Val Gly Asn Asn Asn Ala Glu Leu Thr Gln Phe Lys Lys
                 75                  80                  85

ATA CTC CCC CAA ATT GAG CGA CGC TTT GAG ACG AAT AAA AAC GCT TAT        1536
Ile Leu Pro Gln Ile Glu Arg Arg Phe Glu Thr Asn Lys Asn Ala Tyr
             90                  95                 100

AGC GTT CAA GCC TTG CAA GTG TAT TTG AGT AAT GTG CTT TAT AAC TTG        1584
Ser Val Gln Ala Leu Gln Val Tyr Leu Ser Asn Val Leu Tyr Asn Leu
        105                 110                 115

GTT AAT AAT AGT AAT AAT GGC AGT AAT AAT GGA GTC GTT CCT GAA TAT        1632
Val Asn Asn Ser Asn Asn Gly Ser Asn Asn Gly Val Val Pro Glu Tyr
    120                 125                 130

GTA GGG ATT ATA AAA GTT CTC TAT AAT TCT CAA AAT GAA TTC AGT CTC        1680
Val Gly Ile Ile Lys Val Leu Tyr Asn Ser Gln Asn Glu Phe Ser Leu
135                 140                 145                 150

TTA GCC ACG GAG AGT GTG GCG CTT TTA AAC GCG CTT ACA AGG GTG AAT        1728
Leu Ala Thr Glu Ser Val Ala Leu Leu Asn Ala Leu Thr Arg Val Asn
                155                 160                 165

CTG GAT AGC AAT TCG GTG TTT TTA AAA GGG CTA TTA GCC CAA ATG CAG        1776
Leu Asp Ser Asn Ser Val Phe Leu Lys Gly Leu Leu Ala Gln Met Gln
            170                 175                 180

CTT TTT AAT GAC ACT TCT TCA GCA AAG CTA GGC CAG ATC GCA GAA AAC        1824
```

```
Leu Phe Asn Asp Thr Ser Ser Ala Lys Leu Gly Gln Ile Ala Glu Asn
        185                 190                 195

TTG AAT AAG AGT GGT GGT GCA GGG GCC ATG CTT CAA AAG GAT GTG AAA      1872
Leu Asn Lys Ser Gly Gly Ala Gly Ala Met Leu Gln Lys Asp Val Lys
    200                 205                 210

ACC ATC TCG GAT CGA ATC GCT ACT TAC CAA GAG AAT CTA AAA CAA CTA      1920
Thr Ile Ser Asp Arg Ile Ala Thr Tyr Gln Glu Asn Leu Lys Gln Leu
215                 220                 225                 230

GGA GGG ATG CTG AAT AAT TAC GAT GAG CCT TAC TTG CCC CAA TTT GGG      1968
Gly Gly Met Leu Asn Asn Tyr Asp Glu Pro Tyr Leu Pro Gln Phe Gly
                235                 240                 245

CCA GGC AAA AGC TCT CAG CAT GGG GTT ATT AAT GGC TTT GGC ATT CAA      2016
Pro Gly Lys Ser Ser Gln His Gly Val Ile Asn Gly Phe Gly Ile Gln
            250                 255                 260

GTG GGC TAT AAG CAA TTT TTT GGG AGC AAG AGG AAT ATA GGC TTA CGG      2064
Val Gly Tyr Lys Gln Phe Phe Gly Ser Lys Arg Asn Ile Gly Leu Arg
        265                 270                 275

TAT TAC GCT TTC TTT GAT TAT GGC TTT ACG CAA TTG GGC AGT CTT AAT      2112
Tyr Tyr Ala Phe Phe Asp Tyr Gly Phe Thr Gln Leu Gly Ser Leu Asn
    280                 285                 290

AGC GCT GTT AAA GCG AAC ATC TTT ACT TAT GGC GCT GGC ACG GAC TTT      2160
Ser Ala Val Lys Ala Asn Ile Phe Thr Tyr Gly Ala Gly Thr Asp Phe
295                 300                 305                 310

TTA TGG AAT ATC TTT AGA AGG GTT TTT AGC GAT CAG TCT TTG AAT GTG      2208
Leu Trp Asn Ile Phe Arg Arg Val Phe Ser Asp Gln Ser Leu Asn Val
                315                 320                 325

GGG GTG TTT GGG GGC ATT CAA ATA GCG GGT AAC ACT TGG GAT AGC TCT      2256
Gly Val Phe Gly Gly Ile Gln Ile Ala Gly Asn Thr Trp Asp Ser Ser
            330                 335                 340

TTA AGA GGC CAA ATT GAA AAC TCG TTT AAA GAA TAC CCC ACT CCC ACG      2304
Leu Arg Gly Gln Ile Glu Asn Ser Phe Lys Glu Tyr Pro Thr Pro Thr
        345                 350                 355
```

```
AAT TTC CAA TTT TTA TTT AAT TTG GGC TTA AGG GCT CAT TTT GCC AGC        2352
Asn Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg Ala His Phe Ala Ser
    360                 365                 370

ACC ATG CAC CGC CGG TTT TTG AGC TCG TCT CAA AGC ATT CAG CAT GGT        2400
Thr Met His Arg Arg Phe Leu Ser Ser Ser Gln Ser Ile Gln His Gly
375                 380                 385                 390

ATG GAA TTT GGC GTG AAA ATC CCG GCT ATC AAT CAA AGG TAT TTG AAA        2448
Met Glu Phe Gly Val Lys Ile Pro Ala Ile Asn Gln Arg Tyr Leu Lys
                395                 400                 405

GCG AAT GGG GCT GAT GTG GAT TAC AGG CGT TTG TAT GCG TTC TAT ATC        2496
Ala Asn Gly Ala Asp Val Asp Tyr Arg Arg Leu Tyr Ala Phe Tyr Ile
            410                 415                 420

AAC TAC ACG ATA GGT TTT TAA  GCTCTTTTTA GGGCTTATAA AGAGGTTCTT          2547
Asn Tyr Thr Ile Gly Phe
        425

TACTTTTTTT TGGTATTCTA ACAAGCTTTT AAACCATCCA ATCTACTTTG TTTTAAGGAT      2607

AATATTTTAT GGCAGATGTC GTTGTGGGGA TC                                    2639
```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1710 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 1 to 1179 bp mature peptide (without N-Terminus)

(vi) PROPERTIES: protein HopZ of *Helicobacter pylori*

```
ATC AAA AAC GCC CAA GAA ATC GTC GCA CAA GCT CAA AGC CTT AAC AAC          48
Ile Lys Asn Ala Gln Glu Ile Val Ala Gln Ala Gln Ser Leu Asn Asn
                  5                  10                  15

CCG CAA AAC AAT CAA AAC GCG CCG CAA GAT TTC AAT CCT TAC ACC TCT          96
Pro Gln Asn Asn Gln Asn Ala Pro Gln Asp Phe Asn Pro Tyr Thr Ser
             20                  25                  30

GCT GAT AGG GCT TTC GCT CAA AAC ATG CTC AAT CAC GCG CAA GCG CAA         144
Ala Asp Arg Ala Phe Ala Gln Asn Met Leu Asn His Ala Gln Ala Gln
         35                  40                  45

GCC AAG ATG CTT GAA CTA GCC AAT CAA ATC AAA ACC AAT CTT AGC GCT         192
Ala Lys Met Leu Glu Leu Ala Asn Gln Ile Lys Thr Asn Leu Ser Ala
     50                  55                  60

ATC CCG CAA CAT TTC ACC AAA GAT TAC TTG GCA GCT TGC CGC AAT GGG         240
Ile Pro Gln His Phe Thr Lys Asp Tyr Leu Ala Ala Cys Arg Asn Gly
65                  70                  75                  80

GGT GGG ACA TTA CCT GAT GCA GGG GTT ACT AAC AAC ACT TGG GGA GCC         288
Gly Gly Thr Leu Pro Asp Ala Gly Val Thr Asn Asn Thr Trp Gly Ala
```

```
                85                  90                  95

GGT TGC GCC TAT GTG GAA GAG ACC ATA ACG GCT TTA AAC AAC AGC CTT          336
Gly Cys Ala Tyr Val Glu Glu Thr Ile Thr Ala Leu Asn Asn Ser Leu
            100                 105                 110

GTG CAT TTT GGC ACT CAA GCC GAG CAA ATC AAG CAA TCT GAG TTG CTG          384
Val His Phe Gly Thr Gln Ala Glu Gln Ile Lys Gln Ser Glu Leu Leu
        115                 120                 125

GCG CGC ACG ATA TTT GAT TTT AAA GGC AGC CTT AAG GAT TTA AAC AGC          432
Ala Arg Thr Ile Phe Asp Phe Lys Gly Ser Leu Lys Asp Leu Asn Ser
    130                 135                 140

ACT TAT AAC AGC ATC ACC ACG ACC GCT TCA AAC ACG CCC AAT TCC CCA          480
Thr Tyr Asn Ser Ile Thr Thr Thr Ala Ser Asn Thr Pro Asn Ser Pro
145                 150                 155                 160

TTC CTT AAA AAT TTG ATA AGC CAA TCC ACT AAC CCT AAT AAC CCC GGG          528
Phe Leu Lys Asn Leu Ile Ser Gln Ser Thr Asn Pro Asn Asn Pro Gly
                165                 170                 175

GGC TTA CAG GCC GTT TAT CAA GTC AAC CAA AGC GCT TAT TCG CAA TTA          576
Gly Leu Gln Ala Val Tyr Gln Val Asn Gln Ser Ala Tyr Ser Gln Leu
            180                 185                 190

TTA AGC GCC ACG CAA GAA TTA GGG CAT AAC CCT TTC AGA CGC TTT GGA          624
Leu Ser Ala Thr Gln Glu Leu Gly His Asn Pro Phe Arg Arg Phe Gly
        195                 200                 205

TTA ATC AGC TCT CAA ACC AAC AAT GGT GCC ATG AAT GGG ATC GGT GTG          672
Leu Ile Ser Ser Gln Thr Asn Asn Gly Ala Met Asn Gly Ile Gly Val
    210                 215                 220

CAA ATA GGG TAT AAA CAA TTT TTT GGT GAA AAG AGA AAA TGG GGG GCT          720
Gln Ile Gly Tyr Lys Gln Phe Phe Gly Glu Lys Arg Lys Trp Gly Ala
225                 230                 235                 240

AGG TAT TAC GGC TTT TTT GAC TAT AAC CAT GCT TAT ATC AAA TCC AGC          768
Arg Tyr Tyr Gly Phe Phe Asp Tyr Asn His Ala Tyr Ile Lys Ser Ser
                245                 250                 255

TTT TTC AAC TCC GCC TCT GAT GTG TTC ACT TAT GGG GTA GGA ACA GAT          816
```

```
Phe Phe Asn Ser Ala Ser Asp Val Phe Thr Tyr Gly Val Gly Thr Asp
            260                 265                 270

GTC CTC TAT AAC TTT ATC AAC GAT AAA GCC ACC AAA AAC AAT AAG ATT        864
Val Leu Tyr Asn Phe Ile Asn Asp Lys Ala Thr Lys Asn Asn Lys Ile
        275                 280                 285

TCT TTT GGG GTG TTT GGG GGG ATT GCT TTA GCT GGC ACT TCG TGG CTC        912
Ser Phe Gly Val Phe Gly Gly Ile Ala Leu Ala Gly Thr Ser Trp Leu
    290                 295                 300

AAT TCT CAA TAC GTG AAT TTA GCG ACC TTC AAT AAT TTC TAT AGC GCT        960
Asn Ser Gln Tyr Val Asn Leu Ala Thr Phe Asn Asn Phe Tyr Ser Ala
305                 310                 315                 320

AAA ATG AAT GTG GCG AAT TTC CAA TTC TTA TTC AAC TTG GGC TTG AGA       1008
Lys Met Asn Val Ala Asn Phe Gln Phe Leu Phe Asn Leu Gly Leu Arg
                325                 330                 335

ATG AAT CTG GCT AAA AAC AAA AAG AAA GCG AGC GAT CAT GCG GCT CAA       1056
Met Asn Leu Ala Lys Asn Lys Lys Lys Ala Ser Asp His Ala Ala Gln
            340                 345                 350

CAT GGC GTG GAA CTA GGC GTG AAG ATC CCC ACG ATC AAC ACG AAT TAC       1104
His Gly Val Glu Leu Gly Val Lys Ile Pro Thr Ile Asn Thr Asn Tyr
        355                 360                 365

TAT TCT TTG CTA GGC ACT CAA CTA GAA TAC CGC AGA CTC TAT AGC GTG       1152
Tyr Ser Leu Leu Gly Thr Gln Leu Glu Tyr Arg Arg Leu Tyr Ser Val
    370                 375                 380

TAT TTG AAT TAT GTG TTT GCG TAT TAA   AAGCTTGCGT TAAACCCTTT          1199
Tyr Leu Asn Tyr Val Phe Ala Tyr
385                 390

GTGGAACTCC CTTTTTAAGG GGTTTCTTTT GAAGCCTTTT TTTTGAACCT TTTTTTGGGG      1259

GTCAAGCGTA AAATCCACCC CTATCCCTTT AAGAAAATAA AATAAAACTT TAAGAACTTT      1319

AAGAACTTTA AGAAAATGCG TTTTACAACA AAATAAGATC TAAAACAATA AAACAAAACC      1379

CCATTTTTTA ACAATGAAAT TTTTTAAACA AAAAAGCATT AAATCCTAAT AAGGTTTGTT      1439
```

AGATCTTGAT AAAAACAAAG CTTTTTTAAA ACCCCAAAAA CAATACTAAC CAATAACCAA 1499

AACGCATCTA TTGTGATCCT TATAGCATAA AACCAAGTTT TTATTTAAGC AAAAGCTGTT 1559

ATGCCGTTTT AAGAGCGTTT CGTTTCTATG AAAACCGCAA TATTTTTCAA TTATTCTTGA 1619

CAAGCGTTAA AAAAAATTGT ATCATTATCT TTTTGTGAGA CCCGTTAGCT CAGTTGGTAG 1679

AGCAATTCCC TTTTAAGGAA TGGGAGCGGC C 1710

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2078 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 642 to 692 bp signal peptide
from 693 to 1388 bp mature peptide

(vi) PROPERTIES: 28 kD protein from *Helicobacter pylori*

GATCCATCGC TAAAAGCATT AATTTTGGGC TGGTGTATGG CATGGGGAGT AAGAAATTGA 60

GCGAAACTTT AAACATTTCT TTAAATGAGG CTAAAAGCTA CATAGAAGCG TATTTCAAGC 120

GATTCCCTAG CATCAAAGAT TATTTAAACC GCATGAAAGA AGAGATTTTA AAAACTTCTA 180

AAGCCTTCAC TTTGCTTGGG CGTTATCGGG TGTTTGATTT TACCGGCGCA AATGACTATA 240

TCAAGGGCAA TTATTTGCGA GAGGGCGTGA ATGCGATTTT TCAAGGGAGT GCCAGCGATT 300

TATTGAAATT AGGCATGCTC AAAGTGAGCG AGCGTTTCAA AAATAACCCT TCGGTGAGGC 360

TGCTTTTGCA AGTGCATGAC GAATTGATTT TTGAAATTGA AGAAAAAAAC GCATTAGAGT 420

TGCAGCAAGA AATCCAACGC ATTCTTAATA ATGAAGTGTA TCCTTTGAGG GTGCCACTAG 480

AAACGAGCGC GTTTGTGGCG AATCGTTGGA ATGAACTAAA AGGTTAGTTT TTTAATTGAG 540

TTTAAGAAAA AATATATTAT TATTTCTTTA TAAGTAATAC TTAGCTATTT TAAGCTAATA 600

TAATATAGAG CGATTATCAA AAAATAAAGG GAAAAGACTG A ATG TTG AAA AGA ATT 656
Met Leu Lys Arg Ile

-15

```
ATA TTA TTA GGG GCT TTG GGT GTT TTA GCG AAC GCT GAA GAG AGC GCG          704
Ile Leu Leu Gly Ala Leu Gly Val Leu Ala Asn Ala Glu Glu Ser Ala
        -10                 -5

GCT TTT GTG GGA GTC AAT TAC CAG GTG AGC ATG ATA CAA AAT CAG ACT          752
Ala Phe Val Gly Val Asn Tyr Gln Val Ser Met Ile Gln Asn Gln Thr
  5                  10                  15                  20

AAA ATG GTG AAT GAC AAC GGC TTG CAA AAG CCT TTG ATA AAG TTC CCG          800
Lys Met Val Asn Asp Asn Gly Leu Gln Lys Pro Leu Ile Lys Phe Pro
                25                  30                  35

CCT TAT GCA GGA GCG GGT TTT GAA GTG GGC TAT AAA CAA TTT TTT GGC          848
Pro Tyr Ala Gly Ala Gly Phe Glu Val Gly Tyr Lys Gln Phe Phe Gly
            40                  45                  50

AAG AAA AAA TGG TTT GGT GCG CGT TAT TAT GGG TTT TTT GAC TAC GCG          896
Lys Lys Lys Trp Phe Gly Ala Arg Tyr Tyr Gly Phe Phe Asp Tyr Ala
        55                  60                  65

CAC AAC CGC TTT GGC GTG ATG AAA AAG GGT ATC CCG GTG GGC GAG AGC          944
His Asn Arg Phe Gly Val Met Lys Lys Gly Ile Pro Val Gly Glu Ser
    70                  75                  80

GGG TTT ATT TAC AAT AGT TTT AGT TTT GGA GGG AAC ACT TTA ATG GAA          992
Gly Phe Ile Tyr Asn Ser Phe Ser Phe Gly Gly Asn Thr Leu Met Glu
 85                  90                  95                  100

AGG GAT TCC TAT CAA GGG CAA TAC TAT GTC AAT TTA TTC ACT TAT GGT         1040
Arg Asp Ser Tyr Gln Gly Gln Tyr Tyr Val Asn Leu Phe Thr Tyr Gly
                105                 110                 115

GTG GGG CTA GAT ACG CTG TGG AAT TTT GTG AAT AAA GAA AAC ATG GTT         1088
Val Gly Leu Asp Thr Leu Trp Asn Phe Val Asn Lys Glu Asn Met Val
            120                 125                 130

TTT GGT TTT GTG GTA GGA ATC CAA TTA GCT GGG GAT AGT TGG GCA ACG         1136
Phe Gly Phe Val Val Gly Ile Gln Leu Ala Gly Asp Ser Trp Ala Thr
        135                 140                 145

AGC ATC AGT AAA GAG ATC GCC AGC TAT GCA AAA CAC CAC AGC AAT TCC         1184
```

```
Ser Ile Ser Lys Glu Ile Ala Ser Tyr Ala Lys His His Ser Asn Ser
    150                 155                 160

AGT TAT AGC CCG GCC AAT TTC CAG TTT TTA TGG AAG TTT GGG GTC CGC      1232
Ser Tyr Ser Pro Ala Asn Phe Gln Phe Leu Trp Lys Phe Gly Val Arg
165                 170                 175                 180

ACC CAT ATC GCT AAA CAC AAT AGC CTA GAA TTA GGG ATT AAA GTG CCT      1280
Thr His Ile Ala Lys His Asn Ser Leu Glu Leu Gly Ile Lys Val Pro
                185                 190                 195

ACG ATC ACG CAC CGG CTT TTC TCT CTT ACC AAC GAA AAG GGA TAC ACC      1328
Thr Ile Thr His Arg Leu Phe Ser Leu Thr Asn Glu Lys Gly Tyr Thr
            200                 205                 210

TTA CAG GCT GAT GTG CGC CGA GTT TAT GCG TTT CAA ATC AGT TAC TTG      1376
Leu Gln Ala Asp Val Arg Arg Val Tyr Ala Phe Gln Ile Ser Tyr Leu
        215                 220                 225

AGG GAT TTT TAA CCCCTTTTTA GATACAATCG CACCTAAAAT CAATTTAAAG          1428
Arg Asp Phe
    230

GTGTGAAATG GTGAATTTAG AAAATTTAGA CTGGAAAAAT TTAGGCTTTA GCTACATTAA    1488

AACGGATTTT CGCTTCATCG CTACTTATAA AAACGGCTCT TGGTCGCATG GCGGATTGGT    1548

GAGCGAAAAT GTGCTACAAA TCAGCGAAGG CTCGCCGGTC TTGCACTACG GGCAGGCTTG    1608

TTTTGAAGGC TTGAAGGCTT ACCGCTCTCA AAAGGGGAAG GCTTTACTTT TTCGCCCTTT    1668

AGAAAACGCC AAACGCTTGC AAACTTCATG CGAAAGACTG CTCATGCCCA AAGTGAGCGA    1728

AGAGCTGTTT TTAAGGGCAT GCGCTGAAGT AGTCAAAGCG AATCAAAAAT GGCTCGCTCC    1788

TTATAAAAGC GGGGCGAGTT TGTATTTGCG CCCTTTTGTC ATAGGCGTAG GGGATAATTT    1848

GGGGGTGAAG CCGGCTAATG AATACCTTTT TATCGTGTTT TGCGCGCCTG TGGGGGCGTA    1908

TTTTAAGGGG GGTATAGAAA AAGGAGGGGC TAGGTTTATC ACTACGATTT TTGATAGGGC    1968

CGCGCCTAAA GGCACCGGAG GGGTGAAAGT GGGGGGGAAT TACGCTGCAA GCCTGTTAGC    2028
```

CCATAAAATA GCCACAGAGC AGGGCTATGA TGATTGCATT TATTTAGATC 2078

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3144 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 1149 to 1235 bp signal peptide
from 1236 to 2735 bp mature peptide

(vi) PROPERTIES: 50 kD membrane protein from *Helicobacter pylori*

GATCGGCAGG CAAATACACA TCTTTATTGC AACCCATTCC TTCGTATTTT TCAACTTTCA 60

AGGTCCCCAC CAATAATTCC TTATGCTTGC CTTTCCAAAG CGCGGTCGTG TCGTTGGTGG 120

CATCATTTTT ATTCGCAAAT ACCAGATACA TTTGGTATTC TATGGGCTTA GTTTTAAGGT 180

GTTGTTGGAA TGAAGAAAGC AGATAATTTG AATCTTTTTG CTTTAATTCT TGGGGGTTAA 240

GATACTTAAT GCCCTCTTTA GGCACAAATT TCCATCTCGC AGGCAATAAT TTTTCTTTCT 300

TATCTTTAAA CCTGAACGCA TGCACGCTAT AATAGGGCGT GTTAGCCACG CTTGAGCTAA 360

TCCCTATCGT TTTGGTGTAA GCGGCAAAAT TCTTATAAGA GGGGACTTCT TCATAAAGCT 420

TTTTGATTCT TGCTTCATCC ACCTTGCCAT TTTTAGGGAT TCTCATCTCA AAGAATTGGG 480

CGAATTCGTT AGGGTTTTTG GCAAAATTGA TTTCTGTATT GAGCATCACC ATTGTCCAGC 540

TAGCGTTTTG ATTTTCTAAT TTTAACGCCA TTCCCCTAAC TTTGCTTTTA TCGTCCATTG 600

CCACGCCTCC TAAAGAATAC CTTACAGATG CAGGGATTTC TTTTTCATTG AGTAATGGCA 660

```
CATCTAAATC CTTTTTTGCT TGCGCATTAG GGAGGAACAC GCCTTTAGCA CAAAACCCCT      720

TAGTGTGGTT GATTTTCATT TTAGGCTCTT TGGCGTTGAG CTTGTAGAAA ATATCCGCAA      780

TCTCTTCAGC GCTCACTTCA TGGGCTTTTA AAAAACCCAA GCTAAAAACC AAACACAAGC      840

TCAAACCAAT TTTTTTCATT GTTTCGCTCC TTAATTATTA TTTTTATATA AAACAACGCT      900

TTCTATTGTA TCAGTAAATT CCCTATTGAG CCTTAAAAAA GCCTTTTTTT CAATATATTA      960

TTAGGTCTTA AAAAATTATC CTATCCATTT GCATGCTCAT GAGCAAGCTT TTAAGGATAA     1020

ACTTGTGTTT TAAATTTTGT GATTTTTAAG AAAAATTAGC TTGATTTTAA ACTAATTCTA     1080

TATTCTTTTA TGCTACAATT ATTTCTACAG AGTAATTTAT CTATTCTCAG GTAAAGTAAG     1140

GAAGAGGA ATG AAA TTA AAG AAA CGA AAA GTT GCG GCT ACA TTG CTA AAG       1190
         Met Lys Leu Lys Lys Arg Lys Val Ala Ala Thr Leu Leu Lys
                            -25                 -20

CGT TTT ACC TTA CCA CTA TTG TTC ACT ACG GGT TCA TTA GGG GCG GTT       1238
Arg Phe Thr Leu Pro Leu Leu Phe Thr Thr Gly Ser Leu Gly Ala Val
-15                 -10                 -5

ACT TAT GAA GTG CAT GGG GAT TTT ATC AAC TTC TCC AAA GTG GGT TTT       1286
Thr Tyr Glu Val His Gly Asp Phe Ile Asn Phe Ser Lys Val Gly Phe
             5                  10                  15

AAC CAT TCG CCC ATT AAC CCT GTT AAA GGT ATC TAT CCC ACA GAG ACT       1334
Asn His Ser Pro Ile Asn Pro Val Lys Gly Ile Tyr Pro Thr Glu Thr
         20                  25                  30

TTT GTT AAC CTT ACG GGT AAG CTA GAG GGT TCT GTG CAT TTA GGT AGG       1382
Phe Val Asn Leu Thr Gly Lys Leu Glu Gly Ser Val His Leu Gly Arg
     35                  40                  45

GGA TGG ACC GTG AAT TTA GGC GGT GTT TTG GGC GGA CAG GCT TAT GAT       1430
Gly Trp Thr Val Asn Leu Gly Gly Val Leu Gly Gly Gln Ala Tyr Asp
50                  55                  60                  65

GGC ACT AAG TAT GAT AGG TGG GCG AAG GAT TTT ACC CCC CCA AGC TAT       1478
Gly Thr Lys Tyr Asp Arg Trp Ala Lys Asp Phe Thr Pro Pro Ser Tyr
                70                  75                  80
```

```
TGG GAT AAA ACT TCT TGC GGC ACT GAT TCT ATG AGC CTT TGT ATG AAT        1526
Trp Asp Lys Thr Ser Cys Gly Thr Asp Ser Met Ser Leu Cys Met Asn
            85                  90                  95

GCT ACT AAA ATG TGG CAA CAA TCA GGG CCA GGT GGT GTC ATT AAC CCT        1574
Ala Thr Lys Met Trp Gln Gln Ser Gly Pro Gly Gly Val Ile Asn Pro
        100                 105                 110

AGA GGT ATT GGT TGG GAA TAT ATG GGT GAG TGG AAC GGC TTG TTC CCT        1622
Arg Gly Ile Gly Trp Glu Tyr Met Gly Glu Trp Asn Gly Leu Phe Pro
    115                 120                 125

AAC TAC TAT CCG GCT AAC GCC TAC TTG CCT GGT GGC TCA AGG CGT TAT        1670
Asn Tyr Tyr Pro Ala Asn Ala Tyr Leu Pro Gly Gly Ser Arg Arg Tyr
130                 135                 140                 145

CAA GTC TAT AAA GCA AAT TTG ACC TAT GAC AGC GAC AGA GTC CAT ATG        1718
Gln Val Tyr Lys Ala Asn Leu Thr Tyr Asp Ser Asp Arg Val His Met
                150                 155                 160

GTA ATG GGG CGT TTT GAT ATT ACC GAG CAG GAG CAA ATG GAT TGG ATT        1766
Val Met Gly Arg Phe Asp Ile Thr Glu Gln Glu Gln Met Asp Trp Ile
            165                 170                 175

TAC CAA TTG TTC CAA GGG TTT TAT GGG ACT TTC AAG CTC ACT AAG AAT        1814
Tyr Gln Leu Phe Gln Gly Phe Tyr Gly Thr Phe Lys Leu Thr Lys Asn
        180                 185                 190

ATG AAA TTC TTG CTC TTT AGT GGT TGG GGT CGT GGT ATC GCT GAT GGT        1862
Met Lys Phe Leu Leu Phe Ser Gly Trp Gly Arg Gly Ile Ala Asp Gly
    195                 200                 205

CAG TGG TTG TTC CCT ATC TAT CGT GAA AAG CCT TGG GGG GTT CAT AAA        1910
Gln Trp Leu Phe Pro Ile Tyr Arg Glu Lys Pro Trp Gly Val His Lys
210                 215                 220                 225

GCG GGT ATT ATT TAT CGC CCT ACA AAG AAT TTG ATG ATC CAC CCT TAT        1958
Ala Gly Ile Ile Tyr Arg Pro Thr Lys Asn Leu Met Ile His Pro Tyr
                230                 235                 240

GTG TAT CTT ATC CCA ATG GTA GGC ACA TTG CCC GGT GTT AAA GTA GAG        2006
Val Tyr Leu Ile Pro Met Val Gly Thr Leu Pro Gly Val Lys Val Glu
```

```
            245                 250                 255

TAT GAT ACC AAT CCG GAA TTT AGC GGT AGG GGC ATT AGG AAT AAA ACG        2054
Tyr Asp Thr Asn Pro Glu Phe Ser Gly Arg Gly Ile Arg Asn Lys Thr
        260                 265                 270

ACT TTC TAT GCG TTG TAT GAC TAT CGT TGG AAT AAC GCT GAA TAC GGT        2102
Thr Phe Tyr Ala Leu Tyr Asp Tyr Arg Trp Asn Asn Ala Glu Tyr Gly
    275                 280                 285

CGT TAT GCG CCC GCT CGT TAT AAC ACT TGG GAT CCG TTC TTG GAT AAT        2150
Arg Tyr Ala Pro Ala Arg Tyr Asn Thr Trp Asp Pro Phe Leu Asp Asn
290                 295                 300                 305

GGT AAG TGG CGT GGC TTG CAA GGT CCT GGC GGT GCG ACG CTT CTT TTG        2198
Gly Lys Trp Arg Gly Leu Gln Gly Pro Gly Gly Ala Thr Leu Leu Leu
                310                 315                 320

CGC CAC CAT ATA GAT ATT AAC AAC TAT TTT GTG GTT GGT GGT GCT TAT        2246
Arg His His Ile Asp Ile Asn Asn Tyr Phe Val Val Gly Gly Ala Tyr
            325                 330                 335

CTC AAC ATT GGT AAC CCT AAC ATG AAC TTA GGT ACT TGG GGT AAC CCT        2294
Leu Asn Ile Gly Asn Pro Asn Met Asn Leu Gly Thr Trp Gly Asn Pro
        340                 345                 350

GTG GCT CTT GAT GGT ATC GAA CAA TGG GTC GGT AGT ATC TAC AGC TTA        2342
Val Ala Leu Asp Gly Ile Glu Gln Trp Val Gly Ser Ile Tyr Ser Leu
    355                 360                 365

GGG TTT GCG GGG ATT GAC AAC ATT ACC GAT GCT GAT GCG TTC ACC GAG        2390
Gly Phe Ala Gly Ile Asp Asn Ile Thr Asp Ala Asp Ala Phe Thr Glu
370                 375                 380                 385

TAT GTT AAA GGT GGA GGC AAG CAT GGT AAG TTC AGT TGG AGC GTT TAT        2438
Tyr Val Lys Gly Gly Gly Lys His Gly Lys Phe Ser Trp Ser Val Tyr
                390                 395                 400

CAG CGC TTC ACC ACT GCA CCA AGG GCT TTG GAA TAT GGT ATC GGT ATG        2486
Gln Arg Phe Thr Thr Ala Pro Arg Ala Leu Glu Tyr Gly Ile Gly Met
            405                 410                 415

TAT CTA GAT TAT CAG TTC AGC AAG CAT GTT AAA GCG GGT CTC AAA CTC        2534
```

```
Tyr Leu Asp Tyr Gln Phe Ser Lys His Val Lys Ala Gly Leu Lys Leu
        420                 425                 430

GTA TGG TTA GAG TTC CAA ATT CGT GCG GGT TAC AAC CCT GGA ACC GGT        2582
Val Trp Leu Glu Phe Gln Ile Arg Ala Gly Tyr Asn Pro Gly Thr Gly
    435                 440                 445

TTC CTT GGG CCA AAC GGT CAG CCG CTT AAC TTG AAT ACT GGT TTG TTT        2630
Phe Leu Gly Pro Asn Gly Gln Pro Leu Asn Leu Asn Thr Gly Leu Phe
450                 455                 460                 465

GAG TCT TCA GCG TTC GCG CAA GGC CCT CAA AAC ATG GGC GGT ATC GCA        2678
Glu Ser Ser Ala Phe Ala Gln Gly Pro Gln Asn Met Gly Gly Ile Ala
                470                 475                 480

AAA AGC ATC ACT CAA GAC AGA AGC CAT TTG ATG ACA CAC ATT AGT TAT        2726
Lys Ser Ile Thr Gln Asp Arg Ser His Leu Met Thr His Ile Ser Tyr
            485                 490                 495

AGT TTC TAA GAGAGTTCTC CCCCTATCTC TTAGATATGC CTTTTTTGTA TTTTTATTTT     2785
Ser Phe

AATATCTTTG GGAGTTAGAG TTTTGGAAAT TAAGAAATAT TTTTCTTACT TTCTATTTTT     2845

TTTGCTTTTT TCTAGTCTCT TTTTATCCAA ACTTCAAGCT TATAAATTCA ACATGAGCAT     2905

TGTTGGAAAG GTGAGCAGCT ATACCAAGTT TGGCTTTAAC AACCAAAGAT ACCAGCCTTC     2965

TAAAGACATT TATCCTACAG GTAGCTACAC TTCTTTGCTC GGCGAATTGA ATTTGAGCAT     3025

GGGTTTATAC AAGGGTTTGA GGGCGGAAGT GGGGGCTATG ATGGCAGCGC TCCCCTATGA     3085

CTCTACCGCC TATCAAGGCA ACAATATCCC TAACGGCCAG CCCGGCTCTA GGACCGATC      3144
```

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2825 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 891 to 2090 bp protein

(vi) PROPERTIES: 42 kD protein from *Helicobacter pylori*

GATCGGAAGC GAGAGTGATA AAAGGCATTA GGAGAGTTAG AGACATCACT TCCACAAAAG 60

AAGAAAAAAC CGCCATCAGC ACAAGCAGGA AAAAAACCAT TTTTTCCTTG AAAGTGATGA 120

GCATATAGAT TTGCTTTAAA GAACGCAAAA AGTATTTTAA AGTAGAGATT TTATGTTTTT 180

TTTTCGCCAT AATTTAAGTG TCCATAAATT CTTTTATATG TAATAAGCTT GAGCTGTGTT 240

AAGCCAAATT GAGCTAGATT ATAGCTAAAT TTTAACCATG CTCTGTGCCA TACGAATAAT 300

TTAGCTTTCT GCCATCATTT CTTGACAAGT CAAGTATAAA ACTGCTATAA TCCCAAGTCT 360

TTAATTTGTT TAATTTGTTG CTGGCTTAGC TCAGTTGGTA GAGCAGCTGC CTTGTAAGCA 420

GCAGGTCGGG GGTTCAAGTC CCTTAGCCAG CTCCAGTTGA AATGTTATTG TGCAAAGTTT 480

TTGGTGAGAT ACTCAAGTGG CCAACGAGGG CAGACTGTAA ATCTGCTGAC TATGTCTTCC 540

GTGGTTCGAA TCCACGTCTC ACCACCATTT TGTTTTATAG ATGCGGGAAT AGCTCAGTTG 600

GCTAGAGCAT CAGCCTTCCA AGCTGAGGGT CGCGGGTTCG AGTCCCGTTT CCCGCTCCAT 660

TTTTAGGATA ACATTTTAGT TTTTGAGGCG CCTATATAGC TCAGAGGCAG AGCACTTCCT 720

```
TGGTAAGGAA GAGGTCGGCG GTTCAATTCC GCTTATAGGC TCCAGTTTAT AATCTCTTGA      780

ATGGCGATAA GACAAAAATG TCTTAAATTT TGTGGTAGCA TTTAGGAATA CTTAGGATTT      840

TGTTTAGTAT AATTCTAAAA TCCATTTCAA AAAATTAAGG AGAAATACAA    ATG          893
                                                          Met

GCA AAA GAA AAG TTT AAC AGA ACT AAG CCG CAT GTT AAT ATT GGA ACC         941
Ala Lys Glu Lys Phe Asn Arg Thr Lys Pro His Val Asn Ile Gly Thr
              5                  10                  15

ATT GGG CAT GTA GAC CAT GGT AAA ACG ACT TTG AGT GCA GCG ATT TCA         989
Ile Gly His Val Asp His Gly Lys Thr Thr Leu Ser Ala Ala Ile Ser
         20                  25                  30

GCG GTG CTT TCT TTG AAA GGT CTT GCA GAA ATG AAA GAC TAT GAT AAT        1036
Ala Val Leu Ser Leu Lys Gly Leu Ala Glu Met Lys Asp Tyr Asp Asn
     35                  40                  45

ATT GAT AAC GCC CCT GAA GAA AAA GAA AGA GGG ATC ACT ATC GCT ACT        1085
Ile Asp Asn Ala Pro Glu Glu Lys Glu Arg Gly Ile Thr Ile Ala Thr
50                  55                  60                  65

TCC CAC ATT GAA TAT GAG ACT GAA AAC AGA CAC TAT GCG CAT GTG GAT        1133
Ser His Ile Glu Tyr Glu Thr Glu Asn Arg His Tyr Ala His Val Asp
                 70                  75                  80

TGC CCA GGA CAC GCT GAC TAT GTA AAA AAC ATG ATC ACC GGT GCG GCA        1181
Cys Pro Gly His Ala Asp Tyr Val Lys Asn Met Ile Thr Gly Ala Ala
             85                  90                  95

CAA ATG GAC GGA GCG ATT TTG GTT GTT TCT GCA GCT GAT GGC CCT ATG        1229
Gln Met Asp Gly Ala Ile Leu Val Val Ser Ala Ala Asp Gly Pro Met
        100                 105                 110

CCT CAA ACC AGA GAG CAT ATC TTA TTG TCT CGT CAA GTA GGC GTG CCT        1277
Pro Gln Thr Arg Glu His Ile Leu Leu Ser Arg Gln Val Gly Val Pro
    115                 120                 125

CAC ATC GTT GTT TTC TTA AAC AAA CAA GAC ATG GTA GAT GAC CAA GAA        1325
His Ile Val Val Phe Leu Asn Lys Gln Asp Met Val Asp Asp Gln Glu
130                 135                 140                 145
```

```
TTG TTA GAG TTG GTA GAA ATG GAA GTG CGC GAA TTG TTG AGC GCG TAT        1373
Leu Leu Glu Leu Val Glu Met Glu Val Arg Glu Leu Leu Ser Ala Tyr
                150                 155                 160

GAA TTC CCT GGT GAT GAC ACT CCT ATC GTA GCG GGT TCA GCT TTA AGA        1421
Glu Phe Pro Gly Asp Asp Thr Pro Ile Val Ala Gly Ser Ala Leu Arg
            165                 170                 175

GCT TTA GAG GAA GCA AAG GCT GGT AAT GTG GGT GAA TGG GGT GAA AAA        1469
Ala Leu Glu Glu Ala Lys Ala Gly Asn Val Gly Glu Trp Gly Glu Lys
        180                 185                 190

GTG CTT AAG CTC ATG GCT GAA GTG GAT GCC TAT ATC CCT ACT CCA GAA        1517
Val Leu Lys Leu Met Ala Glu Val Asp Ala Tyr Ile Pro Thr Pro Glu
    195                 200                 205

AGA GAC ACT GAA AAA ACT TTC TTG ATG CCG GTT GAA GAT GTG TTC TCT        1565
Arg Asp Thr Glu Lys Thr Phe Leu Met Pro Val Glu Asp Val Phe Ser
210                 215                 220                 225

ATT GCG GGT AGA GGG ACT GTG GTT ACA GGT AGG ATT GAA AGA GGT GTG        1613
Ile Ala Gly Arg Gly Thr Val Val Thr Gly Arg Ile Glu Arg Gly Val
                230                 235                 240

GTG AAA GTA GGC GAT GAA GTG GAA ATC GTT GGT ATC AGA GCT ACA CAA        1661
Val Lys Val Gly Asp Glu Val Glu Ile Val Gly Ile Arg Ala Thr Gln
            245                 250                 255

AAA ACG ACT GTA ACC GGT GTG GAA ATG TTT AGA AAA GAG CTA GAA AAA        1709
Lys Thr Thr Val Thr Gly Val Glu Met Phe Arg Lys Glu Leu Glu Lys
        260                 265                 270

GGT GAG GCC GGC GAT AAT GTG GGC GTG CTT TTG AGA GGA ACT AAA AAA        1757
Gly Glu Ala Gly Asp Asn Val Gly Val Leu Leu Arg Gly Thr Lys Lys
    275                 280                 285

GAA GAA GTA GAA CGC GGT ATG GTT CTA TGC AAA CCA GGT TCT ATC ACT        1805
Glu Glu Val Glu Arg Gly Met Val Leu Cys Lys Pro Gly Ser Ile Thr
290                 295                 300                 305

CCG CAC AAG AAA TTT GAG GGA GAA ATT TAT GTC CTT TCT AAA GAA GAA        1853
Pro His Lys Lys Phe Glu Gly Glu Ile Tyr Val Leu Ser Lys Glu Glu
```

310 315 320

GGC GGG AGA CAC ACT CCA TTC TTC ACC AAT TAC CGC CCG CAA TTC TAT 1901
Gly Gly Arg His Thr Pro Phe Phe Thr Asn Tyr Arg Pro Gln Phe Tyr
325 330 335

GTG CGC ACG ACT GAT GTG ACT GGC TCT ATC ACC CTT CCT GAA GGC GTA 1949
Val Arg Thr Thr Asp Val Thr Gly Ser Ile Thr Leu Pro Glu Gly Val
340 345 350

GAA ATG GTT ATG CCT GGC GAT AAT GTG AAA ATC ACT GTA GAG TTG ATT 1997
Glu Met Val Met Pro Gly Asp Asn Val Lys Ile Thr Val Glu Leu Ile
355 360 365

AGC CCT GTT GCG TTA GAG TTG GGA ACT AAA TTT GCG ATT CGT GAA GGC 2045
Ser Pro Val Ala Leu Glu Leu Gly Thr Lys Phe Ala Ile Arg Glu Gly
370 375 380 385

GGT AGG ACC GTT GGT GCT GGT GTT GTG AGC AAT ATT ATT GAA TAA 2090
Gly Arg Thr Val Gly Ala Gly Val Val Ser Asn Ile Ile Glu
390 395

TATTAGCAAA AAGAGTTACC ATAAAGGGTC ATTATGAAAG TTAAAATAGG GTTGAAGTGT 2150

TCTGATTGTG AAGATATCAA TTACAGCACA ACCAAGAACG CTAAAACTAA CACTGAAAAA 2210

CTGGAGCTTA AGAAGTTCTG CCCAAGGGAA AACAAACACA CTCTTCATAA AGAAATCAAA 2270

TTGAAGAGCT AGTTCTTTCT TTTGTGTTGT GATTGAAAAG GAGGGGAGGT TAGGTCAGTA 2330

GCTCCAATGG TAGAGCGTCG GTCTCCAAAA CCGGTTGTTG GGGGTTCGAG TCCCTCCTGG 2390

CCTGCCATCT ACTAATTTAT TCTATCAAAT TTTTGTTTCA ATTGGATTGT TTTTGAATTT 2450

TTTAATTTTA GTTTAAGCTA TTTTGGATAA AATTGAAAAT TCTTTTAATG TATAAATATT 2510

AAGTTTAAGT GAGGGCGAAA AGAAACTATG GATAAATGGC TCATGCAATA TAAATTAGCT 2570

AGAGAAGAGC TTTCTAAAGT GATATTTCCT ATTAAGGAGC AGATACGCAA CGCGCTTGTT 2630

TCTGTTTTGG TGGTGGTGAG TGCTATCACG CTGTTTTTAG CTTTGTTGGA TTTTTCTCTG 2690

GGGGCTTTTA TCTCTAGTGT TCTATAGGTT GGTGGCTTTA AATAAGGAGA ATAATGATGG 2750

ATTGGTATGC CATACAAACT TATTCAGGGA GCGAGCAGTC CGTTAAGAAA GCGATTGAGA 2810

ATCTAGCGAA CGATC 2825

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2182 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: ATCC 43504

(v) FEATURES: from 344 to 520 bp signal peptide
from 521 to 1507 bp mature protein

(vi) PROPERTIES: 36/35/32 kD protein from *Helicobacter pylori*

```
GATCGCTCTT TGAGTGATTC CTGTATTCGC TTTATTGGCA AACTCTTCGC CAAACATTTT        60

CTTCACATTA GGGAAAATTA CCCCATCAAA AAACAAGTAG CCAATAAAAA TAATGGCGCA       120

CAATAAATGA ACAACCAACA CATAAGGATA AATCGCATCC ATTTAAAATC CTTTATTCAT       180

GGGAAAATTA AAGAGTTTTT AATCTACTAT AAAAGGGTTT TATTGTCAAG TATCCCACTA       240

TTATGGGAAT TTTAGGGGTG GTTTTTGTTT GACTTTTAAG ATTGCAATTA GCTATAATAA       300

AATAATTAAA AAAGTAACAC TTAAGCGGAG ACCCTAGAGA  GTG ATG CTC AAT TTT        355
                                                Met Leu Asn Phe

ATG ACA AAG AAG AAA AAT AGA ATG CAA GAT TGC AAA ATG GTT GGT AAA          403
Met Thr Lys Lys Lys Asn Arg Met Gln Asp Cys Lys Met Val Gly Lys
-55                 -50                 -45                 -40

AAT TTT AAT CGT AAG GAA TCT GTT TTG ATA GCT CAA TCT TTA GAA ATT          451
Asn Phe Asn Arg Lys Glu Ser Val Leu Ile Ala Gln Ser Leu Glu Ile
                -35                 -30                 -25

TCT AAA AAA GGC TCG GTA ATT TTA GGC GCT CTT TTG AGT TCG TTA TGG          499
```

```
Ser Lys Lys Gly Ser Val Ile Leu Gly Ala Leu Leu Ser Ser Leu Trp
            -20                 -15                 -10

CTG ACA AAC CCC TTA AAT GCC CAT GAA AAG AAT GGC GCG TTT GTG GGG      547
Leu Thr Asn Pro Leu Asn Ala His Glu Lys Asn Gly Ala Phe Val Gly
        -5                                  5

ATT AGC TTG GAA GTG GGT AGG GCT GAT CAA AAG ACC AAC GCT TAT AGA      595
Ile Ser Leu Glu Val Gly Arg Ala Asp Gln Lys Thr Asn Ala Tyr Arg
10                  15                  20                  25

AAC GGC GAG TTG TTT CAA GTG CCT TTT GGC GAT GTT TCA GCC AAT GAT      643
Asn Gly Glu Leu Phe Gln Val Pro Phe Gly Asp Val Ser Ala Asn Asp
                30                  35                  40

GAT GGC AAA GTC CCT AAC GGG CAG ACC GGT GGC TGT CAG CCA GCT TCA      691
Asp Gly Lys Val Pro Asn Gly Gln Thr Gly Gly Cys Gln Pro Ala Ser
            45                  50                  55

GGG ACG CCA GGA ACG CCA GGC TAT ACT AAA GCT AAT TGC GTG GTC AAT      739
Gly Thr Pro Gly Thr Pro Gly Tyr Thr Lys Ala Asn Cys Val Val Asn
        60                  65                  70

TGG ACT TCT CGC ACC ATG CTT AGC ACC AAT AAA AAC ATT CCT GGC CGT      787
Trp Thr Ser Arg Thr Met Leu Ser Thr Asn Lys Asn Ile Pro Gly Arg
    75                  80                  85

AAC CAG CCG ATG TAT GGG CTA GGT GTG ATG ACG GGC TAT AAG CAT TTT      853
Asn Gln Pro Met Tyr Gly Leu Gly Val Met Thr Gly Tyr Lys His Phe
 90                  95                  100                 105

ATC GGT AAA AAA AGG TGG TTT GGG TTG CGC TAT TAC GGC TTT TTT GAT      883
Ile Gly Lys Lys Arg Trp Phe Gly Leu Arg Tyr Tyr Gly Phe Phe Asp
                110                 115                 120

TAT GGG CAT ACC AAT TTC TCT AAC TCC AGG GCC GCT AAC GCT ATA TCG      931
Tyr Gly His Thr Asn Phe Ser Asn Ser Arg Ala Ala Asn Ala Ile Ser
            125                 130                 135

CCT TTC TAT TTG AGC GAT CAA AAA GCG GAC ATG TAT ACT TAT GGT TTT      979
Pro Phe Tyr Leu Ser Asp Gln Lys Ala Asp Met Tyr Thr Tyr Gly Phe
        140                 145                 150
```

```
GGC ACA GAC ATG CTT TTT AAC ATT ATA GAT AAG CCT AAA GCC ACG GCC        1027
Gly Thr Asp Met Leu Phe Asn Ile Ile Asp Lys Pro Lys Ala Thr Ala
    155                 160                 165

GGG TTT TTT GTG GGC GTG AAT TTT GCG GGT AAC ACT TGG ACC AAT AAT        1075
Gly Phe Phe Val Gly Val Asn Phe Ala Gly Asn Thr Trp Thr Asn Asn
170                 175                 180                 185

CGT GTG GGG TAT TTT AAG GAC GGG TAT GTT TAT GGC GTC AAT ACG GAT        1123
Arg Val Gly Tyr Phe Lys Asp Gly Tyr Val Tyr Gly Val Asn Thr Asp
                190                 195                 200

GCT GAC GCT TAC ATG ACT AAC GCT GAT GGC ACA ATC ACA TGC GGG GAC        1171
Ala Asp Ala Tyr Met Thr Asn Ala Asp Gly Thr Ile Thr Cys Gly Asp
            205                 210                 215

ACG ACG CCG GCG AGT TGT GAT GTG GGG ATT AAT CCT AAT AGC GTC TAT        1219
Thr Thr Pro Ala Ser Cys Asp Val Gly Ile Asn Pro Asn Ser Val Tyr
        220                 225                 230

ACC ACA GGA AAA TTG AAC GCT AAA GTG AAT CAC ACG ATT TTC CAA TTT        1267
Thr Thr Gly Lys Leu Asn Ala Lys Val Asn His Thr Ile Phe Gln Phe
    235                 240                 245

TTA GTG AAT GTG GGC ATT AGA ACT AAT ATT TTT GAA CAC CAT GGC ATT        1315
Leu Val Asn Val Gly Ile Arg Thr Asn Ile Phe Glu His His Gly Ile
250                 255                 260                 265

GAG TTT GGT ATC AAA ATC CCC ACG CTC CCT AAT TAC TTT TTC AAA GGC        1363
Glu Phe Gly Ile Lys Ile Pro Thr Leu Pro Asn Tyr Phe Phe Lys Gly
                270                 275                 280

TCT ACT ACC ATA AGA GCG AAA AAA CAA GGC CCG CTA GAG AAT GGC CAA        1411
Ser Thr Thr Ile Arg Ala Lys Lys Gln Gly Pro Leu Glu Asn Gly Gln
            285                 290                 295

CCA ACC ACT ATC ACC GGA GCA GAA ACC AAT TTC AGC TTA ACC CAA ACC        1459
Pro Thr Thr Ile Thr Gly Ala Glu Thr Asn Phe Ser Leu Thr Gln Thr
        300                 305                 310

TTA CGC CGT CAA TAT TCT ATG TAT TTG CGC TAT GTT TAT ACT TTT TGA        1507
Leu Arg Arg Gln Tyr Ser Met Tyr Leu Arg Tyr Val Tyr Thr Phe
    315                 320                 325
```

ATTTGGTAGG GTTTTTAGGC AGGGCTTATA GCTTATATAT GGATATATGA AAGCTTGATT 1567

TGTCAAGCTT TAGGGTTGTC ATTGAGTTGC AATAACTCTG TGCTGTTTTC TACTTTTTTG 1627

ATAAAATCAT TAATGGCATA ACAGCGTATG TTAATATTGT CTTTGAAATG GGCAAATCCC 1687

GCATATTCTT TGGCGTCATC ATGGATATTT GGAGCTACAA AAACACTAAA TTTTTCTCTA 1747

ATATCAGTGC TATTTTTAAT CAATTCTTTT AAATGTCTGG CAATAGGTAT CATTTCCAAG 1807

GTACTTTGAC TTCTATCTCT AATCAAGCTC ACTTCTATAT AACTTTGGGC TTTTGTGTCC 1867

ATAGCTACAA TATCAGGTTT GTTACCGCTT GCTGTGTATA CGGGCAAGCC TTCATCATCG 1927

CTTTTATAAT TGGGTATCAC GCTTAAATTT TCAAAATGTT GTTTCAAGAA AATAGCGCTT 1987

AAAAATTCTA AGCGTAAAGG TTTATCAATG AGTCTTAAAA AACTATCTTT TGATTCTTGC 2047

TTGTTGCAAG TAATGAGTAA TTCTTGCTTG ATAAAATCTT TAGTATAAGT GGTTGCTAGT 2107

TCATTCAATT TGCTTGTTTT AACGCTCTCA TCAGCGCTGA TTGGAGTAAC GCTAACAAGA 2167

AAGCTATCCA CGATC 2182

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 bp
(B) TYPE: nucleotide with deduced protein
(C) STANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULAR TYPE: Genomic DNA

(iii) ORIGIN OF ORGANISM: *Helicobacter pylori*
Direct experimental origin

(iv) NAME OF CELL LINE: Strain 26695

(v) FEATURES: from 843691 bp to 843212 bp mature protein

(vi) PROPERTIES: 17 kD protein from *Helicobacter pylori*

```
ATG GAA CAA AGC CAT CAA AAC TTG CAA TCT CAA TTT TTT ATA GAG CAT      843644
Met Glu Gln Ser His Gln Asn Leu Gln Ser Gln Phe Phe Ile Glu His
                 5                  10                  15

ATC TTA CAA ATT CTA CCT CAC CGC TAT CCC ATG CTT TTA GTG GAT AGA      843596
Ile Leu Gln Ile Leu Pro His Arg Tyr Pro Met Leu Leu Val Asp Arg
            20                  25                  30

ATT ATA GAG TTA CAA GCC AAT AAA AAA ATT GTC GCT TAT AAG AAT ATC      843548
Ile Ile Glu Leu Gln Ala Asn Lys Lys Ile Val Ala Tyr Lys Asn Ile
        35                  40                  45

ACT TTT AAT GAA GAC GTG TTT AAC GGG CAT TTC CCT AAT AAG CCC ATT      843500
Thr Phe Asn Glu Asp Val Phe Asn Gly His Phe Pro Asn Lys Pro Ile
    50                  55                  60

TTC CCG GGC GTT TTG ATC GTA GAG GGC ATG GCG CAA ACG GGA GGG TTT      843452
Phe Pro Gly Val Leu Ile Val Glu Gly Met Ala Gln Thr Gly Gly Phe
65                  70                  75                  80

TTA GCC TTC ACT AGC TTG TGG GGG TTT GAC CCT GAA ATC GCC AAA ACA      843404
Leu Ala Phe Thr Ser Leu Trp Gly Phe Asp Pro Glu Ile Ala Lys Thr
                85                  90                  95
```

```
AAA ATC GTG TAT TTC ATG ACG ATT GAT AAG GTT AAA TTC CGC ATC CCT      843356
Lys Ile Val Tyr Phe Met Thr Ile Asp Lys Val Lys Phe Arg Ile Pro
            100                 105                 110

GTA ACC CCA GGC GAC AGA TTA GAA TAC CAT TTA GAA GTC TTA AAG CAT      843308
Val Thr Pro Gly Asp Arg Leu Glu Tyr His Leu Glu Val Leu Lys His
        115                 120                 125

AAG GGC ATG ATC TGG CAA GTG GGT GGC ACG GCT CAA GTG GAT GGC AAA      843260
Lys Gly Met Ile Trp Gln Val Gly Gly Thr Ala Gln Val Asp Gly Lys
    130                 135                 140

GTG GTC GCT GAA GCC GAA TTG AAA GCC ATG ATT GCA GAG AGA GAT TAA      843212
Val Val Ala Glu Ala Glu Leu Lys Ala Met Ile Ala Glu Arg Asp
145                 150                 155
```

CLAIMS:

1. A protein from *Helicobacter pylori* (*H. pylori*) containing one of the peptide sequences selected from SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 and 19 according to Tables 1a-1c, or parts or homologues thereof having a minimum length of five amino acids.

2. A protein according to Claim 1, characterized in that the peptide sequences are N-terminal sequences.

3. A protein according to Claim 1 or 2, characterized in that the protein containing a peptide sequence having the SEQ ID NO: 1 according to Table 1a has a molecular weight of approx. 250 kD, the protein containing a peptide sequence having the SEQ ID NO: 2 according to Table 1a has a molecular weight of approx. 110 kD, the protein containing a peptide sequence having the SEQ ID NO: 3 according to Table 1a has a molecular weight of approx. 100 kD, the protein containing a peptide sequence having the SEQ ID NO: 6 according to Table 1a has a molecular weight of approx. 60 kD, the protein containing a peptide sequence having the SEQ ID NO: 10 according to Table 1b has a molecular weight of approx. 42 kD, the protein containing a peptide sequence having the SEQ ID NO: 11 according to Table 1b has a molecular weight of approx. 42 kD, the protein containing a peptide sequence having the SEQ ID NO: 12 according to Table 1b has a molecular weight of from approx. 32 to approx. 36 kD, the protein containing a peptide sequence having the SEQ ID NO: 14 according to Table 1c has a molecular weight of approx. 30 kD, the protein containing a peptide sequence having the SEQ ID NO: 15 according to Table 1c has a molecular weight of approx. 28 kD, the protein containing a peptide sequence having the SEQ ID NO: 16 according to Table 1c has a molecular weight of approx. 28 kD, the protein containing a peptide sequence

having the SEQ ID NO: 17 according to Table 1c has a molecular weight of approx. 25 kD, the protein containing a peptide sequence having the SEQ ID NO: 18 according to Table 1c has a molecular weight of approx. 25 kD, and the protein containing a peptide sequence having the SEQ ID NO: 19 according to Table 1c has a molecular weight of approx. 17 kD.

4. A protein according to any one of Claims 1 to 3, characterized in that the protein is a membrane protein or a protein which is firmly associated with the membrane.

5. A protein according to any one of Claims 1 to 4, characterized in that the protein is an integral membrane protein, in particular a Sarkosyl®-insoluble integral membrane protein.

6. A protein according to any one of Claims 1 to 5, which can be obtained in accordance with the following procedural steps:
(a) isolating the proteins by means of differential solubilization;
(b) separating the proteins, which have been isolated in accordance with step (a), by means of gel electrophoretic methods; and
(c) isolating the proteins, which have been separated in accordance with step (b).

7. A protein according to Claim 6, characterized in that the protein can be obtained by means of differential solubilization using Sarkosyl®.

8. A protein according to Claim 6 or 7, characterized in that it can be obtained by means of separation by one or more SDS polyacrylamide gel electrophoreses.

9. A protein according to Claim 8, characterized in that it can be obtained by means of several SDS polyacrylamide gel electrophoreses having different polyacrylamide contents.

10. A protein according to Claim 8 or 9, characterized in that the polyacrylamide content is approximately 8%, 10% or 16%.

11. A peptide having the amino acid sequence according to SEQ ID NO: 1, 2, 3, 6, 10, 11, 12, 14, 15, 16, 17, 18 or 19 according to Tables 1a-1c, or parts or homologues thereof having a minimum length of five amino acids.

12. An antibody against one or more proteins according to any one of Claims 1 to 10 and/or against one or more peptides according to Claim 11.

13. A polynucleotide encoding one or more proteins according to any one of Claims 1 to 10 or one or more peptides according to Claim 11.

14. A process for preparing the proteins according to any one of Claims 1 to 5, characterized in that the following procedural steps are carried out:
(a) isolating the proteins, by means of differential solubilization;
(b) separating the proteins, which have been isolated in accordance with step (a), by means of gel electrophoretic methods; and
(c) isolating the proteins, which have been separated in accordance with step (b).

15. A process according to Claim 14, characterized in that the proteins are isolated in accordance with step (a) using Sarkosyl®.

16. A process for preparing the peptides according to Claim 11, characterized in that a chemical peptide synthesis is carried out.

17. A process for preparing the proteins according to any one of Claims 1 to 10, or the peptides according to Claim 11, characterized in that a polynucleotide according to Claim 13 is expressed.

18. The use of one or more proteins according to any one of Claims 1 to 10, one or more peptides according to Claim 11, one or more antibodies according to Claim 12 or one or more polynucleotides according to Claim 13 for preparing a pharmaceutical composition or a diagnostic agent.

19. A pharmaceutical composition comprising one or more proteins according to any one of Claims 1 to 10 and/or one or more peptides according to Claim 11 or one or more antibodies according to Claim 12 or one or more polynucleotides according to Claim 13 or their expression products.

20. A pharmaceutical composition according to Claim 19, characterized in that the pharmaceutical composition is used as a vaccine.

21. A diagnostic agent comprising one or more proteins according to any one of Claims 1 to 10 and/or one or more peptides according to Claim 11, one or more antibodies according to Claim 12 or one or more polynucleotides according to Claim 13 or their expression products.

22. A protein from *H. pylori* containing one of the peptide sequences deduced from SEQ ID NO: 21, 22, 23, 24, 25, 26 and 27, or parts or homologues thereof having a minimum length of five amino acids.

23. A peptide having the amino acid sequence deduced from SEQ ID NO: 21, 22, 23, 24, 25, 26 or 27, or parts or homologues thereof having a minimum length of five amino acids.

24. A peptide selected from the C-terminal region of the peptide sequence of SEQ ID NO: 20 or homologue thereof.

25. A peptide according to Claim 24, wherein said peptide is selected from RDPKFNLAHIEKEFEVWNWDYRA and EKHQKMMKDMHGKDMHHTKKKK, or parts or homologues thereof.

26. An antibody against one or more proteins according to Claim 22 and/or against one or more peptides according to any one of Claims 23 to 25.

27. A polynucleotide encoding one or more proteins according to Claim 22 or one or more peptides according to any one of Claims 23 to 25.

28. A host cell transformed with the polynucleotide of Claim 13 or 27.

29. An expression product expressed from the host cell according to Claim 28.

30. A pharmaceutical composition comprising one or more proteins according to Claim 22 and/or one or more peptides according to any one of Claims 23 to 25, or one or more antibodies according to Claim 26, or one or more polynucleotides according to Claim 27 or one or more of their expression products.

31. A pharmaceutical composition according to Claim 30, characterized in that the pharmaceutical composition is used as a vaccine.

32. A pharmaceutical composition according to Claim 30 or 31, characterized in that when the pharmaceutical composition comprises a nucleotide sequence, said pharmaceutical composition is used as a DNA vaccine.

33. A diagnostic agent comprising one or more proteins according to Claim 22 and/or one or more peptides according to any one of Claims 23 to 25, or one or more antibodies according to Claim 26, or one or more polynucleotides according to Claim 27 or one or more of their expression products.

34. The use of one or more proteins according to Claim 22, one or more peptides according to any one of Claims 23 to 25, one or more antibodies according to Claim 26, one or more polynucleotides according to Claim 27 or one or more of their expression products as a pharmaceutical composition or as a diagnostic agent.