School of Biomedical Sciences

Proteome Comparison of *Helicobacter pylori* Isolates Associated with Four Disease Groups

Tracy McWilliams

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University of Technology

December 2006

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signed**…….**

Date **…….....**

ACKNOWLEDGEMENTS

Firstly, I wish to thank my supervisors Robert Lock and Warren Grubb. Their support and encouragement made this thesis possible.

Next, I would like to thank John Pearman, Jeff Coombs, Geoffrey Forbes and the staff of the Royal Perth Hospital for their assistance in the *H. pylori* experiments and in obtaining isolates for this study.

Finally, I would like to thank my family for their advice and encouragement throughout my PhD

INDEX

Contents **Page**

Contents **Page**

Contents **Page**

INDEX OF FIGURES

Contents **Page**

Contents **Page**

INDEX OF TABLES

Contents **Page**

ABBREVIATIONS

2D	Two dimensional
2-DE	Two dimensional electrophoresis
2D-PAGE	Second dimension polyacrylamide gel electrophoresis
ACTH	Adrenocorticotropic hormone
APAF	Australian Proteome Analysis Facility
APS	Ammonium persulphate
BPB	Bromophenol blue
Bis	N,N'-Methylene-bis-acrylamide
C	Crosslinker
CBB G250	Coomassie Brilliant Blue G250
CDM	Chemically defined medium
CHAPS	(3-[(3-Chloramidopropyl)dimethylammonio]-1-propanesulfonate
DTT	Dithiothreitol
DU	Duodenal ulceration
ESI-MS	Electrospray ionization mass spectrometry
GI	Gastritis
GU	Gastric ulceration
HCCA	Alpha-cyano-4-hydroxycinnamic acid
HCl	Hydrochloric acid
HDH	Homoserine dehydrogenase

IEF	Isoelectric focusing
IPG	Immobilised pH gradient
MALDI-TOF	Matrix assisted laser desorption/ionization-mass spectrometry time of flight
MSS	Multiple surfactant solution
M_r	Relative molecular mass
NUD	Non-ulcer dyspepsia
pI	Isoelectric point
PDA	Piperazine diacrylamide
PEG	Polyethylene glycol
Q4	*H. pylori* strain NCTC11637
RPH	Royal Perth Hospital
SB3-10	N-Decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate
SDS	Sodium dodecyl sulphate
T	Total polyacrylamide concentration
TBP	Tributyl phosphine
TEMED	N,N,N’,N’-Tetra-methyl-ethylenediamine
TFB	Trifluoroacetic acid
Tris	Tris(hydroxymethyl)- aminoethane
TSB	Tryptic soy broth
UCSF	University of California at San Francisco

Enzyme Commission Numbers

1.1.1.3	Homoserine dehydrogenase
1.1.1.41	Isocitrate dehydrogenase
2.7.4.3	Adenylate kinase
4.1.1.71	Subunit of 2-oxoglutarate oxidoreductase
4.1.2.13	Fructose bisphosphate aldolase
4.2.1.51	Prephenate dehydrogenase

DEFINITIONS

Agglomerative hierarchical cluster analysis A method for creating clusters in which each case starts out as a cluster. In the present instance, a case is equivalent to a *H. pylori* strain and a variable is equivalent to a spot. At every step, clusters are combined until all cases are members of a single cluster. Once a cluster is formed it cannot be split, it can only be combined with other clusters (Everitt B S 1993).

Agglomeration schedule The results of the cluster analysis are summarised in a listing of the cluster numbers (assigned ID number = cluster number) being combined at each stage in the cluster analysis (SPSS version Manual).

Arbitrary analysis sets are composed of any group of spots selected by the analysis program user (PDQUEST Manual).

Auto-matching is the process by which the PDQUEST program automatically matches spots in a number of gel images to those on the Standard Image. The program's algorithm matches spots between images by warping them to a common standard by reference to landmarks.

Between-groups clustering The distance between two clusters is defined as the average of the distances between all pairs of individuals that are made up of one individual from each group (Everitt B S 1993)

A **Boolean Analysis Set** is formed by combining two previously defined Analysis Sets (A and B) using Boolean operators. The Boolean Sets created might include 1) the intersection of A and B; 2) the union of A and B; 3) spots that are unique to set A plus spots that are unique to set B; 4) Spots that are only found in B; 5) spots that are only found in A.

Distance Measure allows a specific distance or similarity measure to be used in clustering. SPSS does not plot actual distances but re-scales them to numbers between 0 and 25. Lines, which are the measure of relatedness on a diagram, join the cases indicate a clustering. Distance can be calculated in two different forms Interval and Binary. Interval measures compare the quantitative values for variables (i.e. the volumes for spots) between cases (i.e. isolates) and binary measures compare the qualitative values for the variables (i.e. presence/absence only) (SPSS version 11 manual).

Euclidean distance (interval measure) is a dissimilarity measure for continuous data. The distance between two items is the square root of the sum of the squared differences in values for each variable (SPSS version 11 Manual).

Gaussian Spot PDQUEST uses Gaussian modelling to create "idealised" spot boundaries. A Gaussian spot is a precise three-dimensional representation of a scanned spot. Gaussian curves are fitted to the scanned spot in the X, Y and Z dimensions, where X and Y represent co-ordinates on the 2-D gel and the Z axis represents spot volume. Additional modelling is then performed to create the final Gaussian spot (PDQUEST Manual).

A **Gel Image File** is a fully processed image file of a gel. This file is a duplicate of the original scan and is utilised in the matchset so it can be edited without losing any of the information contained in the original scan image (PDQUEST Manual).

Gel Spot File contains a synthetic image with Gaussian representation of the spots in the Gel Image File (PDQUEST Manual). At least 6 successive Gaussian spot approximations are required to generate the Gel Spot File.

In a **hierarchical** classification, the data are not partitioned into a particular number of classes or clusters at a single step. Instead the classification consists of a series of partitions, which may run from a single cluster containing all individuals, to *n* clusters

each containing a single individual. Hierarchical clustering techniques may be subdivided into agglomerative methods that proceed by a series of successive fusion of the *n* individuals into groups, and divisive methods, which separate the *n* individuals successively into finer groupings (Everitt B S 1993).

Image Warping is a function that deforms images; all positions in one image plane are mapped to positions in another plane. It is used to bring two or more images into alignment (Gustafsson et al 2002).

K-means Clustering attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. However, the algorithm requires the number of clusters to be specified. The number of clusters can only be specified if this information is known (SPSS Manual).

Landmark spots set reference positions used to compensate for slight positional differences of corresponding spots on different images due to gel distortion. These landmark spots are used by the program to align and position matchset gel images for matching. Spots may be selected as landmark spots if they are well resolved and present in all members of the matchset in corresponding locations. They are particularly valuable if selected at the edges of the image, if isolated from other spots or if located in or near regions of high spot concentrations. A minimum of 20 landmarks were distributed throughout each Standard image.

Manual matching is used if a legitimate spot is present in corresponding positions on one or more member images but is not matched to the Standard Image by automatic matching. The spot is selected in the Standard Image and its corresponding position is highlighted in all the image members of the matchset. If this spot is also located in other images, it may then be selected there and matched.

Matching is the process by which the analytical software program identifies spots in one or more gel images that are found in identical locations to some of those on the Standard

image. This identification may be performed by auto-matching or manual matching of the protein spots.

A **Matchset** is a set of gel images that has been grouped together for the purpose of qualitative and quantitative comparison. Spots on one image can be compared to spots on every other image in the matchset.

There are two types of matchsets 1) level one matchset (see Figure 2.2A) and 2) level two matchset (see Figure 2.2B). A **level one matchset** is created directly from gel image files. It may contain some or all of the gel images of a given experiment. A **level two matchset** is created from the Standard images of a number of level one matchsets.

MS/MS Sequencing data Product ion MS/MS sequence data from one or more peptide – MS/MS mode (Westermeier R and Naven T 2002).

Normalisation Before gels are compared for protein spot differences, the spot volumes of the different gel image files included in the matchset have to be adjusted by normalisation. Normalisation corrects for systematic differences due to variable protein loads and staining effectiveness.

Pattern difference (binary measure) is a dissimilarity measure for binary data that ranges from 0 to 1. Computed from a fourfold table as bc/(n2), where b and c represent the diagonal cells corresponding to cases present on one item but absent on the other and n is the total number of observations (SPSS version 11 Manual).

Peptide mass fingerprinting (PMF) is a technique for searching protein databases for protein identity. The subject protein is cleaved, and the masses of the resultant peptide are used for a database search (Westermeier R and Naven T 2002).

Peptide mass fingerprint and composition information. The molecular mass of each of the peptides derived from the enzyme digestion or chemical cleavage of the protein

can be used alongside some composition information relating to one or more of the peptides (Westermeier R and Naven T 2002).

Peptide mass fingerprint and sequence information. The molecular mass of each peptide derived from enzyme digestion or chemical cleavage of the protein can be used along side some direct sequence information relating to more of the peptides (Westermeier R and Naven T 2002).

Positionally Conserved refers to protein spots that are found in corresponding positions for all isolates included in the study. Positionally conserved spots consists of two subsets 1) spots which vary by less than a factor of three or more (quantitatively conserved) and 2) spots which vary in volume by a factor of three or more.

Protocol For the purposes of the present study the word **"**protocol" refers a set of specific electrophoresis conditions that were used to separate proteins in a particular M_r and pI range.

A **Quantitative Analysis Set** contains spots from two different images or replicate groups that are present in both images or replicate groups but differ in concentration to a predetermined degree (usually by a factor of 3 or more) i.e. it includes spots whose volume has significantly increased or decreased.

A **Qualitative Analysis Set** contains spots from two different images or replicate groups that were detected in one image or replicate group, but not in the other. Such a set might contain, for example, proteins that were expressed under experimental conditions but not expressed under control conditions, or visa versa.

A **Reference image** for a matchset is the gel image chosen as the best representative of all the images in that matchset, i.e. it contains the greatest number of well-resolved spots.

Replicate Groups A replicate group is a set of gels that is prepared from the same protein sample using the same protocol where a **protocol** refers to the specific electrophoresis conditions that were used to separate proteins in a particular M_r and pI range.

There are two reasons for the formation of replicate groups. The first is to determine the validity of spots. In order to be accepted as valid, this study required that, a particular spot must be observable in at least three out four gel images for each replicate group. In addition, slight variations in quantitation values need to adjusted, due to variables such as loading and staining. The volume of a given spot is best estimated by averaging results from several gel images.

Serial charge trains are protein isoforms that appear as a train of spots, usually horizontal, that differ in pI and/or M_r. These may indicate post-translational modifications such as glycosylation or phosphorylation.

Simple matching (binary measure) is the ratio of matches to the total number of values. Equal mass is given to matches and non-matches (SPSS version 11 Manual).

Single cluster solution Displays cluster membership for a single cluster solution with a specified number of clusters.

Spot Volume is a measure of protein quantity related to staining intensity. The spot volume is determined from the average spot intensity (mean density of pixels in the spot) by the spot area (number of pixels in a spot).

Squared Euclidean distance (binary measure) The binary squared Euclidean dissimilarity measure. Computed from a fourfold table as SQRT(b+c), where b and c represent cases present on one item but absent on the other its minimum value is 0 (absent spots), and it has no upper limit (present spots) (SPSS version 11 Manual).

Squared Euclidean distance (interval measure) A measure of distance between pairs of cases. The distance between two cases is the sum of the squared differences between the variables of the cases. This dissimilarity measure is used for continuous data (SPSS version 11 Manual).

	SPOT NUMBER 1	**SPOT NUMBER 2**
GU1	**100**	**90**
NUD1	**280**	**150**

In the above example [values are normalised spot volume (units determined by computer algorithm)] the squared Euclidean distance is $(280\text{-}100)^2 + (150\text{-}90)^2$ which equals $180^2 + 60^2$ which equals 36000. A disadvantage with this measure is that when variables are measured on different scales variables that are measured in larger numbers will contribute more to the computed distance than variables measured in smaller numbers.

A **Standard image** is a composite image that contains all the spots present in all the gel images in a particular matchset. The starting point for generating a Standard image is to enter all the spots detected into the Reference image.

A **Statistical User Set is** composed of spots from replicate groups whose volumes are found to be significant according to a specified statistical test (e.g. Student T-Test, or Mann-Witney Test).

Abstract

The Gram-negative bacterium *Helicobacter pylori* is found in human gastric mucosa. *H. pylori*, one of the most common chronic bacterial infections of humans, is present in almost half of the world population. It is associated with chronic gastritis, non-ulcer dyspepsia, gastric and duodenal ulcers, and malignant neoplasms. The aim of this study was to detect microbial candidate protein markers whose presence might be correlated with the development of four different clinical consequences of *H. pylori* infection, gastric ulceration [GU], duodenal ulceration [DU], non-ulcer dyspepsia [NUD] and gastritis [GI]. Eleven *H. pylori* isolates associated with these outcomes were analysed. The total complement of protein from these *H. pylori* isolates were resolved by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and compared using PDQUEST pattern analysis software. Relationships between the isolates associated with specific disease outcomes were determined by cluster analysis. Fifty six disease specific proteins were then characterised by tryptic peptide-mass fingerprinting using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Up to 1165 protein species were resolved from each *H. pylori* strain. Proteome analysis revealed that only 470 (40%) of the proteins detected were common to all eleven isolates. Twenty six of the 56 disease specific proteins that were selected for identification consisted of spots whose expression is altered in response to stress conditions or those that can affect *H. pylori* cell division and the cell membrane. The remaining 30 proteins had no known function. This study has provided further confirmation of the extensive variation that the bacterium *H. pylori* exhibits at the proteome level. Most significantly this study has found, through the application of cluster analysis and protein matching, that isolates do form disease groups. Comparative proteome analysis is a useful method for highlighting the extensive strain variation that *H. pylori* exhibits and to determine if any disease specific proteins exist.

1.1 *HELICOBACTER PYLORI*

1.1.1 The Discovery of H. pylori

In 1979 Robin Warren, a pathologist in Perth, Western Australia, noticed that curved bacteria were often present in gastric biopsy specimens submitted for histological examination from patients with antral gastritis (Warren, 1983). However, because these organisms were not present within the gastric mucosa but were present in the mucous layer overlying the tissue, they were thought to be irrelevant. Warren found that similar organisms had been described by European pathologists in the late 19th century but, because attempts to culture them had failed, they had been ignored and forgotten by generations of scientists (Fung, 1995).

In 1981, Barry Marshall, a medical registrar working in gastroenterology collaborated with Warren. They reviewed the patients in whom large numbers of gastric spiral bacteria had been seen. One of these patients had been treated with tetracycline; his symptoms resolved and the subsequent endoscopic biopsy showed that the antral gastritis had also resolved (Marshall and Warren, 1984). In addition, they were finally able to visualise organisms in gastric biopsies from high-magnification electron micrographs.

Since the organisms had the appearance of curved, Gram-negative rods, the investigators used methods for the isolation of *Campylobacter* species, which involved inoculating the biopsy specimens onto blood agar and incubating the cultures under microaerobic conditions. The project to culture bacteria from gastric biopsy specimens began in March 1982, with the microbiologist John Pearman supervising (Goodwin and Worsley, 1993a). Gastric specimens from 100 consecutive patients being gastroscoped were examined by direct microscopy and cultured. Among the first 34 cultures, spiral bacteria were seen in six. However, in spite of frequent variations in media and incubation temperatures, they could not be cultured. Since most campylobacters grow within 48 hours under such conditions, plates without visible growth were normally discarded within 3 days (Goodwin and Worsley, 1993a). However, when the 35th culture, by chance, was incubated for 5 days over an Easter holiday, colonies were seen. The date was the 14th April 1982 (Marshall and Warren, 1984). Following this discovery, and taking into account the

length of time required to culture this organism, it was isolated from 11 patients; it was then characterised and called *Campylobacter pyloridis* (later revised to *Campylobacter pylori)* (Marshall and Warren, 1984; Goodwin and Worsley, 1993b; Goodwin, 1994; Anonymous, 1989). The two earliest isolates are now designated NCTC11637 and NCTC11638.

However, based on DNA hybridisation and base composition, ultrastructural studies, and analyses of cellular fatty acids, respiratory quinones, enzyme profiles and growth requirements, the bacterium was transferred in 1989 to a new genus *Helicobacter* and renamed *Helicobacter pylori*. The *Helicobacter* genus now consists of more than 20 species including non-gastric *Helicobacter* species such as *H. pullorum* (Stanley et al, 1994).

Following publication of a report in the Lancet in 1984 (Marshall and Warren, 1984), investigators all over the world rapidly confirmed the presence of these organisms in the gastric mucus (Jones et al, 1984; McNulty and Watson, 1984). By 1984, it was becoming clear that *H. pylori* infection was a major culprit in many gastrointestinal disorders, although this conclusion would still generate controversy for some years (Blaser, 1990).

In 1985 Marshall et al postulated "that *Campylobacter pyloridis* could colonise histologically normal mucosa and induce gastritis". It was shown that "ingested *Campylobacter pyloridis* was able to colonise normal gastric mucosa. That this "colonisation is associated with acute inflammatory changes; polymorphonuclear neutrophil leucocytes infiltration and exudation from the mucosa; mucus depletion; and reversible epithelial cell damage. These changes may now be referred to as being characteristic of acute *Campylobacter pyloridis* gastritis" (Marshall et al, 1985). Marshall also demonstrated how in individuals who are unable to clear the infection chronic gastritis may result.

Prior to the isolation of *H. pylori* the accepted paradigm was "no acid, no ulcers". The discovery that *H. pylori* was the culprit in many gastrointestinal disorders has revolutionised the view on the diseases associated with the gastric environment

(Goodwin and Worsley, 1993a). This major breakthrough has completely changed how some important gastrointestinal disorders are treated.

1.1.2 Phenotypic Characteristics

H. pylori is a spiral, Gram-negative bacterium that has blunt rounded ends in gastric biopsy specimens (Goodwin and Worsley, 1993b). However, when the bacteria are cultured on solid media, they assume a rod-like shape. *H. pylori* organisms are 2.5 to 5.0 μm long and 0.5 to 1.0 μm wide (Goodwin and Armstrong, 1990; Goodwin and Worsley, 1993b). *H. pylori* is motile and usually possess 4-6 unipolar-sheathed flagella that are essential for motility. Each flagellum is approximately 30 μm long and approximately 2.5 nm thick (Geis et al, 1993; Goodwin and Armstrong, 1990) (Figure 1.1). After prolonged culture (4-5 days) on solid or in liquid medium, the bacteria may form spheroid or coccoid bodies. A coccoid microorganism is shaped like or resembles a sphere. The coccoid forms initially appear as U-shaped bacilli with the ends of two arms joined by a membranous structure (Worku et al, 1999) (Figure 1.2). The coccoidal form is metabolically active but is non-culturable (Bode et al, 1993).

Over the period of its evolution, *H. pylori* has become well suited to its environmental niche, the human stomach and duodenum. Currently, no other host is known except for primates such as macaque and baboon (Goodwin and Worsley, 1993b). *H. pylori* is found in gastric juice and under the mucous layer at the surface of gastric epithelial cells. In order to survive in the highly acidic environment of the stomach, the *Helicobacter* genus has evolved a distinctive adaptation, the high expression of the enzyme urease (Goodwin and Worsley, 1993a and 1993b). Urea diffuses freely from plasma into gastric juice. *H. pylori* produces large amounts of nickel-containing urease which degrades each molecule of urea to produce one carbon dioxide and two ammonia molecules, leading to the net production of alkali to neutralise stomach acid in the *H. pylori's* local environment (Hu and Mobley, 1993). Without urea, *H. pylori* is intolerant of acid, but in its presence the bacterium prefers mildly acidic conditions and can withstand a pH as low as 1.5 (Marshall et al, 1990).

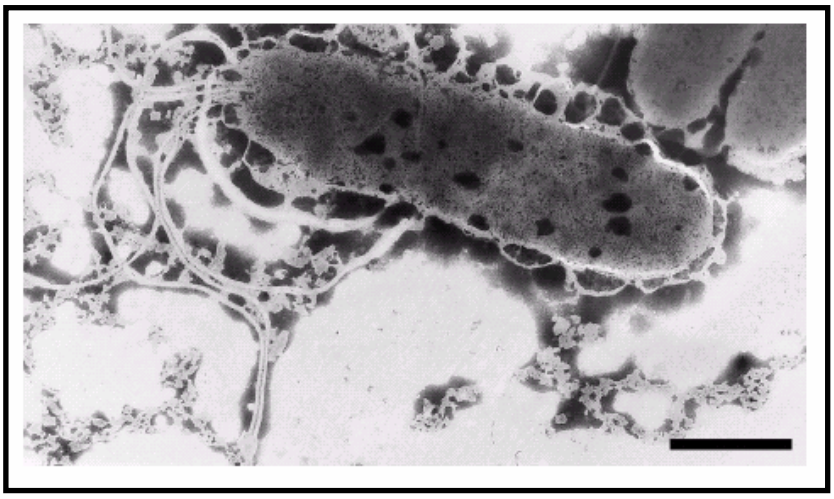


Figure 1.1 ***Helicobacter pylori***

A transmission electron micrograph (negative staining) of a *H. pylori* bacterium. Note the characteristic bundle of sheathed flagellae attached to one pole of the cell. Scale bar, 0.5 μm. (Suerbaum and Josenhans, 1999)

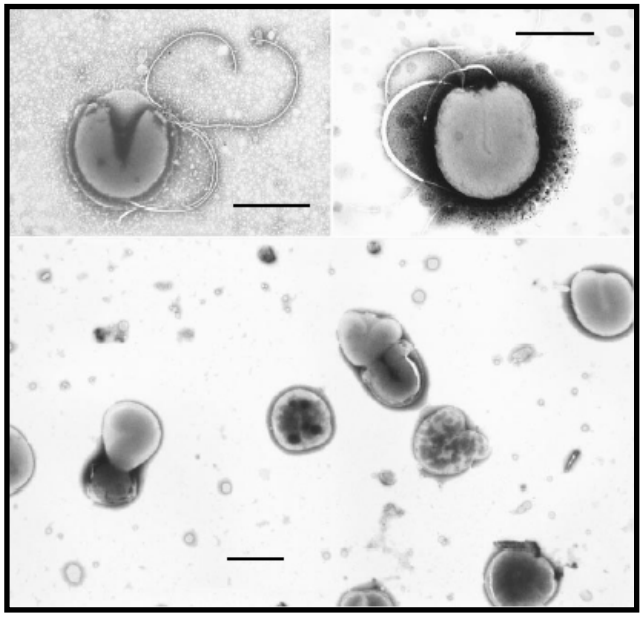


Figure 1.2 *H. pylori* coccoidal form

Electron micrographs of negatively-stained preparations of *H. pylori*, illustrating early (top left) and late (top right) decline phase of growth in the transformation to coccoidal form (bottom). Precoccoidal forms (top left, top right) typically have flagellae which are absent in coccoidal forms. Bars 500 nm (Worku et al, 1999)

H. pylori is microaerophilic and possesses a respiratory type of metabolism. It is a fastidious microorganism that requires nutrient-rich media, an atmosphere enriched in CO_2 (5-12%), and high humidity (96-100 %) and prefers a pH near 7.0 (Velazquez and Feirtag, 1999). It is slow growing, requiring a minimum of 3-5 days incubation time. Culture media are usually supplemented with blood as a reducing agent rather than as a nutrient (Cohen, 1995) and shaking is needed in liquid systems to provide good dispersion of gases throughout the liquid (Sjostrom and Larsson, 1996). Cells grow at 30 and 37^{o}C but not at 25^{o}C and the growth at 42^{o}C is variable (Goodwin and Armstrong, 1990). Atmospheric hydrogen (as much as 5-10%) is either required or stimulates the growth of these organisms. *H. pylori* grows poorly, if at all, in routine aerobic or anaerobic atmospheres (Tominaga et al, 1999).

1.1.3 Geographical Epidemiology of H. pylori

H. pylori is probably the most common chronic bacterial infection of humans, present in more than half of the world population (Figure 1.3). The pattern of *H. pylori* infection in developing countries is different from that seen in developed countries.

Developing Countries

A high level of seroprevalence in the general population has been reported in large studies of developing nations (most percentage figures are available from studies performed prior to 2000). For example the levels in Africa are: 85% in Nigeria (Holcombe et al, 1991; Holcombe et al, 1992), 79% in Zaire (Glupczynski et al, 1992), 79% in Algeria and 71% on the Ivory coast (Teh et al, 1994); In the Far East countries: 75% in Vietnam (Megraud et al, 1989) and in Thailand (Perez-Perez et al, 1990).

Acquisition occurs in about 10% of children every year between the ages of two and eight, with most people infected by their teens (Bardhan, 1997). Increased prevalence in developing countries is believed to be due to overcrowding, poor hygiene and inadequate sanitation. In developing countries, most adults are infected (Calam, 1996).

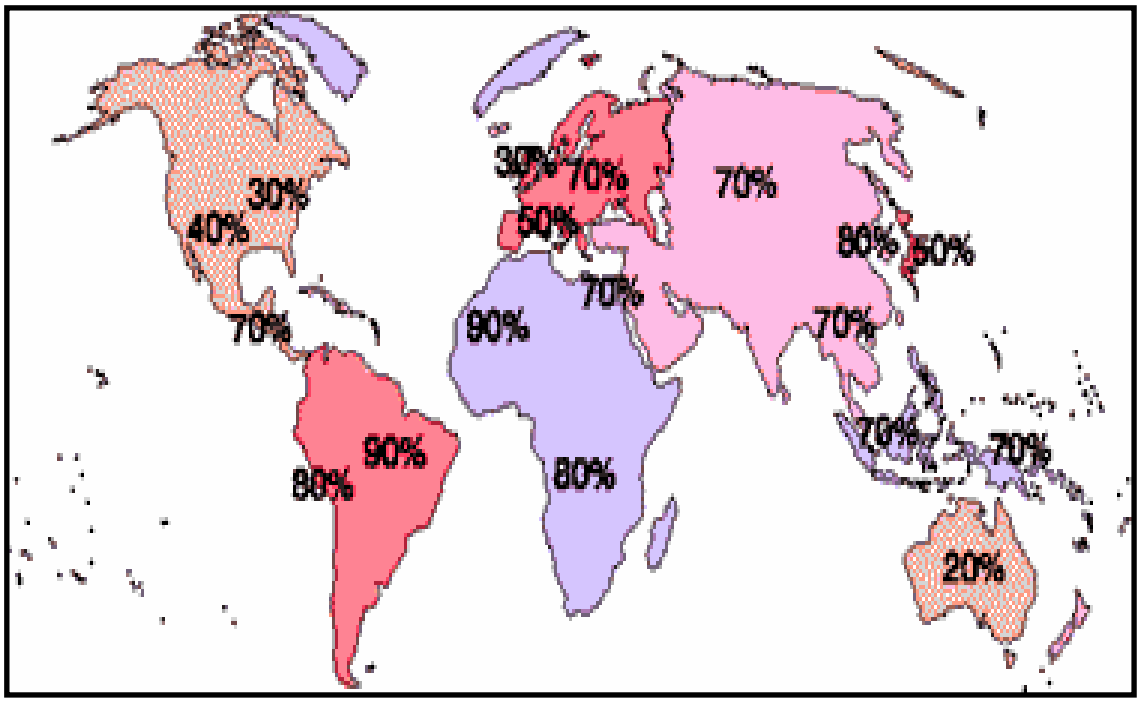


Figure 1.3 Worldwide frequency rates

Average frequency rates for *H. pylori* worldwide in the year 2000 (figure obtained from The Helicobacter Foundation website http://www.helico.com/)

Developed Countries

In the majority of Western countries, the prevalence is lower. For example, Australia, 20%, USA 40%, UK, 30%, Spain, 53% (Martin-de-Argila et al, 1996), Germany, 37% (Ito et al, 2002) and Greece, 49.2% (Apostolopoulos et al, 2002). The exception is Singapore, 70% (Ito et al, 2002).

The prevalence of *H. pylori* infection increases with the age of the host. In the Western world, the percentage of the population that is infected rises gradually from infancy to an age of about 60 years. Infection is uncommon in young children, but present in about 20% of persons under the age of 40 years and about 50% of those over the age of 60 years (Graham et al, 1991a). In Western countries, *H. pylori* infection is associated with low socioeconomic status (Fiedorek et al, 1991; Graham et al, 1991b). Ethnic minority groups such African-Americans and Hispanics, and immigrants from developing countries are believed to be responsible for isolated areas of high prevalence in some Western countries (Replogle et al, 1995). There is no significant difference in *H. pylori* prevalence in relation to gender.

A wide body of evidence now also indicates that once acquired, *H. pylori* usually persists for life, unless eradicated by antimicrobial therapy (Blaser, 1996; Blaser 1997)

1.1.4 Association of H. pylori *with Illness*

Severe pathological consequences of *H. pylori* infection develop in approximately 10% of those infected, and infection is associated with a variety of disease outcomes (Dunn et al, 1997). These outcomes include gastritis, non-ulcer dyspepsia, iron deficiency anaemia, gastric and duodenal ulcer disease, the development of gastric mucosal atrophy (presumed to be a precancerous condition), gastric carcinoma, malignant tumours of the mucosa-associated lymphoid tissue (MALT lymphoma), and an increased susceptibility to other infectious diseases (Dunn et al, 1997). *H. pylori* infections are associated with very high morbidity and mortality, and they impose a major burden upon health care systems worldwide (Correa, 2003; Ofman et al, 1997).

Gastritis (GI)

Gastritis is defined as an inflammatory reaction in the gastric mucosa and can be associated with glandular atrophy and alterations in the differentiation and function of the epithelial cells (Goodwin and Worsley, 1993a). All *H. pylori* infected patients develop chronic gastric inflammation but perhaps surprisingly, this condition is usually asymptomatic (Calam, 1996). This condition is nevertheless important because individuals with it are quite strongly predisposed to more serious disease outcomes (ulcers, adenocarcinoma or lymphoma).

Gastritis in patients who are *H. pylori* negative is rare and it is typically antral-sparing or Type A gastritis associated with an autoimmune disease known as Addisonian pernicious anaemia.

Gastritis frequently progresses from simple inflammation to atrophy (atrophic gastritis) which indicates a loss of normal gastric glands (Sitas et al, 1993; Asaka et al, 1992). This is followed by a growth of new metaplastic (i.e. cells transformed to an abnormal state with evidence of DNA damage) epithelium and glands, which are less differentiated than normal. Chronic gastritis follows a stepwise, slow progression over many years from simple inflammation to more severe and extensive atrophy and metaplasia. Without treatment, gastritis is a long-standing and probably life-long condition (Goodwin and Worsley, 1993a).

Although many patients with *H. pylori* infection develop gastritis, not all patients have the same disease outcome; therefore, it is believed that environmental and microbial or host genetic factors may play a role in the development of disease (Goodwin, 1994).

Duodenitis (DU)

Duodenitis is characterised by changes in the architecture and villous structures of the duodenal mucosa, and alterations in morphology and differentiation of the duodenal epithelium (Goodwin and Worsley, 1993a). In duodenal ulcer disease, duodenitis is limited nearly entirely to the duodenal bulb and is usually very patchy. Inflammatory duodenitis develops in connection with *H. pylori* gastritis. The

progress of acid-induced metaplasia to inflammation is probably the result of the extension of the *H. pylori* infection from the stomach into the bulb (Goodwin and Worsley, 1993a).

Inflammation in the duodenal bulb occurs in other diseases that include Zollinger-Ellinson syndrome, duodenal Crohn's disease, ulcerative jejunitis and pancreatic disorders (Goodwin and Worsley, 1993b).

Non-ulcer dyspepsia (NUD)

Non-ulcer dyspepsia is a broad diagnostic term referring to a complex of symptoms that includes epigastric pain or discomfort which may or may not be related to food intake. Four different syndromes are generally recognised (Thompson and Heaton, 1980):

- ulcer-like dyspepsia which involves classical ulcer-like symptoms, but no ulcer;
- dysmotility-like dyspepsia which involves dyspeptic symptoms which suggest gastric stasis or dysmotility of the small bowel;
- reflux-like dyspepsia which involves symptoms suggestive of gastro-oesophageal reflux; and
- unspecified dyspepsia which is a category reserved for dyspeptics who do not fall into any of the above groups (Calam, 1996).

Symptoms of NUD may be chronic, recurrent, or new in onset and can usually be distinguished from typical gastroesophageal reflux, the irritable bowel syndrome, or biliary disease. This epigastric discomfort may initially be attributed to an ulcer, which is subsequently found not to be present (Goodwin, 1994).

Non-ulcer dyspepsia occurs in 25-40% of the population worldwide. *H. pylori* has been found in 40-60% of patients with non-ulcer dyspepsia (Pantoflickova and Blum, 2001). Although the link between *H. pylori* and peptic ulcer disease and gastric cancer has been accepted for a number of years, there continues to be controversy about the role of this organism in NUD (Pantoflickova and Blum, 2001; McColl, 2000).

Recent studies have indicated that improvement of NUD symptoms following eradication of *H. pylori* (57% recovery rate) is not as significant as the recovery rates of other disease outcomes (Azuma et al, 2001; Talley et al, 1999). It is believed that other unknown factors may be responsible for the symptoms of NUD.

Iron deficiency anaemia

In one study of 30 patients it was found that 91.7% of the patients cured of *H. pylori* subsequently recovered from anaemia and had increased serum ferritin levels (Barabino, 2002). Iron is crucial for *H. pylori* growth. The precise pathological mechanism of the *H. pylori* induced anaemia is still lacking but one possible explanation is that iron is diverted to the *H. pylori*-infected antrum of the stomach. It is not known why iron deficiency anaemia does not develop in all infected individuals (Barabino, 2002).

Gastric and duodenal ulcers (GU and DU respectively).

Ulcers are defined as breaches in the epithelium, which are associated with acute and chronic inflammation. The prevalence of *H. pylori* infection in patients with gastric ulcer disease is reported to be nearly 80% (Labenz and Borsch, 1994) and the rate is nearly 100% (Park et al, 1993) in patients with chronic duodenal ulcer disease. Patients with gastric ulcers typically have atrophic gastritis and corpus-predominant gastritis. Patients with duodenal ulcers have few atrophic changes and have antrum predominant gastritis (Goodwin, 1994).

Eradication of *H. pylori* drastically reduces recurrence of duodenal and gastric ulcers**.** Duodenal ulcers in adults with no evidence of *H. pylori* infection are generally due to Zollinger-Ellison syndrome, Crohn's disease, or cancer (McColl, 2000). About 20% of gastric ulcers are not caused by *H. pylori,* but are due to the corrosive effect of non-steroidal anti-inflammatory medications often taken for arthritis (Dixon, 2000).

Gastric cancers

Gastric adenocarcinomas are often associated with *H. pylori* infection. This is probably related to the observation that *H. pylori* acquisition leads to chronic

gastritis, which may result in atrophy of the gastric epithelium (atrophic gastritis) (Calam, 1996). The clinical importance of atrophic gastritis is related to the fact that it significantly increases the risk of gastric cancer development (Nomura et al, 1991). Most gastric cancers occur in mucosa with atrophic changes, and patients with atrophic gastritis are estimated to have a five- to nine-fold increased risk for developing gastric cancer (Parsonnet et al, 1991).

It has been found that gastric cancers often develop in patients with non-ulcer dyspepsia, active gastric ulcers, and hyperplastic gastric polyps, but gastric cancers rarely develop during the follow-up in patients with active duodenal ulcers (Uemura et al, 2001). The significance of the negative association with duodenal ulcers is not yet known.

Gastric cancer develops in five percent of *H. pylori*-positive persons over a 10 year period. Incidence of gastric cancer is higher in developing countries (Uemura et al, 2001). However, in the developed countries, gastric cancer is receding, probably because of the decreasing prevalence of *H. pylori* (Calam, 1996). It is entirely possible that if *H. pylori* could be eliminated from the developing countries, through the application of antibiotics and improvements in hygiene and sanitation, that the incidence of gastric cancer would approximate that seen in the developed countries.

Mucosa-associated lymphoid tissue (MALT) lymphoma

Lymphoma is defined as any neoplasm (tumour) of the lymphoid tissue, whether benign or malignant. Lymphoid tissue is reticular connective tissue that houses macrophages and a continuously changing population of lymphocytes (Marieb, 1995).

Several studies have found that the most common aetiology of MALT lymphomas in the stomach is *H. pylori* and chronic gastritis, i.e. in approximately 90% of cases (Wotherspoon et al, 1993; Hussell et al, 1993). It is believed that *H. pylori* may provide the antigenic stimulus for sustaining the growth of lymphoma in the stomach. However, the exact mechanism of the transition from *H. pylori* infection to low-grade lymphoma is still unclear.

Following the initial study by Wotherspoon et al (1993), several groups have shown that eradication of *H. pylori* with antibiotics results in regression of low-grade gastric MALT lymphoma in 75% of cases (Wotherspoon et al, 1993; Hussell et al, 1993). However, the elimination of *H. pylori* does not lead to the eradication of high-grade MALT lymphomas (Wotherspoon et al, 1993). This discovery was a major breakthrough in the treatment of low-grade MALT lymphomas.

1.1.5 The H. pylori Genome

Studies of *H. pylori* DNA by polymerase chain reaction (PCR), pulsed-field gel electrophoresis and random amplification of polymorphic DNA (RAPD) have suggested that the species is extremely genetically diverse (Jiang et al, 1996). The use of RAPD-PCR and DNA fingerprinting showed that strains from unrelated infected patients had unique fingerprints whereas strains isolated from family members had very similar, although not identical patterns (van der Ende et al, 1996). These results implied that differences observed between the strains infecting individual family members occurred after the primary infection and suggested that *H. pylori* evolution is rapid, resulting in a genome so highly fluid that it produces highly diverse strains (Jiang et al, 1996).

The recent availability of the genomic sequence for two unrelated *H. pylori* isolates, J99 and 26695 (see www.tigr.org.), has enabled a detailed analysis of the overall level of genetic diversity that had been previously suggested for this organism. *H. pylori* is the first bacterial species for which two complete genome sequences from two unrelated strains was completed.

The complete genome sequence of *H. pylori* strain 26695 was first published in Nature in August 1997 (Tomb et al, 1997). This strain has a circular chromosome of 1,667,867 base pairs and 1,590 predicted coding sequences. *H. pylori* strain 26695 was originally isolated in the early 1980's in the United Kingdom from a patient with gastritis of 'unknown severity' (Tomb et al, 1997). This strain had been passaged extensively in the laboratory prior to its sequence being completed in 1997 (Alm et al, 1999).

The complete genome sequence of strain J99 was released in 1999 (Alm et al, 1999). J99 has a circular chromosome of 1,643,831 base pairs, which is 24,036 base pairs smaller than the 26695 chromosome and has 1,490 predicted coding sequences. *H. pylori* J99 was originally isolated in the United States in 1994 from a patient diagnosed with a duodenal ulcer and had been subjected to minimal passage in the laboratory before genomic sequencing began in 1996 (Alm et al, 1999).

The availability of the two completed genome sequences from two unrelated strains has allowed a precise analysis of their genetic differences. Although the two chromosomes are organised differently in a limited number of discrete regions, the genome size and gene order of these two *H. pylori* isolates was found to be very similar (Alm and Trust, 1999). The regions of organisational differences are associated with insertion sequences, DNA restriction/modification genes, repeat sequences, or a combination of the above. Genes that were unique to only one of the strains comprised between 6% and 7% of the total coding capability of each strain (Alm et al, 1999; Alm and Trust, 1999). Half of the genes unique to each strain were contained within a single hypervariable region of the chromosome called the plasticity zone. A significant level of variation at the nucleotide level was seen across the genome. Alm et al (1999) suggests that this provides an explanation for why the nucleotide-based typing techniques seem to indicate a high degree of apparent genetic diversity among independent *H. pylori* isolates. This nucleotide variation together with the organisational rearrangements is believed to have provided an over-estimation of the gene diversity of *H. pylori* as assessed by pulsed-field gel electrophoresis (Alm and Trust, 1999). As with other bacterial genomes sequenced to date, only approximately 60% of the genes identified encoded a product that can presently be assigned a function (Alm et al, 1999). One-half of the remaining gene products, which have an unknown function, have homologues in other bacteria, and the remainder appear to be *H. pylori*-specific.

Approximately 40% of *H. pylori* isolates contain plasmids, ranging in size from 1.5 to 23.3 kb, but they do not contain recognised virulence factors (Kleanthous et al, 1991; Minnis et al, 1995).

The aim of this study was to detect microbial candidate protein markers whose presence might be correlated with the development of four different clinical consequences of *H. pylori* infection GU, DU, NUD and GI. This was examined by the means of proteome analysis. The proteome of an organism may be defined 'as the total protein complement expressed by an organism or tissue under a given set of conditions'.

1.2 PROTEOMICS

1.2.1 Genomics, Transcriptomics and Proteomics

Comparative molecular biology analyses can be performed at three levels, the genome, the transcriptome and the proteome.

Genome

The genome has been defined as the sum total of all the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs (Yeatman, 2003). Genomics (the study of DNA or the genes of a cell) begins with the gene and makes inferences about the protein products of that gene (Wilkins et al, 1997). A gene has been defined as a specific segment of DNA that controls a specific cellular function.

Genomics is based primarily on sequencing of DNA. However, while the genomic methods have proved to be very successful these methods are unable to identify the relative concentrations of the proteins, to define when or if the protein is produced or the extent of post-translational modifications (Humphrey-Smith et al 1997; Pardanani et al, 2002).

The environment influences gene expression and modifies gene products in ways that may stimulate, accelerate, or inhibit biochemical processes. This does not change the genome, but inevitably changes the complement of transcribed genes (the transcriptome) and the composition of proteins translated from them (the proteome) in affected tissues (Wilkins et al, 1997).

Transcriptome

Cell biology can also be studied at the level of messenger RNA (transcriptome) (Su, 2002). The transcriptome has been defined as the complete profile of transcribed messenger RNA expressed in a given tissue, cell or biological system at a given time under specific conditions. Analysis of the transcriptome in a given sample yields a measurement of the relative level of each mRNA in the mRNA pool (Yeatman, 2003).

Regulation of the pool of mRNA in a cell or organism is the result of a series of complex mechanisms. The concentration of a particular mRNA is dependent on its rates of synthesis, transcription, transport from the nucleus (in eukaryotes) and its rate of decay (Kettman et al, 2001). The half-lives of mRNAs may change in response to hormones, nutrient levels, cell growth rates, viral infection, exposure to toxins or carcinogens and temperature shifts in ways that are not completely understood. In addition, the half-lives of most mRNAs are probably determined by other factors, including their affinity for proteins (Humphrey-Smith et al, 1997; Wilkins et al 1997).

A protein cannot be synthesised without its mRNA being present, and conversely there may be abundant amounts of specific mRNA but no translation of message into protein. For example, studies have shown that there may be a poor correlation between the total mRNA abundance and the amount of a particular protein expressed in a cell at a given time. Anderson and Seilhamer (1997) showed that human liver mRNAs were enriched for secreted proteins, whilst mRNAs for cellular proteins were under-represented. Clements and Foster (1998) showed that *Staphylococcus aureus* produced long-lived mRNAs for starvation recovery proteins, which quickly synthesised growth and replication proteins after the stimulus, indicating that the starvation event was over. Studies such as these show that in order to obtain an accurate estimate of the protein profile produced by an organism or cell under given conditions the proteome needs to be examined directly.

Proteome

Cell biology can also be studied at the level of the proteome. The term proteome was used for the first time in 1995 to describe the total protein complement of the genome (Wasinger et al, 1995). The proteome has been defined as the complete profile of proteins expressed in a given tissue, cell or biological system at a given time under specific conditions (Wilkins et al, 1997).

The proteome differs substantially in concept from the genome. The genome may be analysed in its entirety, from beginning to end and is fixed for each individual. The proteome is not a fixed feature of an organism. Instead, it changes with the stage of development, the tissue and the environmental conditions under which an organism finds itself (Wilkins et al, 1997). In addition, some body fluids such as serum or urine have no DNA or mRNA, which makes these samples incompatible for these types of analysis.

The transcription of DNA into mRNA is only the first step in a long series of events resulting in the production of a protein. For example, in eukaryotes, mRNA is subjected to post-transcriptional control in the form of splicing, polyadenylation and editing prior to translation (Strachan and Read, 2000). The coding into many different protein forms can be generated at this step. Proteins are also subject to post-translational modifications (e.g. phosphorylation, glycosylation or proteolysis) (Kirschner, 1999) and compartmentalisation (Colledge and Scott, 1999). These events cannot be determined from a DNA sequence or from mRNA expression levels; therefore, studies of the expressed proteins are required.

Proteome research endeavours to examine the total protein complement expressed by a particular genome and to address biological problems that cannot be answered by examination of nucleic acid sequences alone (Westermeier and Naven, 2002). The growth of proteomics is a direct result of advances in large-scale nucleotide sequencing of genomic DNA. Without this information, protein identification would be a very difficult process. Protein identification relies on the presence of some form of database for the given organism (Pandey and Lewitter, 1999: Shevchenko et al, 2000). The majority of DNA and protein sequence information has accumulated

within the last 11 years with the completion of many prokaryotic and eukaryotic genome sequences (Broder and Venter, 2000).

1.2.2 Post-translational modifications

In the 1940s Beadle and Tatum proposed that one gene encoded only one protein (Singer and Berg, 2004). It is now known that this is not true since multiple proteins may be expressed from a single gene depending on how the gene is transcribed, how the mRNA is processed and translated, and how the protein is post-translationally modified.

Post-translational modifications are critical to our understanding of physiological protein function. However, while DNA data and mRNA data infer that sequence motifs exist that may be modified they do not indicate which sequence motif will actually be post-translationally modified (Wilkins et al, 1997). It is estimated that up to 200 different types of post-translational modification exist (Krishna and Wold, 1993). The average number of protein isoforms per gene has been predicted to be one or two in bacteria, three in yeast, and three or more in humans (Wilkins et al, 1997). Therefore, it is clear that the tenet of "one gene, one protein" is an oversimplification.

Post-translational modifications of proteins occur after coding and translation; common examples include phosphorylation and glycosylation (Strachan and Read, 1997). Many proteins have multiple potential modifications states. For example, if a protein can be potentially glycosylated and phosphorylated each modification motif present would represent a potential protein form. Each additional state would add a large amount of additional diversity to the expression profile of that protein (e.g. 64 potential forms for just six modifications) (Wilkins et al 1997). All modified forms from one protein can also vary in abundance, activity or location inside a cell, which cannot be predicted from the genome or transcriptome.

Proteins are known to be modified post-translationally in response to a variety of intracellular and extracellular signals (Hunter, 1995). For example, protein phosphorylation is an important signalling mechanism and incorrect regulation of

protein kinases or phosphatases can result in oncogenesis (Hunter, 1995). By using proteomics, changes in the modifications of many proteins expressed by a cell or organism can be analysed simultaneously.

For example, glycosylation is believed to be an important biochemical alteration often associated with malignant transformation (Gorelik et al, 2001). By comparing the changes in protein glycosylation and total protein expression of normal tissue and carcinoma tissue, protein markers may be identified that differ in their glycosylation modification patterns and protein expression levels (Schulenberg et al, 2003; Dwek et al, 2001).

1.2.3 Proteomic Technologies

Thirty years ago when two-dimensional gel electrophoresis (2-DE) was introduced, very few tools existed for proteomics. Since that time, new technologies have emerged and old ones have been improved in areas from protein separation to protein identification. Currently there are three technological approaches dominating the field of proteomics. These are 2-DE, liquid chromatography and protein microarrays (Kersten et al, 2002). These three approaches are often used in conjunction with mass spectrometry. These techniques can be used either separately or jointly with one or more of the other methods.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) proteome analysis

2D-PAGE, first described by O'Farrell (1975) and Klose (1975), was the first approach used to study the simultaneous expression of multiple proteins in a complex mixture. A number of advances since 1975 have improved this established technique for protein separation and it remains a key method for the examination of differential protein expression. The 2D-PAGE approach permits the separation and detection of proteins from a wide variety of sources without the need for any prior knowledge of their function (Kersten et al, 2002; Wilkens et al, 1997).

Until recently, the separation of complex protein samples has been dominated by the techniques of 2D-PAGE, followed by mass spectrometric analysis of protein spots of interest. Proteins are separated according to two independent properties: in the first

dimension according to their isoelectric point and in the second dimension according to their molecular mass. The protein spots are then visualised by staining (e.g. Coomassie Brilliant Blue or silver nitrate). Spot intensity staining may be used to compare relative abundance between the specific proteins of two samples. A number of automated gel-scanning software programs are available to quantify spot intensity and facilitate gel comparisons (Marengo et al, 2005). The identity of a protein spot is established by excising the spot from the gel, performing protease digestion on the gel piece, extracting the resulting peptides, and obtaining peptide mass data or peptide sequence data via mass spectrometry (Thiede et al, 2005). The peptide mass fingerprint or sequence data is used to interrogate protein databases for protein identification.

Liquid Chromatography-Mass Spectrometry (LC-MS) based approaches

An alternative proteomics approach to the 2-DE method is LC-MS. In such cases, the protein mixture is digested and the resultant peptides eluted directly from a reverse phase high performance liquid chromatography (RP-HPLC) column into the mass spectrometer, where multiple peptide product ion MS/MS spectra are acquired (LC-MS) (Westermeier and Naven, 2002). Unfortunately, proteins are not generally amenable to high quality, robust separation of increasingly complex mixtures by RP-HPLC. If the sample is a complex mixture of proteins there is an increased likelihood that the digested peptides will co-elute which results in the MS being unable to select all the peptides for fragmentation. This has been circumvented by a two-dimensional LC/MS (2DLC/MS) approach. In the first dimension, the proteins are first digested typically, by trypsin, then acidified and subjected to cation-exchange chromatography (Westermeier and Naven, 2002). In the second dimension, the bound peptides are eluted with increasing ionic strength onto a RP-HPLC column. The peptides are bound due to their hydrophobic properties and subsequently eluted into the mass spectrometer (Wahsburn et al, 2001).

A further development of the RP-HPLC approach that is now being used is MudPIT (Multidimensional protein identification technology), (Wahsburn et al, 2001). The protein mixture is fractionated into a soluble fraction and an insoluble fraction (containing membrane proteins). The former fraction is digested with trypsin and

subjected to the RP-HPLC approach and the insoluble fraction is cleaved firstly with cyanogen bromide and secondly with trypsin before 2DLC/MS (Wahsburn et al, 2001).

To date the LC-MS approach has been used successfully to characterise relatively simple protein mixtures (Link et al, 1999; Washburn et al, 2001) and to look specifically at protein-protein interactions (Muller et al, 2001). 2-DE-based proteome analysis provides information about protein abundance at the gel level by comparing staining intensities. However, when peptide mixtures are analysed directly by LC/MS/MS techniques, much of the original quantitative information is lost (Peng and Gygi, 2001). Its application to expression proteomics is still at an early stage in its development and time will result in a better understanding of its true potential for the quantitative study of protein expression.

Protein Microarrays

The third technological approach that has been developed to screen complex protein samples rapidly involves the production of microarrays of antibodies, peptides or synthetic mimetic compounds. Similar to DNA microarrays protein microarrays consist of thousands of probes (proteins, peptides, or synthetic compounds) immobilised on a solid support, which may be a glass slide, chip or microtitre wells (Lopez and Pluskal, 2003). Recognition molecules utilised must be capable of binding the individual target proteins or a range of target proteins with a certain degree of affinity and specificity (Lopez and Pluskal, 2003). The array or recognition particles (e.g. recombinant proteins) may be protein/antibodies but do not need to be. The protein arrays are immobilised on a solid support (glass or membrane substrate) such that their ability to bind target protein(s) is not compromised. The target protein mixtures are applied and bind to their appropriate array that are then quantitatively detected (Kusnezow and Hoheisel, 2003). Common detection methods are radiography and fluorescent tags.

1.2.4 Proteomic Challenges

Automation

The study of proteins presents a number of unique challenges such as automation. Although the technology for the analysis of proteins is rapidly progressing, it is still not feasible to study proteins with the high throughput that can be achieved in the study of nucleic acids. Proteomic techniques rely on methods such 2D-PAGE which are not high throughput methods. Even the performance of mass spectrometry can generate stacks of data that may require considerable time in either data acquisition or analysis (Lopez, 2000). Although hundreds of proteins can be analysed quickly, and in an automated fashion, by the different types of mass spectrometers some proteins still cannot be identified. This may be due to poor digestion or the protein in question not being located in any of the protein databases.

Low abundance proteins

Proteomics is still limited in the study of low abundance proteins. In some eukaryotic cells, the amounts of the most abundant proteins can be 10^6-fold greater than those of the low-abundance proteins (Wilkens et al, 1997). Many important classes of proteins (that may be important drug targets) such as transcription factors, protein kinases, and regulatory proteins are low copy proteins. These low-copy proteins (1-10 copies per cell) will not be observed in the cell lysate without additional purification steps (Wilkens et al, 1997).

There is no equivalent of the PCR for proteins, so the analysis of low-abundance proteins remains a major challenge. In DNA studies the amplification of DNA is possible with PCR. In addition, in protein interaction studies native conformation of proteins must be maintained to obtain meaningful results (Wilkens et al, 1997).

Basic proteins

The visualisation of extremely basic proteins in 2-DE proteome analysis is still an area that requires improvement. Extremely basic proteins require sample pretreatement to enrich for basic proteins followed by separation on specialised immobilised pH gradient (IPG) strips (see page 34) with pH ranges extended beyond pH 11 (Bae et al, 2003). Strongly alkaline proteins such as ribosomal and nuclear

proteins with pIs between 10.5 and 11.8 have been separated using narrow IPGs pH 10-12 and 9-12 (strips created in a particular laboratory) (Gőrg et al, 1997). However, 2-DE quantitative analysis of alkaline proteins with pIs between 11-12 is still not possible in most laboratories.

In spite of the above limitations proteomics has enormous potential to provide insight into those questions that DNA and mRNA studies cannot answer. The ability to study the protein expression of organisms will provide new insights into post-translational modifications, differential expression of proteins, protein-protein interactions, protein localisation, the effects of environment factors on organisms, and disease/control studies.

1.3 2-DE PROTEOME ANALYSIS TECHNIQUES

2-DE Proteome Analysis has developed from, and continues to be dependent upon, the core technology of two-dimensional electrophoresis (2-DE) which is currently one of the major methods available capable of simultaneously separating thousands of proteins. 2-DE separation of proteins in the first dimension is based upon a surface charge fractionation by isoelectric focusing. This is followed by a mass-driven separation in the second dimension, by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

The techniques utilised in 2-DE proteome analysis are as follows (see Figure 1.4):

- protein preparation, e.g. cell disruption, inhibition of proteolysis, removal of contaminants and solubilisation of proteins;
- first dimension Separation – isoelectric focusing;
- second dimension Separation – SDS PAGE on polyacrylamide slab gels;
- visualisation – staining of gels;
- image analysis – determination of differences between 2-DE gels;
- excision of protein spots of interest;
- digestion of isolated proteins and mass spectrometry – e.g. Matrix Assisted Laser Desorption Ionisation (MALDI); and
- bioinformatics.

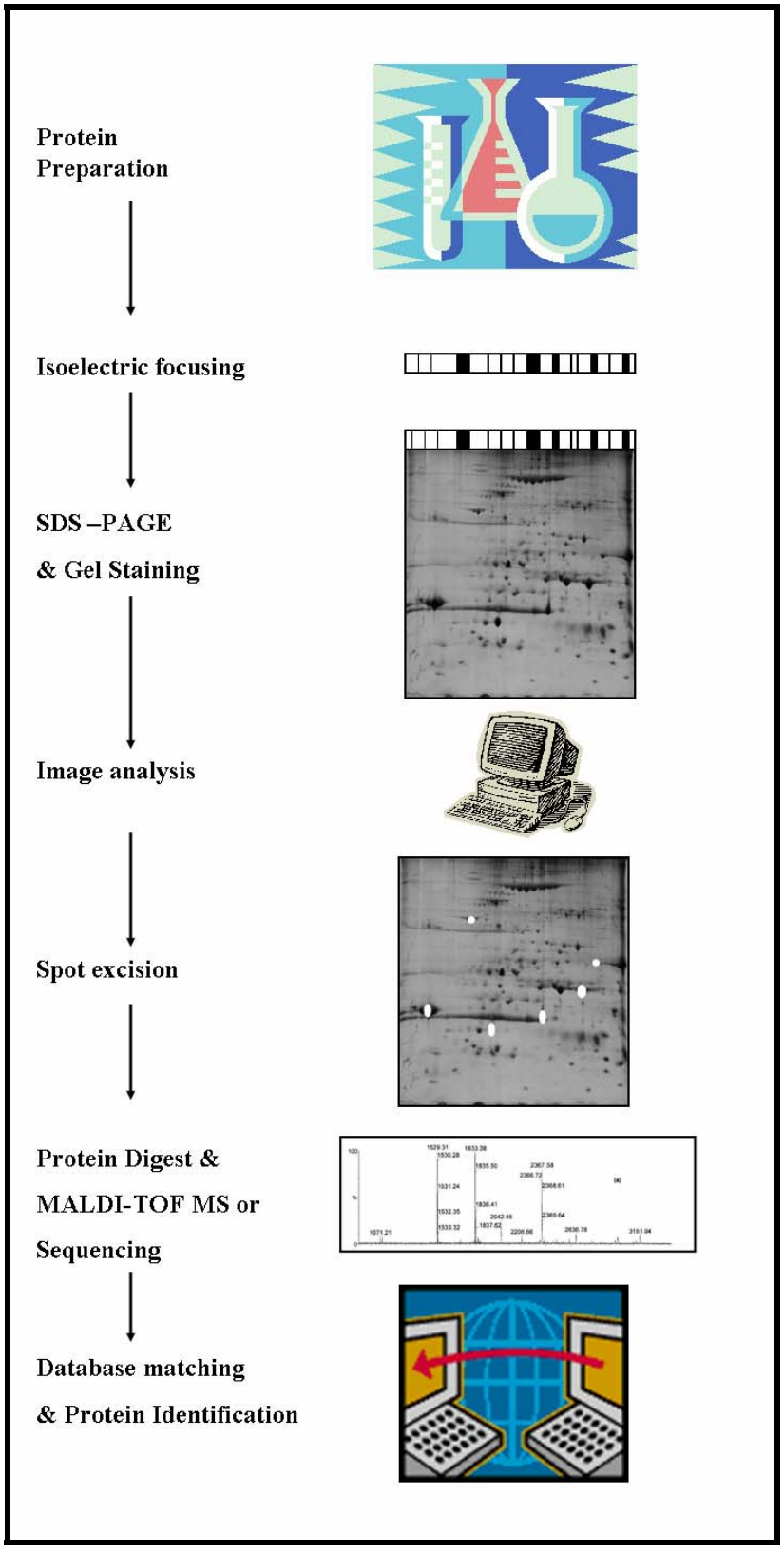


Figure 1.4 2-DE proteome analysis techniques

An example of the techniques that might be utilised in a Proteome analysis project using a combination of 2D gel electrophoresis and MALDI-TOF-MS.

1.3.1 Protein Preparation

Adequate sample preparation is one of the most important prerequisites for a successful 2-D experiment. The sample preparation procedure may include cell disruption, inhibition of proteolysis, removal of contaminants such as salts, RNA, DNA and lipids, and solubilisation of all proteins. In some samples cell disruption may not be required as proteins are already in solution e.g. blood, tears, and milk. Ideally, to avoid protein losses, complete sample preparation should be accomplished in a single step in order to eliminate unnecessary handling. However, this is often not possible. There is no universal protocol for sample preparation.

1.3.2 Cell Disruption

To fully analyse all proteins contained in cells or tissues, the sample must be effectively disrupted. The choice of disruption method is dependent upon the type of sample and if the analysis is targeting all of the material or just a particular subcellular fraction. There are a number of cellular disruption techniques that may consist of gentle (typically mammalian) or more vigorous lysis methods (microbial and plant cells) (Gőrg et al, 2004). Cell disruption techniques include osmotic lysis, freeze thaw cycling, detergent lysis, enzymatic lysis of the cell wall, sonication, grinding with (or without) liquid nitrogen, high pressure (e.g. French press), homogenization with glass beads and a bead beater, nitrogen cavitation, or a rotating blade homogenizer. These techniques may be used on their own or in combination (Wilkens et al, 1997; Westermeir and Naven, 2002).

1.3.3 Protection Against Proteolysis

When cells are lysed, proteases are often liberated or activated. There are a number of methods to prevent the action of these proteases. Most are inactive at low temperatures and at pH above 9 (Olivieri et al, 2001). Therefore proteases can often be inhibited by Tris base, sodium carbonate, or a basic carrier ampholyte mixture, although the use of an alkaline environment (pH 9 or greater) may result in the loss of some basic proteins (Westermeier and Naven, 2002). Strong denaturants such as 8 M urea, 10% TCA or 2% SDS also inhibit proteases (Wu and Wang, 1984).

These approaches alone are sometimes insufficient to prevent proteolysis. For this reason a cocktail of protease inhibitors may be utilised. Such cocktails include a number of compounds designed to inhibit different classes of proteases (e.g. EDTA to inhibit metalloproteases by chelating the metal ions they require for activity).

1.3.4 Removal of Contaminants

A crude extract may be contaminated with non-protein impurities such as salt ions, phospholipids, nucleic acids, nonionic detergents, polysaccharides, phenolic compounds and insoluble material, which interfere with separation and resolution of proteins during 2-DE. However, steps undertaken to remove impurities have the potential to cause losses of proteins that will alter the subsequent protein profile (Westermeier and Naven, 2002; Gőrg et al, 2004). Therefore such steps should only be used where these impurities will significantly decrease the resolution of the proteins.

Salts

Salts, residual buffers, and other charged small molecules disturb the isoelectric focusing (IEF) process and must be limited to as low a concentration as possible (preferably less than 10 mM). In an electric field salt contaminants start to migrate towards the electrodes. Due to their size these contaminants will migrate at a faster rate than the proteins in the applied sample (Westermeir and Naven, 2002; Gőrg et al, 2004). During the first dimension separation contaminants are forming ion fronts between the region with low ion concentration and regions with high ion concentrations, which are moving towards the electrodes (Gőrg et al, 2004). These differences in conductivity generate heat. If the voltage applied is too high during the migration of these contaminants it will generate regions of high heat that will burn the strips where the ion front is located (Westermeir and Naven, 2002).

Desalting can be performed by tube dialysis, mini-desalting columns, spin dialysis, gel filtration or protein precipitation/resuspension using organic solvents such as methanol or acetone (Westermeier and Naven, 2002; Oh-Ishi and Maeda, 2002). In some instances, proteins cannot be desalted prior to electrophoresis in the first dimension. For example, proteins in halobacteria lysates will not be soluble if the

salt is removed (Bandyopadhyay and Sonawat, 2000). In such cases, the sample can be electrophoreticaly desalted during the first IEF phase by extended application of a low voltage.

Phospholipids

Many proteins, especially membrane proteins, form complexes with lipids. This reduces their solubility and can affect both their apparent pI and M_r. Strongly denaturing conditions and detergents minimise lipid-protein interactions. Lipids can often be removed with detergent or by precipitation with acetone (Westermeir and Naven, 2002).

Nucleic acids

Nucleic acids can increase sample viscosity which causes background smearing in first dimension electrophoresis. Nucleic acids can also bind to proteins through electrostatic interactions and inhibit isoelectric focusing or cause protein precipitation (Link, 1999; Gőrg et al, 2004). Nucleic acids are visualised with silver staining as horizontal streaks in the acidic part of the gel. Ultracentrifugation can be used to remove large DNA. However, this technique may also result in loss of very high molecular mass (M_r) proteins from the sample (Westermeier and Naven, 2002). Nucleic acids can also be removed by treatment with endonuclease, or with sonication, that breaks them into small fragments. Protein precipitation with methanol or acetone also removes nucleic acids, which remain in the supernatant.

Ionic detergent

Ionic detergents (usually SDS) are often used during protein extraction; however, this can strongly interfere with protein migration in the first dimension. SDS forms complexes with proteins, and the resulting negatively charged molecules will not properly focus (Ames and Nikaido, 1976). SDS can partially be removed by acetone precipitation. The protein is precipitated by the acetone and the SDS remains in solution. Protein precipitation will be more complete if conducted at $-20^{o}C$ (Guy et al, 1994).

Polysaccharides

Polysaccharides can clog gel pores in the first dimension, causing protein precipitation and extended focusing times, which result in horizontal streaking (Gőrg et al, 2004). Some polysaccharides contain negative charges and can complex with proteins by electrostatic interactions. Polysaccharides can be removed by precipitating the sample in TCA, ammonium sulphate or phenol/ammonium acetate. Ultracentrifugation will also remove high M_r polysaccharides. However, loss of very high M_r proteins may occur.

Phenolic compounds

Phenolic compounds are present in many plant tissues and can modify proteins through an enzyme-catalysed oxidative reaction (Gőrg et al, 2004). They may be treated by a variety of methods:

1. prevention of phenolic oxidation by the use of reductants during extraction;
2. rapid removal of phenolics by precipitation;
3. use of oxidative inhibitors such as diethyldithiocarbamic acid or thiourea; and
4. removal by adsorption to polyvinylpyrolidone (PVP) and/or polyvinylpolypyrolidone (PVPP) (Flengsrud and Kobro, 1989; Cremer and van de Walle, 1985).

Insoluble material (cell debris)

Insoluble material in the sample can clog gel pores in the first dimension and result in poor focusing. Samples should always be centrifuged prior to application in the first dimension (Westermeier and Naven, 2002).

1.3.5 Protein Solubilisation

The sample solution (IPG rehydration solution) must contain certain components to ensure complete solubilisation and denaturation of proteins prior to their separation in the first dimension. A particular challenge in proteomics is the solubilisation and separation of proteins such as membrane and membrane-associated proteins and proteins from insoluble samples such as hair and skin (Wilkins et al, 1997). Enhanced protein solubility has largely been achieved by including appropriate

reagents, which can be divided into three groups - chaotropes, surfactants and reducing agents.

Chaotropic agents

Chaotropic agents such as **urea** enable proteins to unfold and thus expose their hydrophobic cores. This is achieved by changing the hydrogen bond structure in the solution, thus decreasing the energy penalty for contact of the hydrophobic residues with the solution (Herbert, 1999). However, this can increase the potential for hydrophobic interactions, especially when lipids are present in the sample (Rabilloud 1996; Rabilloud, 2000). In practice, with urea denaturation it is normal to have surfactants such as 3-(3-cholamidoproyl) dimethylammonio-1-propane sulfonate (CHAPS) present to aid in protein solubilisation (Herbert, 1999).

Thiourea

Thiourea is used in combination with urea to increase the solubility of proteins. Thiourea is an efficient chaotrope, although it is poorly soluble in water and requires high concentrations of urea for solubility, the optimal conditions being solutions of 2 M thiourea in 5-7 M urea (Herbert, 1999; Link, 1999). Although the use of urea and thiourea improve solubilisation of hydrophobic proteins, it is only in combination with surfactants such as CHAPS and sulfobetaine 3-10 that improved sample solutions for 2-DE may be achieved (Herbert, 1999).

Surfactants

It is normal to have at least one surfactant present in the first dimension sample solution to solubilise the hydrophobic regions that are exposed because of denaturation in chaotropes. Detergents act on hydrophobic interactions by providing a stable dispersion of hydrophobic proteins in the aqueous medium, in the form of micelles (Link, 1999).

Surfactants that are suitable for use in the IEF buffer are restricted to nonionic (Triton X-100, Nonidet P-40) or zwiterionic surfactants (CHAPS) however, the use of nonionic or zwiterionic detergents in the presence of urea presents some problems. In concentrated urea solutions, urea can bind linear alkyl chains to form inclusion

compounds (Link, 1999). These complexes are much less soluble than free solute, and may precipitate (Link, 1999). This precipitation is stronger with increasing alkyl chain length and higher urea concentrations. Therefore, many nonionic or zwiterionic detergents with linear hydrophobic tails cannot be used in the presence of high concentrations of urea (Link, 1999).

This limits the choice of surfactants mainly to those with non-linear alkyl tails (e.g. Triton X-100, Nonidet P-40, CHAPS) or with short alkyl tails (e.g. octyl glucoside). Commonly used surfactants such as Triton X-100, Nonidet P-40 and the sulfobetaine CHAPS are soluble in high concentrations of urea (Wilkins et al 1997; Herbert, 1999). However, they are not efficient at quenching hydrophobic interactions in high concentrations of chaotropes and their solubilising power is further reduced in the presence of the highly chaotropic thiourea (Gőrg et al, 2004). In contrast, sulfobetaines with long linear alkyl tails such as N-decyl-N, N-dimethyl-3-ammonio-1-propane sulfonate are more efficient at protein solubilisation but suffer from poor solubility in high concentrations of urea (Herbert, 1999). In order to achieve maximum efficiency for those proteins that require strong surfactants, low concentrations of the chaotropes may have to be used (Herbert, 1999; Link, 1999). For example, SB3-10 may only be used with urea concentrations $\leq$ 5 M.

Reduction and alkylation

Proteins contain a number of cysteine residues that are often linked in pairs to form disulphide bridges. In order to allow complete unfolding of many proteins it is necessary to reduce disulphide bonds. Protocols to do this have been utilised in proteome analysis since its early beginnings.

Reduction of protein disulphide bridges is usually achieved with a free-thiol-containing reducing agent such a 2-mercaptoethanol or dithiothreitol (DTT). However, free-thiol-containing reagents such as these are charged, especially at alkaline pH, and migrate out of the pH gradient during the IEF (Galvanni et al 2001; Gőrg et al, 2004). This results in reformation of disulphide bonds and a consequent loss of solubility for some proteins, especially those that are prone to interaction by disulphide bonding (Wilkins et al, 1997; Herbert et al, 1998).

As reduced polypeptide chains migrate to their isoelectric point (pI), they sometimes leave behind the reducing agent allowing either the reformation of disulphide bridges or a variable degree of cysteine-alkylation by residual acrylamide monomers in the gel (Herbert, 1999). The regeneration of disulphide bridges could occur between unlike polypeptide chains (i.e. not only within but also between polypeptide chains). If this were to occur, artifactual spots could be generated in the 2-DE gel.

In peptide mass spectrometric analysis, the identity of cysteine-containing peptides is difficult to establish if different mass values are generated from cysteine oxidation products and acrylamide adjuncts (Yan et al, 1998). This results in a range of possible masses being obtained which may generate a number of possible protein identities during protein database searches. Therefore, it is advantageous for proteomic applications to modify cysteine to completion by reduction with tributyl phosphine (TBP) or dithiothreitol (DTT) followed by alkylation, prior to separation in the first dimension (Yan et al, 1998).

However, free-thiol-containing reagents, such as DTT, react with acrylamide or other alkylating agents (iodoacetamide, N-ethylmaleimide) in preference to the cysteine-thiol. This may lead to incomplete alkylation of cysteine residues. Herbert et al (1998) replaced DTT with an uncharged reducing agent, TBP. Because phosphines do not contain a thiol, they cannot be alkylated by acrylamide. This leads to a simplified IPG strip re-equilibration protocol incorporating reduction and alkylation in a single step and cysteine modification proceeds to completion (Herbert, 1999). This greatly enhances protein solubility during the IEF, which increases the transfer of protein to the second dimension with a range of samples.

1.3.6 First dimension

Proteins are amphoteric molecules with both acidic and basic buffering groups that become protonated or deprotonated depending on the pH environment. In a basic environment, the acidic groups become negatively charged, while in the acidic environment the basic groups become positively charged. The net charge of a protein is the sum of all negative and positive charges of the amino acid side chains

and amino- and carboxy-termini. The isoelectric point (pI) is the specific pH at which the protein has no net charge and therefore stops migrating in an electric field (Righetti, 1983; Berkelman et al, 1998). Proteins are positively charged at pHs below their pI and negatively charged at pHs above their pI.

Isoelectric focusing

Isoelectric focusing is performed in a pH gradient. When a protein is placed at a certain pH on the gradient and an electric current is applied, the protein will start to migrate towards the electrode of the opposite sign to its net charge (Righetti, 1983). Because it migrates inside a gradient, it will eventually arrive at a pH value equivalent to its isoelectric point where it will stop (Berkelman et al, 1998). Should it diffuse away from this point it will become charged again and migrate back to its pI. The effect is the focusing of each protein to a tight band centred on its pI. The efficiency of focusing is improved by increased voltage. Isoelectric focusing can produce high-resolution separation of complex mixtures of proteins.

There are two kinds of pH gradient gels (Righetti, 1983; Westermeier and Naven, 2002):

- those which are formed in a gel by the effect of an electric field upon amphoteric buffers (carrier ampholytes), which are free in solution; and
- immobilised pH gradients (IPG) in which the buffering groups are attached to the gel matrix.

Carrier ampholyte-generated pH gradients

Carrier ampholytes are small, soluble, amphoteric molecules that have a high buffering capacity near their pI. They have good and regular electric conductivity and an absence of biological effects (Righetti, 1983; Westermeier and Naven, 2002). Almost all the carrier ampholytes are charged; those with the higher pI positively, those with the lower pI negatively.

Carrier ampholyte-generated pH gradients were the first to be developed for isoelectric focusing (O'Farrell, 1975; Klose, 1975). IEF was performed using polyacrylamide rod gels with pH gradients generated using carrier ampholytes (Link,

1999). In this technique, the pH gradient is established by the electric field. At the beginning, the gel with carrier ampholyte has a uniform average pH value. When an electric field is applied however, the negatively-charged carrier ampholytes migrate towards the anode, the positively-charged ones to the cathode, with velocities depending on their net charges (Berkelman et al, 1998). The carrier ampholytes then align themselves in between the two electrodes according to their pI, forming a pH gradient.

To maintain a gradient that is as stable as possible, electrode solutions (anode and cathode buffers) are used between the gel and the electrodes. An acid is used at the anode and a base at the cathode (Righetti, 1983). Should an acidic carrier ampholyte reach the anode, its basic buffering group would acquire a positive charge from the acidic solution there and it would be attracted back to the cathode (Berkelman et al, 1998; Link, 1999). Similarly, basic carrier ampholytes would acquire a negative charge from the basic solution and be attracted back to the anode. Although theoretically the carrier ampholytes should have functioned as previously mentioned in fact, given enough time, they eventually suffer from gradient drift (Berkelman et al, 1998).

Isoelectric focusing is a relatively slow separation method because, close to their isoelectric points, proteins have only a low net charge and therefore low mobility. Also high resolution requires long separation distances, which results in extended migration times (Link, 1999). Long runs present difficulties in reproducibility for carrier ampholyte-generated pH gradients because of gradient drift. Because the carrier ampholytes are in free solution, the gradient is unstable during an extended IEF run and they usually drift towards the cathode (Link, 1999). This gradient drift causes a flattening of the pH gradient at both ends of the gel. This effect is particularly noticeable above pH 9 and most of the basic proteins eventually drift out of the gel together with the basic part of the gradient (Berkelman et al, 1998). Because of this, the protein pattern is time dependent.

Another problem is that the proteins of the sample behave like additional carrier ampholytes and modify the pH profile of the gradient. This means that the gradient is sample dependent (Link, 1999; Westermeier and Naven, 2002).

The pH gradients generated by such procedures in the first dimension are inherently variable, which makes it difficult to obtain reproducible 2-D gel runs in large numbers (Gőrg et al, 2004; Berkelman et al, 1998). The subsequent results are therefore highly dependent on the skill of the operator.

Immobilised pH gradients (IPG)

Due to the limitations of the carrier ampholyte method, an alternative technique for pH gradient formation was developed, IPG (Bjellqvist et al, 1982). An IPG strip is created by covalently incorporating a gradient of acidic and basic buffering groups into a polyacrylamide gel at the time it is cast. The buffers, called acrylamido buffers (e.g. Amersham Biotech *Immobilines™*), are a set of well-characterised molecules, each with a single acidic or basic buffering group covalently linked to an acrylamide monomer (Link, 1999; Westermeier and Naven, 2002). The pH gradient is generated by casting a gradient gel using a continuously varying mixture of acrylamido buffers with the desired pH extremes. The main advantage of the IPG system is that during polymerisation the buffering groups forming the pH gradient are immobilised via vinyl bonds to the polyacrylamide backbone of the strip (Link, 1999; Westermeier and Naven, 2002). This results in pH gradients that are effectively stable, thus eliminating the effects of pH gradient drift (cathodic drift) encountered in conventional IEF using free ampholytes (Link, 1999; Westermeier and Naven, 2002). After the advent of immobilised pH gradients stabilised on a stiff plastic support strips, the practicality of utilising 2-DE gels to resolve complex protein mixtures in a quick and reproducible manner became apparent.

Advantages of immobilised pH gradients

There are a number of advantages of IPG strips compared to gels with pH gradients generated from carrier ampholytes. These are (Westermeier and Naven, 2002; Berkelman et al, 1998):

- standardised reproducibility;

- the chemistry of the buffering acrylamide derivatives is better controllable;
- the film-supported gel strips are easy to handle;
- the fixed gradients are not modified by the sample composition, and they do not drift with IEF time;
- stable basic pH gradients allow reproducible separation and display of basic proteins;
- higher protein loads are achievable;
- less protein is lost during equilibration in SDS buffer, because the fixed charged groups of the gradient retain the proteins like a weak ion exchanger; and
- reproducible protein patterns can be prepared on gradients engineered according to special needs. They are commercially available in a variety of pH ranges (wide range, pH 3-10 and narrow range, pH 4.5-5.5).

Carrier ampholytes continue to have an important function in IPG strip sample application as they improve the solubility of proteins considerably. The addition of carrier ampholytes in the sample solution enhances protein solubility by reducing protein aggregation due to charge-charge interactions (Berkelman et al, 1998). In a solution, they are charged but they do not disturb isoelectric focusing like ionic buffers, because they migrate to their pIs, where they become uncharged. Therefore, small quantities of free ampholyte are included in the rehydration solution (Berkelman et al, 1998).

1.3.7 Second dimension

Sodium dodecyl sulphate (SDS) electrophoresis

With SDS-PAGE the polypeptides are separated on gels in the presence of SDS according to their apparent molecular masses.

Polyacrylamide composition

A polyacrylamide gel is formed from the polymerisation of acrylamide monomers into long chains and the cross-linking of these chains by bifunctional compounds such as N, N-methylenebisacrylamide (Link, 1999). Different cross-linkers (e.g. piperazine diacrylamide and bis-acrylamide) impart different characteristics to

polyacrylamide gels (e.g. susceptibility to swelling, tensile strength and background staining characteristics). Polymerisation of acrylamide is initiated by the addition of either ammonium persulphate or riboflavin. N,N,N',N'-Tetramethylethylenediamine (TEMED) may be added to accelerate the polymerisation process by catalysing the formation of free radicals from persulphate (Link, 1999). These in turn initiate polymerisation. Oxygen inhibits polymerisation and thus gel mixtures are usually deoxygenated by reduced air pressure before gels are poured.

The pore size can be exactly and reproducibly determined from the total acrylamide concentration *T* and the degree of cross-linking *C*.

$$T = \frac{(a + b) \times 100}{V}\ [\%], \qquad C = \frac{b \times 100\ [\%]}{a + b}$$

Where **a** is the mass of acrylamide in g
b the mass of crosslinker in g
V the volume in ml

When *C* remains constant and *T* increases, the pore size decreases. When *T* remains constant and *C* increases, the pore size follows a parabolic function. At high and low values of *C*, the pores are large (Westermeier and Naven, 2002).

SDS and TBP

SDS binds to proteins to form complexes with a necklace-like structure composed of protein-decorated micelles connected by short flexible polypeptide segments (Ibel et al, 1990) As a result, large amounts of SDS are incorporated in the SDS-protein complex in a ratio of approximately 1.4g SDS/g protein (Berkelman et al, 1998). Because SDS masks the native charge of the proteins, the SDS-protein complexes have essentially a constant net negative charge per unit mass. SDS solubilises proteins by blocking hydrogen bonding and inhibiting hydrophobic interactions, and largely unfolds the protein molecules, eliminating the secondary and tertiary structures (Link, 1999).

Usually a reducing agent such as TBP is added to the sample to reduce the disulphide bridges between cysteines (Herbert et al, 1998). Following separation in the first dimension, in order to allow complete unfolding of many proteins, it is necessary to reduce disulphide bonds. In addition, cysteine residues need to be completely alkylated prior to second dimension separation to prevent difficulties associated with incomplete alkylation of proteins by free monomers of acrylamide in the gel (see Reduction and Alkylation page 30).

Essentially, the electrophoretic mobility of proteins treated with SDS and TBP depends only on the molecular mass of the protein.

Homogeneous versus gradient gels

Slab gels may be prepared with either a single concentration of acrylamide or gradient-polyacrylamide forms that can be optimised to separate proteins over a specific M_r range.

Large format single percentage (homogeneous) gels offer good resolution within a specific M_r range but poor resolution outside of this M_r range (see Table 1.1). For example, a 7.5% acrylamide gel might efficiently resolve proteins with M_r between 24-200 kDa (this is partially dependent upon the pore size of the gel and the complexity of the sample) (Berkelman et al, 1998). A commonly used second-dimension gel for 2-DE is a homogeneous gel containing 12.5% *T* and 3% *C*.

Gradient gels have an increasing concentration of acrylamide from top to bottom. For example, in a 10-20% gradient gel there is 10% polyacrylamide concentration at the top that increases to 20% at the bottom of the gel. When a gradient gel is used, the overall separation interval is wider and the linear separation interval is larger (Berkelman et al, 1998). The entire M_r range may be displayed on one gradient gel where two or more homogeneous gels would be required to obtain good resolution of the same M_r range. In addition, bands are sharper on gradient gels because the decreasing pore size functions to minimise diffusion. However, a gradient gel requires more skill to cast, so unless precast commercial gels are utilised they may present problems with reproducibility (Link, 1999). Another disadvantage of gradient

% Acrylamide in resolving gel	Separation Size Range (M_r x 10^3)
Single Percentage	
5%	36-200
7.5%	24-200
10%	15-200
12.5%	7-100
15%	0-60
Gradient	
5-15%	14-200
5-20%	10-200
10-20%	0-150

Table 1.1 Acrylamide Gel Concentration Guide

An approximate guide to the acrylamide concentrations suitable for protein separation by M_r (Berkelman et al, 1998).

gels is that they compress the protein pattern that leads to spatial resolution difficulties, especially for high M_r proteins.

1.3.8 Visualisation of spots

There are many techniques employed for the detection of protein on 2-D gels. The four most commonly used methods are staining with either coomassie brilliant blue, silver nitrate, or fluorescent dyes and autoradiography.

In choosing the chemical entity to bind to the proteins the key issues which must be considered are, sensitivity (detection threshold), linearity of response, homogeneity (i.e. consistency of dye biding by different proteins), and reproducibility (Berkelman et al, 1998; Westermeier and Naven, 2002). Homogeneity and reproducibility are generally less important for 1-D electrophoresis, in which different samples are loaded in parallel lanes of the slab gels. The situation is quite different in 2-D slab gels, in which there is usually only one sample per gel (exception is fluorescent dual labelling of proteins [Bernhardt et al, 1999]), because the abundance of different protein spots separated on one gel have to be compared to those on other gels.

There is a list of desirable properties for the spot detection techniques in 2-DE gels for proteomics (Westermeier and Naven, 2002) they are:

1. be sensitive enough for low copy number proteins;
2. allow quantitative analysis;
3. have a wide linearity of range;
4. have a wide dynamic range;
5. be compatible with mass spectrometry;
6. be non toxic;
7. be environmentally friendly;
8. have ease of handling (time, complexity of procedure);
9. be reproducible; and
10. be inexpensive

Unfortunately, no staining method has all of these features. The choice of staining technique for any particular project will often be dependent on the amount of the protein sample available. For example, in the study of human tissue biopsies it may be difficult to obtain large quantities of protein. However, when studying bacterial proteomes, it is often easy to acquire large protein quantities. Low protein amounts will require dyes with higher sensitivities and this will determine the staining technique utilised.

Coomassie Stains

There are a number of Coomassie Brilliant Blue (CBB) dyes including R250, R350 and G250, all of which belong to the triphenylmethane family. CBB forms complexes with basic amino acids such as arginine, tyrosine, lysine and histidine. CBB staining requires an acidic medium for electrostatic attraction to be exerted between the dye molecules and the amino groups of the proteins (Link, 1999). This ionic attraction, together with the van der Waals' forces, binds the dye-protein complex together. The binding however, is fully reversible under the appropriate conditions.

Coomassie R250. The Coomassie R250 dye was the first of the triphenylmethane dyes to be introduced. This dye has a detection limit of 30-100 ng of protein per spot (Patton, 2002). The classical approach involves the use of a regressive staining approach where gels are saturated with CBB R250 that has been dissolved in an aqueous solution containing methanol and acetic acid, followed by destaining in a similar solution devoid of the dye (Patton, 2002). The proteins have a higher affinity for the dye molecules than the gel matrix and therefore a point is reached where the background staining is minimal but the protein spots are well defined (Patton, 2002). However, the classical CBB R250 staining procedure is often not appropriate for 2D gel electrophoresis. Gel-to-gel reproducibility in staining is difficult to control because during gel destaining, with solutions containing alcohol, the protein spots are partly destained to varying extents along with the gel background. Some proteins such as collagens may even destain before the gel itself (Westermeier and Naven, 2002). Because no steady state of staining is reached, quantification is not reliable and not reproducible.

Coomassie G250. In 1985, a new colloidal technique for the staining of 2-DE gels was developed (Neuhoff et al, 1988). Coomassie G250 solution utilises 3% phosphoric acid. In the presence of a strong acid CBB G250 has diminished solubility in methanol permitting its use as a colloidal dispersion. Colloidal CBB G250 stain also contains alcohol, but in the presence of ammonium sulphate. Ammonium sulphate increases the strength of the hydrophobic interactions between proteins and dye (Neuhoff et al, 1988). Initially trichloracetic acid (TCA) was utilised instead of phosphoric acid, but it was found that this led to irreversible, acid-catalysed esterification of glutamic and aspartic acid side chain carboxyl groups, an effect that can complicate interpretation of peptide mapping data from mass spectrometry due to variable esterification of proteins. The use of phosphoric acid with methanol prevented this problem (Peisker, 1988; Mitra et al, 1994).

During staining, equilibrium is achieved between colloidal particles and freely dispersed dye in solution. The low concentration of free dye penetrates the gel matrix and preferentially stains the proteins, but colloidal dye particles are excluded from the gel, thus preventing matrix staining (Neuhoff et al 1988; Patton, 2002). This permits rapid staining of proteins without an undesired background. Because of this, the G250 procedure has the advantage of not requiring the gel alcohol-destaining step. The limit of protein detection for colloidal CBB stain is approximately 8-10 ng of protein and, since the destaining step is unnecessary, quantitation is more reliable (Patton, 2002). However, the linear dynamic range of the dye is relatively small.

CBB colloidal stained gels are compatible with mass spectrometry because the dye can be completely removed from the proteins prior to analysis. It is important to use a dye of high quality in order to avoid interference due to contaminants. The CBB G250 colloidal stain has 5 of the 10 desirable staining properties listed on page 39. These are that it allows quantitative analysis, has a wide dynamic range, is compatible with mass spectrometry, is highly reproducible and is the most affordable of the four detection methods.

This technique is an improvement on previous Coomassie staining methods although it still has a lower sensitivity than silver.

Silver Stains

The silver ion (Ag^+) complexes with lysine residues of proteins. The basic mechanism underlying all silver staining involves reduction of ionic silver to its metallic form (Switzer et al, 1979; Oakley et al, 1980). Detection of proteins in the gel requires a difference in the oxidation-reduction potential between sites occupied by proteins and adjacent sites of the gel. If the protein site has a higher reducing potential than the surrounding gel or matrix, then the protein will be positively stained (Westermeier and Naven, 2002). Conversely, if the protein site has a lower reducing potential than the surrounding gel, the protein will appear to be negatively stained. Negatively stained protein spots give the appearance of a lighter stained centre surrounded by a darker stained border. The detection limit for silver nitrate stains is about 0.5 – 1.5 ng per spot (Patton, 2002).

Silver staining methods are quite complex, multi-step procedures that must be stopped at some arbitrary time point in order to avoid over-development. Therefore, gel-to-gel reproducibility can be difficult to achieve, one study found that there are variations of 20% in spot intensities on duplicate gels (Quadroni and James, 1999). The addition of a specific quantity of a known protein can be used as a reference point for normalisation of gel spot intensity.

This means that silver staining is one of the most sensitive non-radioactive methods for detecting proteins separated by gel electrophoresis. It is 20-fold more sensitive than Coomassie staining for most proteins and equivalent to that provided by fluorescent stains such as SYPRO Ruby (White et al, 2004). However, small traces of contaminants may cause a loss of sensitivity and result in staining artefacts. Therefore, care is required in selecting reagents and during handling of gels in order to obtain high sensitivity and good images.

Silver stains can be modified for compatibility with mass spectrometry by omitting glutaraldehyde from the sensitising solution (Shevchenko et al, 1996; Scheler et al,

1998). The detection sensitivity then decreases to about one fifth that of the unmodified procedure.

Silver staining has 3 of the 10 desirable properties listed on page 39, these being that it is sensitive enough for low copy number proteins, has a wide linearity of range, and is a medium-cost staining technique (more expensive than CBB G250 but less expensive than SYPRO or autoradiography detection methods).

SYPRO™ fluorescent dyes

Certain d-block and lanthanide transition metal complexes such as those containing ruthenium, rhenium, osmium, platinum, europium, or terbium are intensely luminescent. Early SYPRO fluorescent dyes such as SYPRO Red and Orange bind protein through noncovalent interactions with the SDS-shell complex on SDS-protein complexes (Steinberg et al, 1996a; Steinberg et al, 1996b). However, the staining techniques of these dyes are not compatible with electroblotting or electroeluting and they are less sensitive than the more recent SYPRO stain SYPRO Ruby (Patton, 2002). SYPRO Ruby is a ruthenium-based metal chelate that binds avidly to proteins by a CBB-type mechanism primarily involving lysine, arginine, and histidine residues (Patton, et al 1999). The SYPRO Ruby stain is more sensitive than CBB and exhibits a broader linear dynamic range.

The SYPRO Ruby staining method is rapid, requiring approximately 4 hours in comparison to the 24 hours required by the colloidal CBB stain. Like colloidal CBB G250, but unlike silver nitrate, SYPRO Ruby is a staining procedure that approaches a fixed end-point (Patton et al, 1999). Thus, staining times are not critical and the procedure can be performed for extended periods without gels overdeveloping. Proteins that stain poorly with silver are often readily detected by SYPRO Ruby (Nishihara and Champion, 2002). However, a disadvantage of SYPRO Ruby is that it is expensive in comparison to colloidal CBB and silver nitrate stains. The SYPRO Ruby method has 6 of the 10 desirable staining properties listed on page 39, these being that it is sensitive enough for low copy number proteins, allows quantitative analysis, has a wide linearity of range, has a wide dynamic range, is compatible with mass spectrometry, and is highly reproducible. However, due to the cost of this

staining technique large-scale proteome analysis studies may become too expensive. Fluorescent dyes are not visible to the naked eye, which makes it difficult to ascertain if a particular experiment has been successful until the gel is scanned by a fluorescence imager.

Radiolabelling

Radiolabelling is commonly accomplished by incorporating ^{3}H, ^{14}C, ^{35}S, ^{32}P, or ^{125}I into proteins (Link, 1999; Springer, 1996). After electrophoresis, signal detection may be accomplished using film, direct autoradiography (for the γ-emitting isotopes) or by fluorography (for the β-emitting isotopes). Autoradiography has the ability to detect proteins below the pg range. Therefore, this is one of the most sensitive detection techniques. Fluorography employs fluorescent enhancers impregnated into the gel matrix to improve detection of the radioactive emissions through the generation of light. Intensifying screens may be used to enhance the signal of high-energy isotopes such as ^{32}P or ^{125}I (Link, 1999). These screens absorb the radioactive emissions that pass through the film, fluoresce and generate light that exposes the film. Over the past decade, fluorescent imagers and microchannel plate analysers have begun to replace autoradiography film for detecting radiolabelled proteins (Link, 1999). The new technologies offer a wider dynamic range and better detection sensitivity than film.

Although sensitive detection is achieved, radiolabelling is both hazardous and expensive. Expenditures involved include disposal of the radioactive waste, inventory control, training laboratory personnel and monitoring their exposure. Reducing reliance on radioactivity in biomedical research had become a high priority in many laboratories over the past several years as it makes economic good sense and provides a safer work environment. Autoradiography has 3 of the 10 desirable staining properties listed on page 39, these being that it has a wide linearity of range, has a wide dynamic range and is potentially the most sensitive technique available for the detection of low copy number proteins.

1.3.9 Computer Analysis

Scanning

2-DE-gels can be used to resolve thousands of proteins from one cell type or organism. This mass of information makes it difficult, if not impossible to discern all the differences between samples with the unaided eye, particularly when low abundance proteins may barely be visible and the eye is poor at quantitating. Therefore, a scanning densitometer or digital camera must be used to prepare quantitative, digitised gel images to be analysed with a computer. The best results may be achieved by acquiring the gel images as a grey-scale TIFF file with adequate resolution and intensity measured preferably using a 16 bit scale providing 2^{16} levels of intensity (i.e. over 65,000) (Westermeier and Naven 2002). The ability to utilise a 16-bit intensity scale is conditional on the ability of the analytical software to access this format. However, the use of a 16-bit scale results in large files (typically 14 Mb for 18.5 x 18.5 cm gel). This may make it difficult to analyse a large number of image files, and therefore it may be more advisable to use an 8-bit scale, which produces much smaller files (typically 1.4 Mb for 18.5 x 18.5 cm gel). In addition increases in dots per inch or bit scale may not necessarily improve the degree of information obtained from the gel images to be analysed.

Image Analysis

Analysis of complex 2-DE patterns has been facilitated by the development and utilisation of sophisticated image analysis software. In this manner, one can detect and quantify even faint spots, quantitatively compare 2-DE gel images and identify protein expression changes across sets of 2-DE gel images (Link, 1999).

There are three types of 2-DE imaging software available (some programs may fall into more than one category):

- public domain tools for basic image analysis e.g. Flicker http://www-lecb.ncifcrf.gov/flicker/
- professional programs which evaluate with high reproducibility and offer many valuable functions for image analysis, statistical evaluation and data reporting e.g. PDQUEST and Melanie.

- fully automated software solutions integrated with robotic hardware for high-throughput hands-free analysis involving a large number of gels e.g. Progenesis, PDQUEST (Westermeier and Naven 2002).

The development of software for 2-DE gel image analysis is a continuously evolving process. The functions become more reliable, reproducible and automated from year to year. With the latest programs, it is already possible to compare gels of different sizes, shapes, and even damaged gels. However, if the protein spots on the gels are not reproducible, even the most advanced software cannot fix the problem. Currently the most advanced software packages still require a certain degree of manual correction of protein matches. The day of fully automated software analysis is approaching, but it is not here yet.

1.3.10 Types of Analyses

PDQUEST Analysis

An important part of databasing is the ability to compare data between experiments. The spots in the PDQUEST program consist of Quantitative, Qualitative, Boolean, Arbitrary and Statistic. This enables Analysis sets to be created which contain statistically and biologically related spots.

Arbitrary analysis sets are composed of any group of spots selected by the program user (PDQUEST manual 6.2.1).

A Boolean analysis set is formed by combining two previously defined Analysis Sets (A and B) using Boolean operators. The Boolean Sets created might include: 1) the intersection of A and B; 2) the union of A and B; 3) spots that are unique to set A plus spots that are unique to set B; 4) Spots that are only found in B; and 5) spots that are only found in A (PDQUEST manual 6.2.1).

A quantitative analysis set contains spots from two different images or replicate groups that are present in both images or replicate groups that differ in concentration to a predetermined degree (usually by a factor of 3 or more) i.e. it includes spots whose volume has significantly increased or decreased (PDQUEST manual 6.2.1).

A qualitative analysis set contains spots from two different images or replicate groups that were detected in one image or replicate group, but not in the other. Such a set might contain, for example, proteins that were expressed under experimental conditions but not expressed under control conditions, or vice versa (PDQUEST manual 6.2.1).

A statistical user set is composed of spots from replicate groups whose volumes are found to be significant according to a specified statistical test (e.g. Student T-Test, or Mann-Witney Test) (PDQUEST manual 6.2.1).

Cluster analysis

The aim of classification is to group together a number of objects based on their attributes or variables so that the object is more similar to other objects in its group than objects in the other groups (Everitt, 1972; Everitt, 2004). Two methods that may be used to classify objects are discriminant analysis and cluster analysis. Discriminant analysis requires that the parameters used to define group membership are known for the cases used to derive the classification rule. However, in cluster analysis the parameters used to define group membership for all cases is unknown. In fact, even the number of groups is often unknown. The goal of cluster analysis is to identify homogeneous groups or clusters.

In biology, cluster analysis is often used to classify animals and plants. This is called numerical taxonomy. In medicine, cluster analysis is often used to identify diseases and their stages. For example, by examining patients who are diagnosed as depressed, it might be found that there are several distinct subgroups of patients with different types of depression (Everitt, 1972; Everitt, 2004). Two of the most commonly used types of cluster analysis are k-means and hierarchical clustering techniques (Everitt, 1972; Everitt, 2004).

K-means Clustering attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. However, the algorithm requires the number of clusters to be specified. The

number of clusters can only be specified if this information is known (SPSS Manual Version 11).

In a **hierarchical** classification, the data are not partitioned into a particular number of classes or clusters at a single step. Instead the classification consists of a series of partitions, which may run from a single cluster containing all individuals, to *n* clusters each containing a single individual. Hierarchical clustering techniques may be subdivided into agglomerative methods that proceed by a series of successive fusion of the *n* individuals into groups, and divisive methods, which separate the *n* individuals successively into finer groupings (SPSS Manual Version 11).

Agglomerative hierarchical cluster analysis A method for creating clusters in which each case starts out as a cluster. In the present instance, a case is equivalent to a *H. pylori* strain and a variable is equivalent to a spot. At every step, clusters are combined until all cases are members of a single cluster. Once a cluster is formed it cannot be split, it can only be combined with other clusters (SPSS Manual Version 11).

The application of most methods of cluster analysis depends on the availability of suitable computer software. There is a large range of cluster analysis software packages available, these include SPSS, SAS, SYSTAT, CLUSTAR and CLUSTAN. These programs differ in the types of statistical analysis that can be performed ranging from a wide range of applications (SPSS and SAS) to programs that are specially designed for cluster analysis (SYSTAT, CLUSTAR, CLUSTAN).

1.3.11 Protein identification methods in proteomics

Spot Volume is a measure of protein quantity related to staining intensity. The spot volume is determined from the average spot intensity (mean density of pixels in the spot) by the spot area (number of pixels in a spot) (PDQUEST Manual).

A protein cannot be identified based on its position in a gel alone. Identification is only possible by further analysis of the protein, for instance using mass spectrometry. Furthermore it is incorrect to conclude that a protein is up or down regulated on the

basis of an increased or decreased spot volume because the protein may have changed its pI due to post-translational modification and may now be co-migrating with another protein. Therefore, proteins of interest have to be analysed further. This is done mainly by mass spectrometry (MS). Generally, MS data can be used in four approaches for protein identification. These approaches are peptide mass fingerprinting, peptide mass fingerprinting and composition information, peptide mass fingerprint and sequence information, and MS/MS sequencing data.

Other methods, such as Edman degradation, were the principal methods of protein sequencing in the eighties but have largely been superseded by mass spectrometry sequencing.

Peptide mass fingerprinting (PMF) is a technique for searching protein databases for protein identity. The subject protein is cleaved, and the masses of the resultant peptides are used for a database search (James et al, 1993; Westermeier and Naven, 2002).

PMF and composition information is a technique where each peptide's molecular mass combined with its composition information can be used for a database search (Westermeier and Naven, 2002).

PMF and sequence information for each peptide can be used along side some of the direct amino acid sequence information to search a peptide sequence database (Westermeier and Naven, 2002).

The product ion MS/MS sequence data from one or more peptide – MS/MS mode (Isobe et al 1990; Westermeier and Naven, 2002).

Spot Cutting. The gel plugs containing the proteins of interest are excised from the gel after image analysis. Cutting spots from a gel manually is a very tedious job. The spots of interest, marked on the gel image printout for further analysis, have to be matched to the actual gel and transferred to the correct tube or a microtitre plate. Additionally, contamination with exogenous proteins such as skin keratins must be

avoided. Robot spot pickers not only increase the speed of the excision process but also minimise contact with the gel and consequent keratin contamination from hair, skin and dust particles (Lopez, 2000).

Protein Cleavage, proteolysis. In this approach the proteins have been purified, in many cases, by 2-D gel electrophoresis and are either enzymatically or chemically cleaved. An aliquot of the peptide mixture is then analysed by mass spectrometry. The proteins are digested either in-gel or after transfer to membranes (PVDF or nitrocellulose) with an enzyme or chemical reagent which specifically cleaves at certain amino acid residues. The exact masses of the peptides that result from this cleavage are then measured in a mass spectrometer. A commonly used enzyme is trypsin that cleaves only at the C-terminal side of arginine or lysine (Wilkins et al, 1997; Link, 1999). However, there may be occasions where digestion with an alternative enzyme will be advantageous, specifically in the case of post-translational modifications, and small proteins, which may have few trypsin cleavage sites (Fountoulakis et al, 1997).

Chemical methods of protein cleavage are not commonly used. When these methods are applied, it is often for specific applications where no suitable enzyme is available or enzymatic cleavage is not appropriate. An example is cyanogen bromide cleavage of insoluble or membrane proteins. Cyanogen bromide cleaves specifically at methionine residues, often yielding relatively large peptides (Westermeier and Naven, 2002).

Mass Spectrometry. Mass spectrometry is an analytical technique that measures the molecular mass of molecules based upon the motion of a charged particle in an electric field (Westermeier and Naven, 2002; Link, 1999). Sample molecules are converted into ions in the gas phase and separated according to their mass:charge ratio (*m/z*). Positive and negatively charged ions can be formed (Westermeier and Naven, 2002; Link, 1999).

Biological MS is fast and very sensitive. The two main lines of technological development that have dominated the field to date are based upon two different

methods of ionising analytes, matrix-assisted laser desorption ionisation/time of flight (MALDI-TOF) MS and electrospray ionisation (ESI) MS (Wilkins et al, 1997). Consequently, associated sample preparation procedures are substantially different. While ESI is mainly linked to liquid chromatographic instrumentation, MALDI-TOF is more suitable for high-throughput approaches and has until recently been better suited for large-scale proteomics (Wilkins et al, 1997).

In comparison to other ionisation techniques such as ESI, MALDI tolerates moderate buffer and salt concentrations in the analyte mixture and almost exclusively produces singly charged ions (Wilkins et al, 1997). For these reasons it has become the preferred ionisation technique for PMF analysis.

1.3.12 Bioinformatics

To date (June 2006), the complete genome of at least 409 organisms have been sequenced. The gene sequence data is available at http://www.tigr.org/tdb/mdb/mdbcomple.html. Comprehensive studies at the proteome level have been published for *Mycobacterium tuberculosis* complex, *H. pylori*, *Haemophilus influenzae*, *Mycoplasma pnuemoniae*, *Pseudomonas aeruginosa*, *Staphylococcus aureus,* and *Escherichia coli*.

In sequence homology, proteins in the database are theoretically digested with the relevant cleavage reagent, generating many hundreds of thousands of theoretical peptides resulting in a "virtual fingerprint". Each protein of interest under investigation is digested to give an actual PMF. The actual PMF data derived from the mass spectrometry experiment is subsequently compared to these "virtual fingerprints" and the best theoretical/actual match is retrieved as the possible candidate proteins.

Several on-line programs are available to perform this type of search, (including MASCOT at www.matrix-science.com, Profound at www.prowl.com; MS-FIT at www.prospector.ucsf.edu). Accuracy, reliability and speed will vary depending on the program chosen. Regardless of which program is used, four user variables are important for a PMF search. These are

- peptide mass range;
- specification of the cleavage agent;
- error tolerance - the accuracy of mass measurement is determined by the calibration, the more accurate the mass spectrometer the greater the specificity; and
- knowledge of amino acid residue modifications likely to be encountered e.g. methionine oxidation.

In combination, high resolution and high mass accuracy allow the probable identification of proteins present in databases from as few as 3-5 peptides, depending on the size of the protein. However, for this approach to give an unambiguous result, ≥ 6 peptides of experimentally determined peptide masses should match the theoretical masses (Enroth et al, 2000). The number of matched peptides was selected to cover at least 15% of the total amino acid sequence for high molecular mass proteins (≥ 100 kDa) and 60% of low molecular mass proteins (≤ 30 kDa). Low molecular mass proteins have fewer cleavage sites than high M_r proteins (Fountoulakis et al, 1998) and as such may result in an increased number of matches with other proteins with similar peptide masses, therefore a sequence match of ≥60% is required. In high M_r proteins there is a greater number of peptide masses which reduces the probability of matches with other proteins.

Each on-line search engine is capable of searching a variety of databases such as NCBI and SWISS-PROT. However, each database differs in the number of proteins it lists. For example by March 2003, the NCBI database has over 6000 *H. pylori* proteins listed whereas SWISS-PROT has only about 550. The advantage of SWISS-PROT is that its entries are very thoroughly annotated. However, given that a species like *H. pylori* which has approximately 1600 ORFs a search of SWISS-PROT alone would mean that many peptide mass fingerprints identities would be missed.

1.4 Previous Studies versus the Present Study

1.4.1 Jungblut's study of H. pylori

Jungblut et al (2000) was the first group to conduct a comparative 2-DE proteome analysis of *H. pylori* strains. This study performed a preliminary global comparison of proteins from three different strains of *H. pylori* to identify immunodominant antigens and to compare effects of different biological conditions. The Jungblut group separated *H. pylori* (whole cell) proteins extensively by 2-DE gel electrophoresis and 152 proteins were identified by mass spectrometry. The resulting 2-DE map is available at http://www.mpiib-berlin.mpg.de/2D-PAGE/. Jungblut compared three *H. pylori* strains, 26695 (a gastritis strain), J99 (a duodenal strain) and the 'Sydney strain' SS1 (a prominent strain used in animal models). Approximately 1800 proteins were resolved on large format 2-DE gels (23 cm x 30 cm), and proteins were separated over the pI range of 4-10 and the M_r range of 5-150 kDa and stained with silver. In addition to the comparative analysis of these three strains, Jungblut et al (2000) also studied the effects of different pH conditions on protein expression of *H. pylori* strain 26695 and identified the immunodominant antigens of this strain. Strain 26695 was examined after culture in four different media at pH 5, 6, 7 and 8 and the proteins were separated using small 2-DE gel (size dimensions were not mentioned in the article).

Jungblut's group found that there was a high degree of proteomic variability between the three strains included in their study. This contrasted with the genomic comparison of J99 and 26695 by Alm et al (1999) which indicated that there was a high degree of conservation (85%) in genomic organisation and gene order of these two strains. The Jungblut study found that a number of proteins were subject to positional pI shifts and identified at least four proteins with amino acid substitutions, which affected the pI of these proteins. These proteins were alkyl hydroperoxide reductase, superoxide dismutase, ribosomal protein L7/L12 chaperonin and GroEL.

1.4.2 Enroth's study of H. pylori

Enroth et al (2000) was the second group to perform a comparative analysis of *H. pylori* isolates. The Enroth study set out to determine whether *H. pylori* isolates isolated from three different disease groups could be grouped according to their

protein patterns. This study also searched for disease-specific protein spots, which might be useful for *H. pylori* strain characterisation (protein markers). Twelve different *H. pylori* isolates from three different patient groups (gastritis, duodenal ulceration and adenocarcinoma) were extracted and analysed by 2-D PAGE. Enroth studied the whole cell protein extracts and water extracts of *H. pylori* isolates by intermediate precast gradient 2-DE gels (Bio-Rad Criterion format 110 mm width). Proteins were separated over the pI range 4-7 (water extracts) and 3-10 (whole cell extracts) and the M_r range of 5-150 kDa. Enroth's (2000) study compared twelve silver stained gels (i.e. a single gel per strain, four isolates for each disease group) by the use of specialised analytical software (BioImage 2-DE analyser version 6.1) and through the application of cluster analysis. Eight spots were identified that were absent in certain isolates or significantly different in intensity between isolates.

Cluster analysis of the twelve isolates was conducted by two different methods: i.e. average linkage and neighbour joining. Cluster analysis utilised the number of spots matched i.e. qualitative rather than quantitative spot information. The Enroth (2000) study found that some *H. pylori* isolates might be more associated with a specific disease than other isolates but that this may not apply to all disease related isolates.

In addition, this study compared the immunodominant proteins of one adenocarcinoma strain and one duodenal ulceration strain through the application of large format gels (18 x 18 cm).

The Enroth study found that less than 50 percent of the proteome could be matched between any two disease groups. However, Enroth's group only visualised approximately 600 spots on gels stained by silver. This number was less than the 2000 proteins visualised by Jungblut (silver stained gels). The reason for the reduction in the number of proteins visualised was due to the use of small format 2-DE gels, which would have decreased the spatial resolution of the protein spots. This would have resulted in co-migration of many protein spots, which would alter quantitative spot information and many spots may appear to be absent or unique spots may not have been resolved.

1.4.3 Current study of H. pylori

This current study compares eleven *H. pylori* isolates isolated from patients with DU, GU and NUD as well as GI isolates (NCTC11637 and 26695) through the application of proteome analysis and cluster analysis. Each disease outcome group consisted of three isolates except for gastritis that consisted of two isolates. This study aims to determine whether *H. pylori* isolates can be sorted into disease groups according to their protein patterns. Whole cell protein extracts of *H. pylori* isolates were examined using large format (18.5 cm x 18.5 cm) 2-DE gels using Tris-tricine and Tris-glycine buffer systems. Proteins were separated over the pI range 4-11 and the M_r range 5-150 kDa.

The study incorporated a minimum of four gels for each isolate under examination. There are 44 gel images included in the comparative image analysis for each protocol, with exception of Tris-glycine pH 6-11 which contained 40 gel images. An important feature of this study is the application of replicate groups. Note that the Enroth (2000) study did not appear to utilise replicate groups. It consisted of twelve gels: only one for each strain in the comparative analysis. Neither was it apparent from the paper by Jungblut et al (2000) if replicate groups were utilised in their study. There are two important reasons for the use of replicate groups. The first reason is to confirm the validity of spots. In order to be accepted as valid, the current study required that a particular spot should be observable in at least three out of four gel images for each replicate group. The second reason is because of slight variation in quantitation due to experimental variables (i.e. loading and staining). Because of this, the volume of a given spot is best estimated by averaging results from several gel images.

The present study is differs from the previous two studies in a number of ways:

- compares eleven rather than the three isolates examined by Jungblut et al (2000);
- the current study compares four disease groups;
- it is the first study to compare gastric ulceration (GU) and non-ulcer dyspepsia (NUD); neither of these disease groups has been included in any other published study to date (December 2005);

- it greatly improves upon the spatial resolution and the number of observable proteins studied by Enroth et al (2000); and
- in addition, this study separates proteins in a pH range (10-11) which has not been previously examined by the other *H. pylori* comparative proteome studies.

1.4.4 Comparisons between H. pylori *isolates*

The number of spot matches between gels that may be obtained from a 2-DE proteome analysis of *H. pylori* is dependant on a number of factors. These include:

1. the number of isolates that are included in the study;
2. the degree of relationship between the isolates;
3. the resolution of the proteome maps; and
4. the staining technique utilised.

Isolates included in this study and their relationship

Two variables that may affect the percentage of spots matching between isolates are the number of isolates that are included in an *H. pylori* study and the degree of relationship between the isolates. If two isolates cause the same disease it is likely that they will share a greater percentage of common proteins (e.g. virulence proteins) than isolates associated with dissimilar outcomes. Conversely, isolates from different disease groups might produce a lower percentage of matches due to differences in virulence factors.

The current study compares four disease groups. Three disease groups (GU, NUD and DU) consisting of three isolates each and a fourth (GI) containing two isolates. This will expand the study that was performed by the Jungblut et al (2000) group, but is one isolate short of the number included in the Enroth et al (2000) study. The Jungblut group consisted of two isolates from two different disease groups, duodenal ulceration (J99) and gastritis (HP26995), and a third isolate which was the mouse-adapted *H. pylori* strain `Sydney strain' SS1 (Lee et al., 1997). The Enroth group had twelve isolates from three different disease groups, gastritis, duodenal ulceration and gastric adenocarcinoma.

Laboratory history of isolates

Putative virulence factors of *H. pylori* include colonisation proteins, such as those involved in motility, adhesion and cytotoxocity. A study by Sung et al (2002) examined the effect of repeated serial laboratory subculturing of *H. pylori*. This study found that, following the 64th serial subculturing of strain SS1, adhesion, motility and cytotoxicity had all decreased in comparison to the original culture. In addition, gastric inflammation produced in a Mongolian gerbil by the final subcultured strain was less severe (Sung et al 2002). Therefore, given that GI1 and GI2 are gastritis isolates which have been maintained for a long period in laboratory culture it might be argued that these may give different results to those seen in the Enroth (Enroth et al, 2000) study. However, the Microbiology Department at the Royal Perth Hospital has strict guidelines to limit the subculturing of *H. pylori* and therefore repeated subculturing is unlikely to be a factor in this study.

Spot resolution

The resolution of proteins influences the results obtained from any proteome study. The greater the spatial resolution (both with respect to pI and M_r) on 2-D gels the greater the possibility that slight differences in pI and M_r will be observable. However, if the resolution area is not large enough this will result in spots comigrating in both the x- and y-dimensions. Overlapping spots, which may be a mixture of two or more proteins, make it difficult to obtain an accurate estimate of the quantitative and/or qualitative differences between isolates (Westbrook et al, 2001). Increasing the spatial resolution of the protein spots increases the number of spots visualised, provides additional data for comparative pattern analysis and subsequent protein identification (Gőrg et al, 2004; Westbrook et al, 2001). The quality of comparative pattern analysis is dependent both on the complexity of the 2-D gel images (e.g. the proteome of human cell lines is more complex compared to the comparative simplicity of the *H. pylori* proteome) and the quality of the program that analyses the images (e.g. advanced ProGensis, PDQUEST versus relatively unsophisticated freeware such as Flicker (see page 45).

The resolution area is also an important factor in the detection of low abundance proteins. The proportion of the total protein complement that can be seen on a gel for any cell or tissue will depend on the protein copy number per cell, on the quantity of material loaded on the gel and on the method of detection (Wilkins et al, 1997). Low abundance proteins are often proteins of great interest, but difficult to detect on 2-D gels because more abundant proteins predominate. Abundant proteins (e.g. albumin in serum samples) have the appearance of large high-stain-intensity spots. Enrichment of specific subcellular fractions for low abundance proteins prior to the loading of the sample onto the first dimension will increase the number of detectable low-abundance proteins (Langen et al, 2000). One procedure is sequential extraction of proteins from whole cell lysates or subcellular subfractions according to their differing solubilities (Molloy et al, 1998). This is even more effective if proteins are then separated with narrow range pH IPG strips.

1.5 Aims of the Investigation

Primary aims of this study

- To determine whether *H. pylori* isolates can be sorted into disease groups according to their protein patterns.
- To identify qualitative and quantitative protein differences between the disease groups.

Secondary aim of this study

- To further expand the number of *H. pylori* proteins identified in previous studies.
- To gain an insight into the proteins that may be involved in the disease groups e.g. virulence factors.

2.1 Materials

Reagent	Distributor
Acetic acid	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Acetone	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Acrylamide 40%	Plus One Pharmacia Biotech Upsala, Sweden
Agarose	Sigma Chemical Co St Louis, USA
Ammonium persulphate	Bio-Rad Laboratories Hercules, CA, USA
Ammonium sulphate	ICN Biomedicals Inc Aurora, Ohio, USA
Ampholytes	Amersham Biosciences Upsala, Sweden
Bis-acrylamide	Bio-Rad Laboratories Hercules, CA, USA
Bromophenol Blue	Sigma Chemical Co St Louis, USA
Campygen plastic pouches	Oxoid Ltd Detroit, MI, USA
Campygen paper sachets	Oxoid Ltd Detroit, MI, USA
3-(3-cholamidoproyl)-dimethylammonio-1-propane sulfonate	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Coomassie Brilliant Blue G250	Bio-Rad Laboratories Hercules, CA, USA
Endonuclease	Sigma Chemical Co St Louis, USA
Filter paper wicks	Bio-rad Laboratories Hercules, CA, USA

Reagent	**Distributor**
Formaldehyde	Sigma Chemical Co St Louis, USA
Glutaraldehyde	Sigma Chemical Co St Louis, USA
Glycine	Bio-Rad Laboratories Hercules, CA, USA
Glycerol AnalaR®	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
HCl	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Immobilised pH gradient gel strips	Pharmacia Biotech Upsala, Sweden
Isobutanol	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Lint free tissues	Kimberly Clark Milsons Point, NSW, Australia
Methanol	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Orange G	ICN Biomedicals Inc Aurora, Ohio, USA
Paraffin Liquid	Ramprie Laboratories Welshpool, WA, Australia
Parafilm	Pechiney Plastic Packaging Chicago, IL, USA
Piperazine diacylamide	Bio-Rad Laboratories Hercules, CA, USA
Pyroneg	Diversey Sydney, Australia
Phosphoric acid 85%	BDH Laboratory Supplies Poole, England

Reagent	**Distributor**
Polyethylene glycol	Sigma Chemical Co St Louis, USA
Ready-cast gradient gels	Bio-rad Laboratories Hercules, CA, USA
Sulfobetaine 3-10	Sigma Chemical Co St Louis, USA
SDS	Bio-Rad Laboratories Hercules, CA, USA
SDS M_r standards	Sigma Chemical Co St Louis, USA
Silver nitrate AnalaR	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Sodium acetate anhydrous AnalaR®	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Sodium azide	ICN Biomedicals Inc Aurora, Ohio, USA
Sodium carbonate anhydrous AnalaR®	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Sodium chloride	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Sodium glutamate	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Sodium thiosulphate	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Tributyl phosphine	ICN Biomedicals Inc Aurora, Ohio, USA
N,N,N',N'- Tetramethylethylenediamine (TEMED)	Plus One Pharmacia Biotech Upsala, Sweden
Thiourea	BDH Merck Pty Ltd Kilysyth, Victoria, Australia

Reagent	**Distributor**
Tris(hydroxymethyl)-aminoethane	Sigma Chemical Co St Louis, USA
N,tris(hydroxymethyl)-methyl glycine	BDH Merck Pty Ltd Kilysyth, Victoria, Australia
Urea AnalaR®	BDH Merck Pty Ltd Kilysyth, Victoria, Australia

2.2 List of Solutions

All buffers solutions and agars were prepared using Millique high pure water. All buffers and solutions were freshly made for day of use, except were specified.

Tris-glycine Buffer Systems (section 2.8.3)

This buffer system was utilised for proteins>20 kDa.

5x Tris/HCl Buffer

1.9 M Tris

Adjusted to pH 8.8 with conc. HCl

Stored at -20^{O}C

PDA-Acrylamide Stock Solution (2.5% crosslinker)

200 ml 40 % acrylamide solution

51 mM Piperazine diacrylamide

Anode Electrode Buffer

0.75 M Tris

0.77 mM Sodium azide

Adjusted to pH 8.8 with conc. HCl

Cathode Electrode Buffer

0.19 M Glycine

3.5 mM SDS

10% Polyacrylamide Gel

27 ml H_2O

10 ml 5x Tris/HCl Buffer

12.5 ml Acrylamide stock solution

250 µl 10% Ammonium persulphate

10 µl Sodium thiosulphate

16.5 µl TEMED

Tris-tricine Buffer System (section 2.8.4)

This buffer system was utilised for proteins < 20 kDa

5x Gel buffer

2.9 M Tris

8.7 mM SDS

Adjusted to pH 8.45 with conc. HCl

Stored at -20^{O}C

Cathode buffer

0.1 M Tricine

3.5 mM SDS

Adjusted to pH 8.3 with Tris

Anode buffer

0.19 M Glycine

3.5 mM SDS

Adjusted to pH 8.3 with Tris

Bis-Acrylamide gel Stock solution (6% crosslinker)

200 mls Acrylamide (40%)

0.16 M Bis

8.5% Polyacrylamide Gel

9 ml 5x Gel Buffer

12 ml Acrylamide stock solution

3.1 M Urea

400 µl 10% Ammonium persulphate

400 µl Sodium thiosulphate

59 µl TEMED

H_2O make up to 50 ml

Miscellaneous Solutions

Multiple surfactant solution (5 ml) (section 2.5)

40 mM Tris

33 mM CHAPS

65 mM SB3-10

2 M Thiourea

5 M Urea

3.2 ml H_2O

25 µl Ampholytes

2.5 µl TBP

5 µl Orange Dye

SDS Equilibration solution (100 ml) (section 2.8.5)

20% 5x Tris/HCl Buffer (v/v)

6.7% Acrylamide (40%) (v/v)

20% Glycerol (v/v)

6 M Urea

70 mM SDS

125 µl TBP

make up to 100 ml with H_2O

1x Tris/HCl Buffer

5 ml 5x Tris/HCl Buffer

20 ml H_2O

Stains for Gels (section 2.10)

Colloidal Coomassie G250 Stain

1.3 M Ammonium sulphate

3.6% Phosphoric acid (85%) (v/v)

34% Methanol (v/v)

1.17 mM CBB G-250

Silver Staining Solutions

Fixer 1

40 % Methanol (v/v)

10% Acetic acid (v/v)

Fixer 2

0.55 M Sodium acetate

2 g Sodium thiosulphate

30% Methanol (v/v)

0.5% Glutaraldehyde (v/v)

Silver

11.8 mM Silver nitrate

Developer

0.28 M Sodium carbonate anhydrous

0.1% Formaldehyde (v/v)

Stop

5 % Acetic acid (v/v)

Drying Solution (section 2.11)

50% Methanol (v/v)

2.5 mM PEG

Table 2.1 Isolates included in this study

Strain	Isolate Characteristics		Patient Characteristics			
	Metronidazole	Biopsy Year	Disease Outcome	Age	Gender	Country of origin
GU1	NA	2001	Gastric Ulceration	50	Male	UK
GU2	Resistant	1997	Gastric Ulceration	58	Female	UK
GU3	Resistant	1997	Gastric Ulceration	46	Male	Yugoslavia
DU1	Resistant	1996	Duodenal Ulcerations	33	Male	UK
DU2	Sensitive	1997	Duodenal Ulcerations	66	Male	Australia
DU3	Sensitive	1997	Duodenal Ulcerations	32	Male	Hong Kong
NUD1	Sensitive	1997	Non-Ulcer Dyspepsia	47	Male	Australia
NUD2	Sensitive	1997	Non-Ulcer Dyspepsia	64	Female	UK
NUD3	Resistant	1996	Non-Ulcer Dyspepsia	64	Female	Australia
GI1	NA	1982	Gastritis	NA	NA	Australia
GI2	NA	1985	Gastritis	NA	NA	UK

NA = Not Available

2.3 Bacterial isolates

There were eleven *H. pylori* isolates in this study. They were obtained from the Department Microbiology and Infectious Diseases at Royal Perth Hospital. All of these isolates had originally been obtained from patient biopsies. The isolates are divided into four disease outcomes: GU, DU, NUD and GI (see Table 2.1). Each disease outcome group consisted of three isolates except for gastritis that consisted of two isolates. The nine isolates comprising the GU, DU and NUD groups were obtained recently and had been subjected to minimal passage in the laboratory. The other two isolates NCTC11637 (GI1) and 26695 (GI2) had been extensively cultivated in the laboratory since 1982 (Goodwin and Worsley, 1993a and 1993b) and 1987 (Alm et al, 1999), respectively. A large proportion of the proteomes of isolates NCTC11637 (Lock et al, 2001) and 26695 (Jungblut et al, 2000; Myung-Je et al, 2001), had already been evaluated, therefore these were used as reference isolates. None of the other nine isolates had previously been analysed in any published study.

2.4 Culture conditions

Each *H. pylori* stock culture was prepared by inoculation into 1 ml of horse blood and then stored at –80^{o}C. When *H. pylori* isolates were required, subcultures were obtained from the stocks by inoculation onto gonococcal heated blood (chocolate) agar plates. Bacterial cells were incubated at 37^{o}C in Campygen plastic pouches and Campygen paper sachets for a period of three days.

Confluent growth on at least three chocolate agar plates was needed to obtain sufficient cell numbers for adequate protein extraction. To achieve confluent growth, plates were subcultured at least three times. For a preparation of whole cell proteins, cells were removed from three confluent plates and placed in a vial containing 1 ml 20% sodium glutamate. The bacteria were then lyophilised and stored at –80^{o}C. Each vial contained about 20 mg of bacterial cells (dried weight).

2.5 Sample preparation

A single sample preparation method starting with 1, 2 or 3 vials was tested. The vial(s) of each strain were rehydrated in a total of 3.2 mls of a 0.48% Tris-base solution then sonicated for 12 x 60 sec bursts at 50% duty cycle with a power setting of 6 using a Branson Sonifier 250 sonicator (Danbury, CT, USA) fitted with a microtip probe. Nine volumes of cold methanol were added to the sonicated sample to precipitate the proteins and the material was stored overnight at -80 ^{o}C. Each protein sample was placed in 1.5 ml centrifuge tube and centrifuged for 1 hour at 0 ^{o}C at 21,000 *g* using a Beckman microfuge (Fullerton, CA, USA). Pellets were air dried at room temperature for 40 minutes then resuspended in a combined total of 5 mls multiple surfactant solution (MSS) and stirred for 1 hour at room temperature. In the final 15 minutes, 2 µl (250 units per µl) of endonuclease was added to each tube. The resuspended sample was centrifuged at 15 ^{o}C for 1 hour at 21,000 *g*, and the pellets discarded. The supernatants were divided into 500 µl aliquots and stored at -80 ^{o}C until required.

2.6 Wash-up

Scrupulously clean glassware and apparatus is essential in electrophoresis as contaminants may affect the migration of proteins, the staining of gels and the subsequent identification of proteins.

Glass Plates. Glass plates were thoroughly scrubbed in deionised water and liquid Pyroneg. All previous traces of polyacrylamide were removed. Prior to use glass plates were washed with a small amount of methanol and wiped clean with low-lint tissues.

Glass Gel-Staining Dishes. Following each use, sealable flat glass dishes were thoroughly washed using liquid Pyroneg and deionised water. After the majority of the stain was removed from the glass dishes, a small amount of methanol was used to remove any residual trace of Coomassie G250. Finally, glass dishes were carefully wiped with low-lint tissues.

2.7 First dimension electrophoresis (isoelectric focusing)

IEF was performed using 18 cm IPG strips with pH ranges of either 4.0-7.0 or 6.0-11.0. Rehydration of strips with sample was performed as follows: both ends of a Sarstedt (Numbrecht, Germany) 2ml disposable plastic pipette were clipped off and one end sealed with parafilm. An IPG strip was placed in the pipette and 500 µl of protein sample (in MSS) was introduced in such a way as to remain in contact with the upper (gel) surface of the strip. The other end of the pipette was sealed with parafilm and strips were left to rehydrate for 24 hours at room temperature.

IEF was performed using a Pharmacia Multiphor II (Pharmacia Biotech, Upsala, Sweden) apparatus. After rehydration was completed, strips were placed in the channels of an Immoboline dry strip kit (Pharmacia Biotech, Upsala, Sweden) on the Multiphor II. Damp (with H_2O) filter paper wicks were placed at the acidic and basic ends of the IPG strips, electrodes were placed in contact with the wicks and the tray was covered in paraffin oil prior to electrophoresis. Strips were electrophoresed using three programmed phases (Lock et al, 2001) for each method. The first method utilised a PowerPac 3000xi (Bio-rad, Hercules, CA, USA) and second method a power supply EPS 3501 XL (Amersham-Pharmacia, Upsala, Sweden). Initially two alternative methods were investigated.

1st Method

Phase	Constant	Time	Safety Limit (per six strips)
1	300 V	2 1/2 hrs	1 mA/1 Watt
2	1,000V	2 1/2 hrs	1 mA/1 Watt
3	3,000V	20 hrs	1 mA/1 Watt

Total Volt hours 63,200

2nd Method

Phase	Constant	Time	Safety Limit (per six strips)
1	300 V	2 1/2 hrs	1 mA/1 Watt
2	1,000V	2 1/2 hrs	1 mA/1 Watt
3	3,500V	21 1/2 hrs	1 mA/1 Watt

Total Volt hours 78,450

Generally the current did not exceed 150 μA per strip. After IEF, strips were stored at –80^{o}C in a plastic Petri dish sealed with parafilm until required for the second dimension separation.

2.8 Second dimension electrophoresis (SDS-PAGE)

This study examined the separation of proteins in the second dimension using two different protocols (see Figure 2.1). The first used gradient polyacrylamide gels. The second involved homogeneous (single concentration) polyacrylamide gels using two different buffer systems, the Tris-glycine buffer system and the Tris-tricine buffer system.

2.8.1 Gradient Gels

Gradient gels have an increasing concentration of acrylamide from top to bottom, for example in a 10-20% gradient gel a 10% polyacrylamide solution at the top increases to 20% at the bottom of the gel. With gradient gels, the overall separation interval is wider, and the linear relation interval between the logarithm of relative molecular mass (M_r) and the migration distance is wider than for homogeneous gels. Also the pore sizes are continuously decreasing (Westermeier and Naven, 2002). The manual preparation of gradient gels is much more time consuming than for homogeneous gels, and it is more difficult to obtain reproducible gel properties. To obtain reproducible gradient gels this study utilised commercially available Bio-Rad Ready-cast™ (Hercules, CA, USA) gradient gels *T* 10-16% (T = total polyacrylamide concentration). The gradient gels were run at a constant 25 mA per gel for 8 1/2 hours at 10^{o}C.

2.8.2 Preparation of Homogeneous (Single-Concentration) Gels

Homogeneous slab gels were prepared manually. They were poured either using a Bio-Rad (Hercules, CA, USA) gel-casting stand (which makes one or two gels per stand) or a Bio-Rad Multi-Caster (Hercules, CA, USA) (which makes up to 12 gels).The glass plates between which the gels were poured consisted of an outer plate

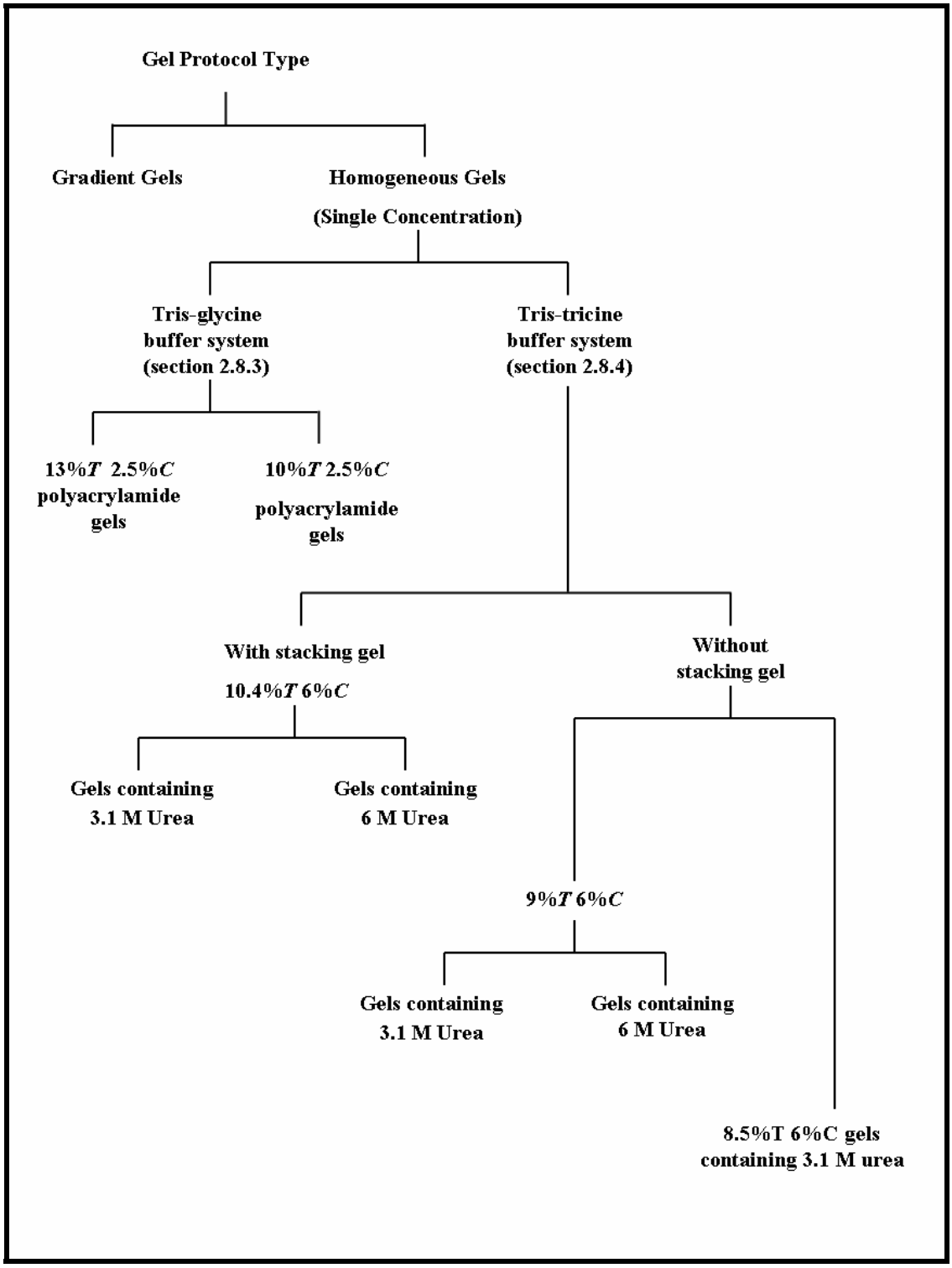


Figure 2.1 Second dimension protocols
Diagrammatic representation of the relationship between the second dimension protocols utilised in this study.

T = total polyacrylamide concentration (as weight/vol percentage)

C = percentage of crosslinker

(22 x 20 cm) and smaller inner plate (20 x 20 cm). Plates were separated by 8 mm wide Bio-Rad (Hercules, CA, USA) 2-DE gel plate spacers (IPG adaptor kit) with a 1 mm thickness, and clamped together with Bio-Rad (Hercules, CA, USA) clamps. Final gel size was ≈18.5 x 18.4 cm. Where a Bio-Rad Multi-Caster was utilised, clamps were applied after the gels had been removed from the casting apparatus and prior to 2-DE. After the polymerisation solution had been poured between the glass plates, approximately 1 ml of water-saturated isobutanol was overlaid on the top of the polymerisation solution. This ensured a straight polymerisation line at the top of the gel. Gels were left to undergo initial polymerisation for 45 minutes at room temperature. After this, a 1.5 cm gap remained between the top of the glass plates and the upper surface of the gel. The isobutanol overlay was poured off and the gel surface was given three quick washes with high purity H_2O. A 1x gel buffer solution was poured on top as an overlay and the gels stored at 4^{o}C overnight. Storage overnight allowed for more complete polymerisation of the gels.

2.8.3 Tris-glycine buffer system

The methods for the separation of proteins by a Tris-glycine buffer system in this study are those used by Lock et al (2001).

Buffers. The Cathode Buffer (Tris-glycine) was prepared fresh for each electrophoresis run. The Anode Buffer was a Tris-HCl solution that was re-used for a maximum 20 electrophoretic runs.

Gel composition. This study used two formulations of homogeneous (single concentration) polyacrylamide gels with piperazine diacrylamide as the crosslinker. These formulations were 1) 13%*T*, 2.5%*C* (where T = total polyacrylamide concentration and C = % of crosslinker), with subsequent electrophoresis at 25 mA per gel for 5 1/2 hours, and 2) 10%*T*, 2.5%*C* gels at 20 mA per gel for 5 1/2 hours at 10^{o}C. The gels of both formulations were electrophoresed until the bromophenol blue tracker dye had migrated to the bottom of the gel. The 13% polyacrylamide gels were used to resolve low M_r proteins (7-20 kDa) whereas the 10% polyacrylamide gels were used for the optimal resolution of proteins with M_rs ≥ 20 kDa.

2.8.4 Tris-tricine buffer system

Buffers. The Cathode Buffer (Tris-tricine) was prepared fresh each day. The Anode Buffer was a Tris-glycine buffer that was re-used for a maximum of six electrophoretic runs.

Gel composition. A number of different gel compositions were prepared, based on that of Fountoulakis et al (1998). The Fountoulakis method uses the combination of a separating gel and a stacking gel for the separation of proteins. The stacking gel has a low concentration (*T* value) % polyacrylamide. The pH of the stacking gel is also lower than the separating gel (6.9% vs 8.9%). This causes the proteins to run quickly through the stacking gel and pile up in front of the stack/separating gel interface. This sharpens the protein bands. Fountoulakis recommended two different separating gels for the improved resolution of low M_r proteins. In both cases, the crosslinker for these gels was N N'methylenebisacrylamide.

Fountoulakis et al (1998) Procedure with Stacking Gel. The first formulation of the separating gel consisted of a 10.4%*T*, 6%*C* containing 3.1 M urea. The second consisted of a 10.4%*T*, 6%*C* gel containing 6.2 M urea. The final separating gel size was 16.5 x 18.4 cm. The stacking gels for both separating gel formulations had a composition of 5.4%*T*, 3%*C*, without urea, and was 2 cm x 18.5 cm in size. The gels were electrophoresed at 40 mA per gel for 7 1/2 hours at 10^{o}C.

Modified Fountoulakis et al (1998) Procedure without Stacking Gel This study also used a modified Fountoulakis procedure in which the stacking gel was omitted. There were two gel formulations utilised.

The first formulation had a lower separating gel *T* value, 9.5%*T* and 6%*C*. Three different urea concentrations were explored, 0 M, 3.1 M and 6.2 M. Final gel size was 18.5 x 18.4 cm. The 9.5% polyacrylamide gels, without stacking gels were electrophoresed at 40 mA per gel for 8 1/2 hours at 10^{o}C

A second formulation of the modified Fountoulakis procedure without a stacking gel consisted of a separating gel composition of 8.5%*T*, 6%*C* and contained 3.1 M urea.

The gels were electrophoresed at 40 mA per gel for 7 hours and 50 minutes at 10^{o}C. This procedure was the one finally used in this study for the resolution of proteins with M_rs between 7-20 kDa.

2.8.5 Second Dimension Electrophoresis

Following IEF, and prior to electrophoresis in the second dimension, each IPG strip was soaked in 20 mls of SDS equilibration buffer in a plastic Petri dish and shaken on a reciprocating platform for 20 minutes at room temperature. A solution of 0.5% agarose, briefly boiled in Cathode Buffer, was poured on top of the polyacrylamide gel and the re-equilibrated IPG strip was slid through the agarose solution into position on top of the gel. The agarose was left to set for five minutes. All gels were electrophoresed in the second dimension using a Protean II (Bio-Rad, Hercules, CA, USA) apparatus and a 1000 kV Bio-Rad power pack (Hercules, CA, USA). The conditions which were dependent on the gel formulation (see sections 2.8.3 and 2.8.4). Cathode Buffer was freshly made for each electrophoresis run.

2.8.6 After 2D-PAGE

After the electrophoresis run was completed, the gels were removed and placed into a staining solution in Pyrex glass dishes with lids. Two types of stain were utilised in this study, Coomassie Brilliant Blue Colloidal G250 and silver nitrate. (See Section 2.10)

2.9 M_r and pI

2-D SDS-PAGE standards (Bio-Rad, Hercules, CA, USA) were used for estimating molecular masses and isoelectric points of sample spots in the gel. Three calibration 2-DE gels were run for each isolate, using the following samples:

A. 25 µl 2-D SDS-PAGE standards alone (this sample was only electrophoresed once);
B. 15 µl 2-D SDS-PAGE standards plus 10 µl bacterial protein sample ; and
C. 25 µl bacterial protein sample alone.

The first dimension was electrophoresed using pI 3-10 IPG strips (see section 2.7). The second dimension used the Tris-Tricine buffer system (see section 2.8.4). Spots were visualised by staining with silver nitrate (see section 2.10). In order to determine the pI and M_r of all the protein spots in a 2-DE gel the known values for a few spots need to be entered into PDQUEST (Bio-rad, Hercules, CA, USA). With this data PDQUEST can calculate the pI and M_r for all the protein spots. In order to do this a 2-DE gel was electrophoresed which consisted of standards of known pI and M_r. A second 2-DE gel was performed which contained both standards and isolate protein sample. A third gel with only isolate protein sample was run to assist in the determination of which protein spots would be utilised as landmark proteins. The known values of the standards was entered into PDQUEST and all three 2-DE gel types were compared. Gel type A and B were matched by utilising the protein standards as points of reference (landmarks). By comparing Gel type A and B landmark proteins spots were identified that were seen in the same corresponding position for all isolates. PDQUEST was then able to calculate the M_r and pIs for all the spots seen in Gel type C.

Landmark spots were used to set reference positions to compensate for slight positional differences of corresponding spots on different images due to gel distortion. These landmark spots are used by the program to align and position matchset gel images for matching. Spots may be selected as landmark spots if they are well resolved and present in all members of the matchset in corresponding locations. They are particularly valuable if selected at the edges of the image, if isolated from other spots or if located in or near regions of high spot concentrations. A minimum of 20 landmarks were distributed throughout each Standard Image.

Matching is the process by which the analytical software program identifies spots in one or more gel images that are found in identical locations to some of those on the Standard Image. This identification may be performed by auto-matching or manual matching of the protein spots.

Manual matching is used if a legitimate spot is present in corresponding positions on one or more images but is not matched to the Standard Image by automatic

matching. The spot is selected in the Standard Image and its corresponding position is highlighted in all the images of the matchset. If this spot is also located in other images, it may then be selected there and matched.

Auto-matching is the process by which the PDQUEST program automatically matches spots in a number of gel images to those on the Standard Image. The program's algorithm matches spots between images by warping them to a common standard by reference to landmarks.

Image Warping is a function that deforms images; all positions in one image plane are mapped to positions in another plane. It is used to bring two or more images into alignment (Gustafsson et al, 2002).

A **Matchset** is a set of gel images that has been grouped together for the purpose of qualitative and quantitative comparison. Spots on one image can be compared to spots on every other image in the matchset.

There are two types of matchsets: 1) level one matchsets (see Figure 2.2A) and 2) level two matchsets (see Figure 2.2B). A level one matchset is created directly from gel image files. It may contain some or all of the gel images of a given experiment. A level two matchset is created from the Standard Images of a number of level one matchsets.

A **Gel Image File** is a fully processed image file of a gel. This file is a duplicate of the original scan and is utilised in the matchset so it can be edited without losing any of the information contained in the original scan image (PDQUEST Manual).

A **Standard Image** is a composite image that contains all the spots present in all the gel images in a particular matchset. The starting point for generating a Standard Image is to enter all the spots detected into the Reference Image.

A **Reference Image** for a matchset is the gel image chosen as the best representative of all the images in that matchset, i.e. it contains the greatest number of well-resolved spots.

2.10 2-DE Gel Staining

Colloidal Coomassie G250 stain (Neuhoff et al, 1988; Lock et al, 2001) This was prepared at least 24 hours ahead of use. CBB G250 was dissolved in 50 mls of methanol and stirred at approximately 50^{o}C for 20 minutes. The ammonium sulphate, remaining methanol, phosphoric acid and high purity H_2O were placed in a beaker and stirred at approximately 80^{o}C until the ammonium sulphate was completely dissolved. The CBB G250 solution was then added to the ammonium sulphate solution and stirred for the next 24 hours. Immediately prior to use it was left stirring at approximately 80^{O}C for four hours. Following gel electrophoresis, each gel slab was stained in 300 mls of the CBB G250 suspension. Gels were shaken at room temperature for 24 hours then the CBB G250 suspension was poured off and the gels were placed in a 1% acetic acid solution at room temperature for 8 hours during which the 1% acetic acid solution was replaced twice in the first two hours. Gels could then be stored in a 1% acetic acid solution in either glass trays or press-seal plastic bags.

Silver nitrate staining of gels. It was especially important that high purity water (≥ 18 MΩ/cm) was utilised at all steps of silver nitrate staining and that the glass dishes were thoroughly cleaned. Gels were stained in Pyrex glass dishes, with lids, at room temperature. Each gel was soaked in Fixer 1 for at least one hour, and then placed in Fixer 2 for 1 hour. The gel was then given four 15-minute washes with H_2O and then placed in silver nitrate stain for 45-60 minutes. The silver nitrate was poured off and the gel was given two quick rinses in H_2O, then a three-minute wash in H_2O. The gel was then transferred to a second dish containing the Developer and shaken until the majority of spots appeared to have developed. The Developer was then poured off and Stop Solution added. The gel was left in the Stop Solution for 5 minutes then washed twice in H_2O. For storage, prior to scanning, the gel was

placed in a 1% acetic acid solution. In this solution the gels could be either stored in glass dishes or in press-seal plastic bags.

2.11 Gel Drying

Gels could be dried for long-term storage. Gel drying was performed after the gels had been scanned (see section 2.12.1). The gel was soaked in the Drying Solution for five hours. Two sheets of cellophane were cut to a size 2 cm larger than the gel and then soaked with the gel in the last hour. Gels were dried using a Bio-Rad (Hercules, CA, USA) Drying apparatus set to linear temperature ramping, at 55^{o}C for 9 1/2 hours.

2.12 Image Analysis

The comparative analysis software utilised in this study was PDQUEST™ version 6.2. PDQUEST is a specialised software package for the analysis and databasing of 2-DE gel images. The steps involved in image analysis include scanning, spot detection, spot editing, matching, formation of replicate groups, normalisation and comparison.

Replicate Groups. A replicate group is a set of gels that is prepared from the same protein sample using the same protocol where a **protocol** refers to the specific electrophoresis conditions that were used to separate proteins in a particular M_r and pI range.

There are two reasons for the formation of replicate groups. The first is to determine the validity of spots. In order to be accepted as valid, this study required that a particular spot must be observable in at least three out of four gel images for each replicate group. In addition, slight variations in quantitation values needed to be adjusted, due to variables such as loading and staining. The volume of a given spot is best estimated by averaging results from several gel images.

2.12.1 Scanning

Gels were imaged by a Bio-Rad (Hercules, CA, USA) Scanning Densitometer GS-700 using Multianalyst software version 1.02. Gels were placed on a neutral-density diffuser plate prior to scanning by the densitometer. The following scanning parameters were used for all gels within this study:

Light source	Red (electronic filter) for Coomassie G250 stained gels. Greyscale for silver nitrate stained gels.
Bits per pixel	8
Resolution	150 dpi
Mode	Transmission
Optical filter	Neutral- density diffuser plate

Using these parameters file sizes were usually ≈ 1.4 Mb.

2.12.2 Automated spot detection

Images were subjected to automatic spot detection with the *Automatic Spot Wizard* feature using the following parameters:

The **faintest spot** was selected (this spot was different for each isolate). This set the minimum intensity threshold for spot detection. Next, the smallest spot was selected. This set minimum spot size. It usually consisted of a high molecular mass protein seen in a corresponding position for all isolates. The smallest spot was not necessarily the faintest spot. The largest spot was then selected on the gel image. The radius of the largest spot is used in the program methodology to determine background subtraction and level of streak removal required. Background subtraction involved selecting a background region of the image, which contained no spots or streaks. This step set the threshold that was required to differentiate between the spot and background. The gel image files usually displayed **streaks** in areas of

high spot concentration and at the edge of the gel image. Streaks were removed from the image by selecting the *remove vertical and horizontal streak* option. Filtering methods were left at the default settings for all gel images.

2.12.3 Spot editing

Gaussian Spot. PDQUEST uses Gaussian modelling to create "idealised" spot boundaries. A Gaussian spot is a precise three-dimensional representation of a scanned spot. Gaussian curves are fitted to the scanned spot in the X, Y and Z dimensions, where X and Y represent co-ordinates on the 2-D gel and the Z axis represents spot volume. Additional modelling is then performed to create the final Gaussian spot (PDQUEST Manual).

After spot detection was completed a processed gel image file and Gaussian gel spot image file was generated by the PDQUEST software. These image files were utilised in matchsets at a later point.

Parameters are set in order to obtain the maximum detection of spots. However, this usually results in a number of artifactual spots. Therefore the next step involves manual spot editing.

The majority of spot editing involves erasure of spots. These are often clustered around the edges of the gel and are easily removed as groups. Other artifactual spots such as air bubbles or streaks that are not deleted by the automated detection steps are individually deleted.

Additionally some spots of low intensity or partially merged spots may be missed by the automatic spot detection algorithm. These spots are manually added to gel spot files prior to file matching.

Gel Spot File contains a synthetic image with Gaussian representation of the spots in the Gel Image File (PDQUEST Manual). At least 6 successive Gaussian spot approximations are required to generate the Gel Spot File.

2.12.4 Matchsets

Four replicate gels were electrophoresed for each isolate. A replicate group (which contained the images for these four gels) was formed for each isolate. The protocols used in the present study were as follows (see Figure 2.2):

Protocols optimising for low M_r proteins between 7-62 kDa

- Proteins separated in the first dimension within the pH 4-7 range and in the second dimension using the Tris-tricine electrode buffer method.
- Proteins separated in the first dimension within the pH 6-11 range and in the second dimension using the Tris-tricine electrode buffer method.

Protocols optimising for high M_r proteins ≥62 kDa

- Proteins separated in the first dimension within the pH 4-7 range and in the second dimension using the Tris-glycine electrode buffer method.
- Proteins separated in the first dimension within the pH 6-11 range and in the second dimension using the Tris-glycine electrode buffer method.

First Matching Strategies (level one matchsets). The first matching strategy consisted of preparing a level one matchset containing all the gel images in a particular protocol. The level one matchset consisted of 44 gels for each protocol (11 isolates x 4 gels per strain) excepting Tris-glycine 6-11 gels in which there were 40 gels (10 isolates x 4 gels per strain). Due to the limited sample available for the eleventh isolate (GI2 - 26695), this sample was not separated by the Tris-glycine electrophoresis conditions for the pH range 6-11. Isolate 26695 (GI2) was included in the study near the end and only a limited amount of sample was available. In order to perform comparative proteome analyses a Standard Image needed to be created.

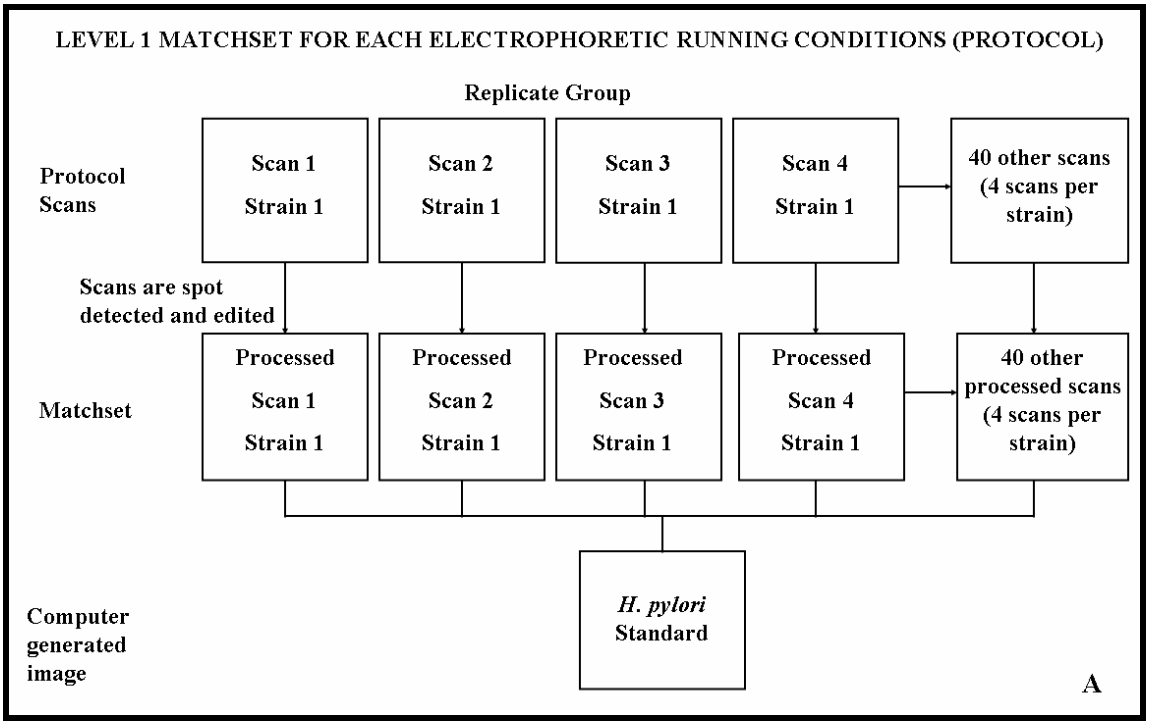


Figure 2.2 Construction of Matchsets

A. Organisation of experiments into a low level matchset.

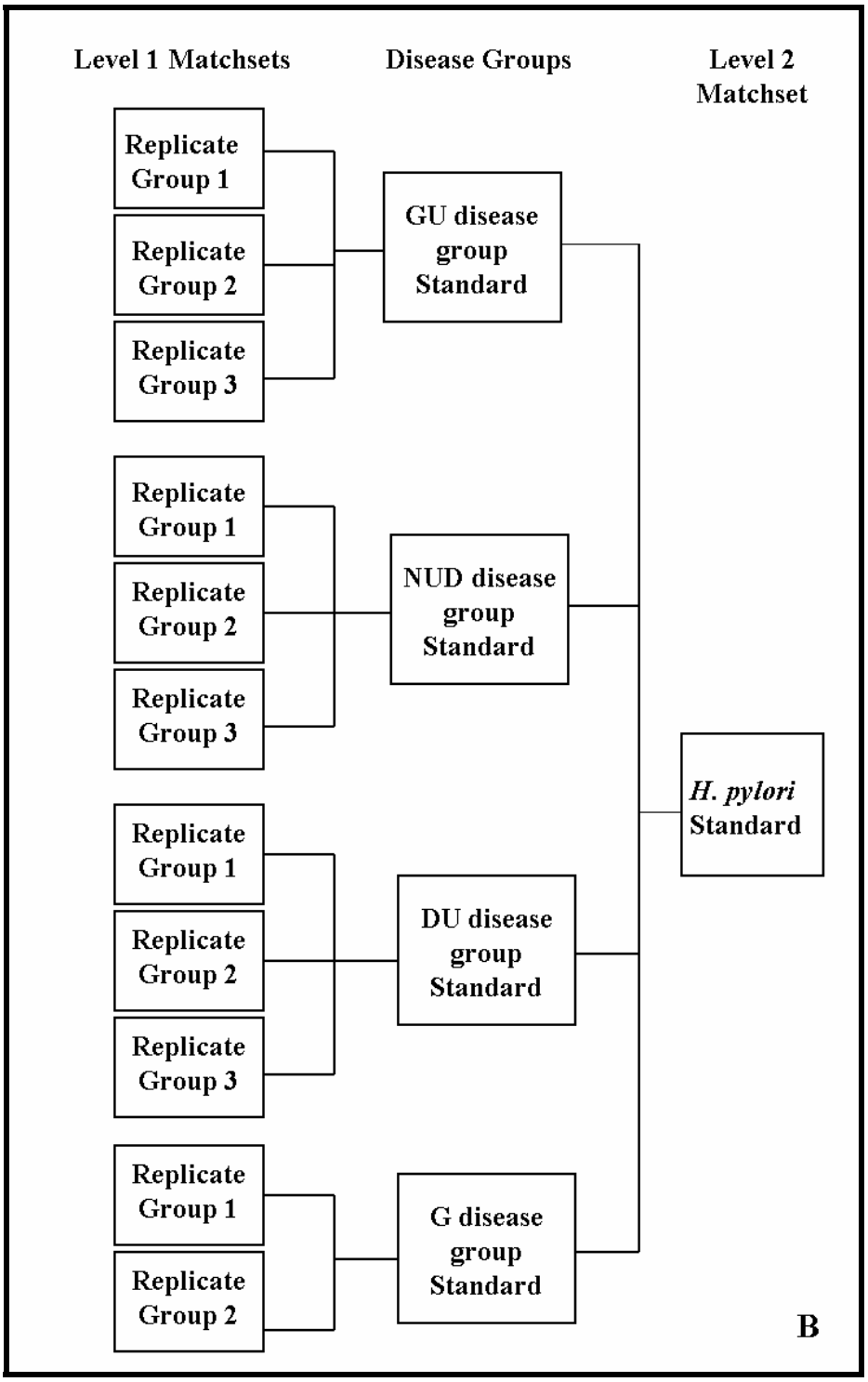


Figure 2.2 Continued

B. Creation of a higher level matchset. The higher level matchset is made from four lower level matchsets, GU (gastric ulceration), DU (duodenal ulceration), and NUD (non-ulcer dyspepsia) consisting of three members (isolates) and GI (gastritis) containing only two members. Each member includes four gel images.

Firstly, a gel was selected both visually and by PDQUEST as the Reference Image for the Standard. The criteria for this selection were which gel contained

- the largest number of spots as detected by PDQUEST;
- the least streaks both horizontally and vertically;
- no warping of gel image; and
- the best separation of spots both horizontally and vertically.

GU1 was the Reference Image for all matchsets in this study. See the Appendix for examples of all replicate gel images in this study.

The second step involved selecting landmark proteins. A minimum of 20 landmarks was selected.

The third step was the placement of replicate gel images files into a replicate group (gel images GU1 [1-4] formed replicate group GU1).

The fourth step involved the auto-matching of the Standard Image spots to the corresponding spots of all the gel images included in the matchset. Those proteins that could not be auto-matched to the Standard were manually matched if necessary.

After all the protein spots in the Standard had been matched, those spots which remained unmatched in the other three replicate images (i.e. spots present in all three images) of replicate group GU1 were added to the Standard. The matchset was then auto-matched and manually matched again.

The fifth step was to match spots from the gel image files of GU2 to those in the Standard. Those spots that were located in at least three out of four gel image files of this strain's replicate group were accepted as valid and added to the Standard. When

the spot addition was completed, the spots in the matchset were auto-matched and manually matched to the Standard.

The addition, auto-matching and manual matching of protein spots to the Standard was repeated for gel images of each replicate group in turn. The order of spot addition to the Standard was GU1 (reference image), GU2, GU3, GI1, NUD1, NUD2, NUD3, DU1, DU2 and DU3, G3.

The final Standard Image was representative of all the spots in the matchset.

Second Matching Strategy (level 2 matchsets). A level two matchset was generated from the Standard Images of level one matchsets. The first step in this strategy involved the construction of four level one matchsets, one for each of the four disease outcomes (i.e. GU, DU, NUD and GI). Isolates that were selected as the reference images for each disease group matchset were:

Disease Outcome	**Reference image**
Gastric ulceration	GU1
Non-ulcer dyspepsia	NUD1
Duodenal ulceration	DU2
Gastritis	GI1

The next step was to place landmarks for each level one matchset. After this was completed, spots from each image in the matchset were auto-matched and then manually matched to the Reference Image as described for the first matching strategy. Spots that could not be matched to the initial reference spots in the Standard were auto-added to the Standard then auto-matched to the other gels in the matchset. This was repeated until every spot in the matchset was matched to a spot in the Standard. Each completed matchset Standard Image consisted of all the spots present in all gel images for each disease outcome. Replicate groups were formed for every level one matchset.

After the four level one matchsets were completed, the next step was to use them to generate a level two matchset. To generate a high-level matchset a dialog box was opened, but instead of selecting the Matchset type as level one, level two matchset was selected. The program then displayed those level one matchsets that were available for inclusion in the level two matchset. A reference image for the level two matchset Standard was selected. This was generated from the Gastric Ulceration matchset. The procedure for matching spots and auto-adding spots in the level two matchset was the same as for level one matchsets.

2.12.5 Entering M_r and pI Data

M_r and pI data were entered by using PDQUEST's *enter MrpI data tool* selecting a specific spot then entering the data in a dialog box. At least 12 reference spots were used for each protocol. The M_r and pI of these sample spots were estimated by reference to 2-DE standards (known M_r and pIs: see Section 2.8). The sample spots utilised to enter M_r and pI data were present in corresponding positions for all isolates. The program then used these sample spots as internal references to automatically assign M_r and pI to all other spots.

2.12.6 Normalisation

Before isolates were compared for differences in their proteomes (comparative spot analysis), the spot volumes of the different gel image files included in the matchset were adjusted by normalisation. Normalisation corrects for systematic differences due to variable protein loads and staining effectiveness.

The normalisation procedure utilised was *Total of All Valid Spots*. In this method, the raw volume of each spot in a member gel is divided by the total volume of all the spots in that gel that have been included in the Standard.

2.12.7 Analyses

After replicate groups had been formed, matching completed, and the protein spots normalised, analysis sets were set up to compare differences. In the first matchset

strategy, analysis sets were constructed for each of the four experimental protocols to compare differences between the disease outcome groups of each protocol matchset. Each disease outcome group, except gastritis, consisted of three replicate groups – one for each isolate. In level 1 matchsets of the second matchset strategy, analysis sets were constructed to compare replicate groups within each disease group's matchset for each experimental protocol.

Quantitative and qualitative spot comparisons were undertaken in this study. In order to compare quantitative or qualitative differences between two disease groups, within a level 1 matchset, differences had to be determined and Boolean sets constructed for both qualitative and quantitative differences. The construction of analysis sets was the same for both matching strategies.

Quantitative spot comparisons are composed of spots whose quantitation has increased or decreased "x" fold (where x is a user-determined value) or whose quantitation has changed above or within the fold change factor that was selected.

Qualitative spot comparisons are composed of spots that are present in one gel but not in another. A protein spot could appear to be absent because it is either a low copy number protein or the gene is turned off.

An example of a Comparative Quantitative Analysis between two disease groups (Figure 2.3). A number of analysis sets were constructed to determine which protein spots had a three-fold increase in concentration in the GU disease group when compared to the NUD disease group. The first step was to construct quantitative difference sets between each GU replicate group and the NUD replicate groups (i.e. GU1vsNUD1, GU1vsNUD2, GU1vsNUD3, GU2vsNUD1, etc). Boolean sets were then constructed to determine which protein spots displayed a greater than three-fold increase in concentration for all isolates within a disease group. This required the construction of intersecting Boolean sets that were then used to construct a Master Boolean set i.e. the intersection of all Analysis sets. The construction of analysis sets was also performed for both qualitative and quantitatively conserved proteins.

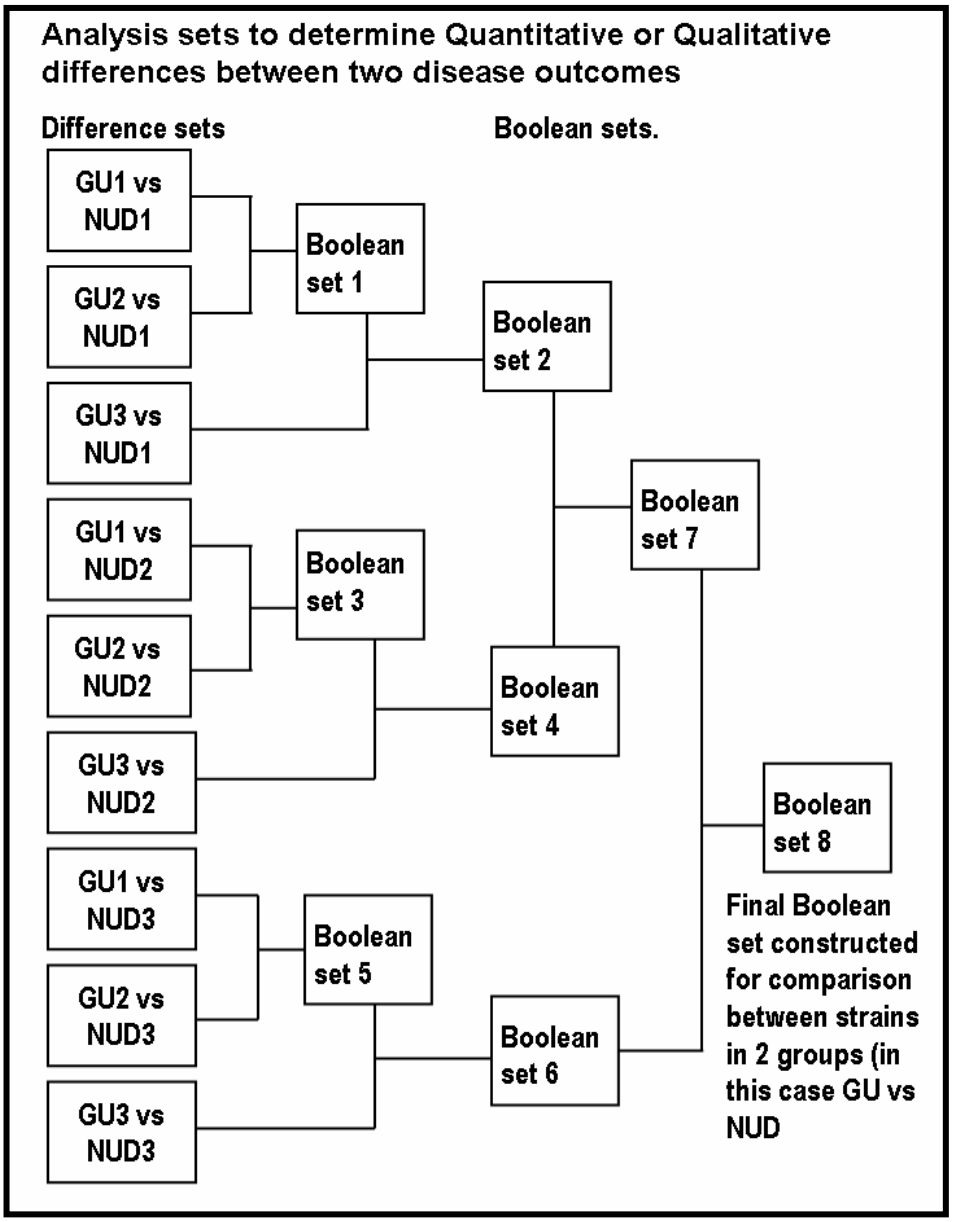


Figure 2.3 Example of Analysis sets

Representation of the construction of analysis sets to compare differences between two disease outcomes in a level 1 matchset. In this representation GU (gastric ulceration) isolates are compared with NUD (non-ulcer dyspepsia) isolates. This comparison was undertaken using both quantitative and qualitative analysis sets. This procedure was used to compare GU vs GI, vs DU vs NUD.

Difference Sets refers to quantitative or qualitative analysis sets. Each Boolean set represents the intersection of two other sets.

2.12.8 Exporting Matchset Data

Raw data, generated by PDQUEST, for the spots of each isolate's image consisted of protein spot volume (normalised), M_r and pI and the spot's number. The raw data was exported into an Excel (MAC version 10) file. This raw data was then processed by SPSS Manual Version 11.

2.12.9 Excel merging of overlapping pI and M_r ranges

Each electrophoresis running condition for separating low M_r and high M_r proteins resolved them in two pH ranges pH 4-7 and 6-11, each range had a corresponding region of pH 6-7. In order to obtain a complete pH 4-11 master map of *H. pylori* this pH 6-7 region in both pH ranges needed to be combined. The spatial resolution on the x-axis was greater in the pH 4-7 region in comparison to the 6-11, therefore accurate matching of the pH 6-7 region could not be obtained. Additionally because of the increased spatial resolution there was an increase in the spot numbers. In order to obtain the most accurate spot data for a master pH 4-11 gel only the pH 7 protein spots, from each pH region (4-7 and 6-11) was merged. The 6-6.99 protein spots information was manually deleted from the pH 6-11 gels. The pH 7 protein spots displayed in the 2-DE map were determined by PDQUEST, merged, and then the spot volumes were averaged. This spot volume information was then included in the Excel files.

Two different sets of electrophoresis conditions were used to separate proteins in the second dimension. The current warping routines for image analysis programs are unable to match corresponding protein spots that have been separated on two different logarithmic scales. Under these circumstances, volume information for best spatial resolution and separation of the particular spots under consideration. Data for spots with M_r greater than 62 kDa were taken from the Tris-glycine gels because of the improved spatial resolution of these spots on such gels in comparison to those on Tris-tricine gels. Data for spots with M_r below 62 kDa were taken from the Tris-tricine gels; these protein spots were best resolved in the Tris-tricine gels, especially proteins with M_r between 7-20 kDa.

2.13 Cluster analysis

Dendrograms indicating isolate similarities were derived by compiling all the spot data for the eleven isolates for each protocol into an Excel file. Cluster analysis was performed using SPSS MAC Version 11 (Chicago, IL, US). The first step was to select the type of classification to be utilised. In the present case, similarities between the isolates were calculated using hierarchical cluster analysis. The parameters used were as follows:

Parameter	**Selected Type**
Variable	Spot volume
Cases	Isolates
Clustered according to	Cases
Statistics	
Agglomeration schedule	Yes
Single cluster solution	11 cases
Dendrogram	Vertical orientation
Cluster Method	Between-groups linkage
Distance Measure	Interval 1. Squared Euclidean distance 2. Euclidean distance Binary • Simple matching

Agglomeration schedule. The results of the cluster analysis are summarised in a listing of the cluster numbers (assigned ID number = cluster number) being combined at each stage in the cluster analysis (SPSS Manual Version 11).

Single solution. Displays cluster membership for a single cluster solution with a specified number of clusters.

Between-groups clustering. The distance between two clusters is defined as the average of the distances between all (SPSS Manual Version 11).

Distance Measure allows a specific distance or similarity measure to be used in clustering. SPSS does not plot actual distances but re-scales them to numbers between 0 and 25. Lines, which are the measure of relatedness on a diagram, join the cases to indicate a clustering. Distance can be calculated in two different forms Interval and Binary. Interval measures compare the quantitative values for variables (i.e. the volumes for spots) between cases (i.e. isolates) and binary measures compare the qualitative values for the variables (i.e. presence/absence only) (SPSS Manual Version 11).

Interval measure involves the similarity measures for interval data (Euclidean distance and Squared Euclidean distance). The data utilised were the protein volume for each variable (spot) present in each isolate. Spot quantities ranged from 0-10,000 (normalised), where 0 represents an absent protein. The distance between two variables was measured using both Squared Euclidean distance and Euclidean distance.

Euclidean distance (interval measure) is a dissimilarity measure for continuous data. The distance between two items is the square root of the sum of the squared differences in values for each variable (SPSS Manual Version 11).

Squared Euclidean distance (interval measure). A measure of distance between pairs of cases. The distance between two cases is the sum of the squared differences between the variables of the cases. This dissimilarity measure is used for continuous data (SPSS Manual Version 11).

	SPOT NUMBER 1	SPOT NUMBER 2
GU1	100	90
NUD1	280	150

In the above example [values are normalised spot volumes (units determined by a computer algorithm)] the squared Euclidean distance is: $(280\text{-}100)^2 + (150\text{-}90)^2$ which equals $180^2 + 60^2$, which equals 36,000. A disadvantage with this measure is that when variables are measured on different scales variables that are measured in

larger numbers will contribute more to the computed distance than variables measured in smaller numbers.

In **binary measure** (Squared Euclidean, Pattern difference and Simple matching) the spot volume values were converted to either 1 or 0, where 1 represented a spot present in a isolate and 0 represented an absent spot. The distance measure used to calculate the dissimilarity between isolates was Pattern difference. Other measures in binary measure analysis (Simple matching and squared Euclidean distance) were applied in order to determine if the resulting dendrogram changed.

Simple matching (binary measure) is the ratio of matches to the total number of values. Equal weight is given to matches and non-matches (SPSS Manual Version 11).

2.14 Peptide mass fingerprinting

2.14.1 Spot Excision

The majority of proteins of interest selected for mass spectrometry analysis were high-intensity spots (with the exception of 20 low-abundance proteins). The more intensely stained a spot the higher the protein amount, which enhances the possibility of protein identification by mass spectrometry.

The person performing the excision was gowned, gloved and masked, and hair was covered with a paper cap in order to help minimise keratin contamination of the gel. Protein spots were excised from the gel by hand with scalpels (Paragon, Sheffield, England). The scalpel blade was flamed between each excision in order to burn off any fragments of gel that might carry over. Each spot was transferred into a 1.5 ml centrifuge tube (Sarstedt, Numbrecht, Germany). The majority of spots excised were from the corresponding positions of at least two isolates (see Table 2.2). Exceptions were spots 1-96 which were positionally conserved reference spots from strain NCTC11637.

Table 2.2 Proteins excised.

Except where indicated all proteins were searched at 50 ppm and 1 missed cleavage.

Spot ID	Spot excision Isolates	Spot ID	Spot excision Isolates
1	GI1	42	GI1
2	GI1	43	GI1
3	GI1	44	GI1
4	GI1	45	GI1, DU2
5	GI1	46	GI1, DU2
6	GI1	47	GI1, DU2
7	GI1	48	GI1, DU2
8	GI1	49	GI1, DU2
9	GI1	50	GI1
10	GI1	57	GI1, GU1, GU2
11	GI1	52	GI1, DU2
12	GI1	53	GI1
13	GI1	54	GI1
14	GI1	55	GI1
15	GI1	56	GI1, DU2
16	GI1	57	GI1
17	GI1	58	GI1
18	GI1	59	GI1
19	GI1	60	GI1, DU2
20	GI1	61	GI1, DU2
21	GI1	62	GI1, DU2
22	GI1	63	GI1
23	GI1	64	GI1, DU2
24	GI1	65	GI1, DU2
25	GI1	66	GI1
26	GI1	67	GI1, DU2
27	GI1	68	GI1, DU2
28	GI1	69	GI1
29	GI1	70	GI1
30	GI1	71	GI1, DU2
31	GI1	72	GI1, DU2
32	GI1	73	GI1, DU2
33	GI1	74	GI1, DU2†
34	GI1	75	GI1
35	GI1	76	GI1, DU2
36	GI1	77	GI1
37	GI1	78	GI1, DU2‡
38	GI1	79	GI1
39	GI1	80	GI1, DU2
40	GI1	81	GI1
41	GI1	82	GI1, DU2

§ Not Identified

† 60 ppm †† 70 ppm ‡ 2 missed cleavages

Table 2.2 Continued

Spot ID	Spot excision Isolates	Spot ID	Spot excision Isolates
83	GI1, DU2	125	GI1, DU1, DU3
84	GI1, DU2	126	GI1
85	GI1, DU2	127	DU1, NUD2†
86	GI1, DU2	128	DU2, NUD2†
87	GI1, DU2	129	DU1, DU2
88	GI1	130	DU2, DU3
89	GI1, DU2	131	DU1, DU2, NUD1, NUD3†
90	GI1	132	DU1, DU3‡
91	GI1, DU2	133	NUD3, NUD1, GI1†‡
92	GI1, DU2	134	GU1, GU2, NUD2†††‡
93	GI1	135	DU1, DU2
94	GI1	136	DU1, DU3
95	GI1	137	DU1, GI1‡
96	GI1	138	NUD3, GU1, GU3
97	GI1	139	DU1, DU2
98	GI1	140	GU3, GU2, DU2
99	GI1, DU2	141	GU3, GU2, GU1
100	GI1	142	GU1, GU2
101	GI1	143	NUD2, GI1‡
102	GI1, DU2	144	NUD1, NUD2
104	GI1	145	GU1 NUD2
105	GU1, GU2, GU3	146	GU1, GU2
106	GU1, GU2, GU3	147	NUD2, NUD3
107	GU1, GU2	148	GI1, NUD2, NUD3
108	GU1, GU2, GU3	149	GI1, GU2, NUD3
109	GU1, GU3	150	GI1, GU3, DU3
110	NUD1, NUD2, DU1, DU3	151	GI1, DU2, NUD3
111	NUD1, NUD3, DU2	152	GI1, DU2, NUD1
112	NUD1, NUD3, DU2	153	GU1, NUD1
113	DU1, NUD3	154	GI1, GU1, NUD1
114	NUD3, DU1	155	GI1, GU1, GU2
115	NUD1, NUD3	156	GI1, GU2, GU3
116	NUD2, NUD3	157	GI1, GU2, DU3
117	NUD1, NUD2	158	GI1, GU1, NUD3
118	GU1, DU2	159	GI1, GU2, DU1
119	GU2, DU1	160	GI1, GU1, NUD2
120	GU1, NUD1	161	GI1, GU2, NUD1
121	GU2, NUD3†‡	162	GU1, NUD3
122	GU1, NUD2	163	GU1, NUD2
123	DU1, DU2†	164	GU1, NUD1
124	DU2, DU3	165	GU2, NUD3

§ Not Identified

† 60 ppm †† 70 ppm ‡ 2 missed cleavages

Positionally conserved refers to protein spots that are found in corresponding positions for all isolates included in the study. Positionally conserved spots consist of two subsets 1) spots which vary by less than a factor of three or more (quantitatively conserved) and 2) spots which vary in volume by a factor of three or more.

Mass spectrometry data was generated by either the Australian Proteome Analysis Facility (APAF) (Lock et al, 2001) or Proteomics International (Perth, Western Australia).

2.14.2 The Australian Proteomics Facility (Lock et al, 2001) Procedure

In-gel protein digests. Gel pieces that had been stained with Coomassie G250 were washed for 1 hour in 60%, 0.05 M ammonium bicarbonate in 40% acetonitrile. After washing, the gel pieces were dried under vacuum to allow maximum adsorption of the trypsin digest solution. Gel pieces were rehydrated in 15µl of trypsin digest solution (12.5 µg/ml of trypsin solution containing 0.05 M ammonium bicarbonate) and incubated for 45 min at 4^{o}C. Unabsorbed trypsin solution was removed with a micropipette and the gel pieces further incubated overnight at 37°C. Ten µl of 100% acetonitrile containing 1% trifluoroacetic (TFA) acid was added to the digest solution, and incubated at 25 °C for 20 min. The extraction was repeated and pooled extracts were dried under vacuum to concentrate the samples to 1-2 µl final volume.

Mass spectrometry. Samples were analysed in a Perseptive Biosystems Voyager-DE STR MALDI-TOF mass spectrometer. A nitrogen laser tuned to 337 nm was used to irradiate the sample in 0.5 µl of matrix (α-cyano-4-hydroxycinnamic acid [HCCA]), 10 mg/ml in 70% acetonitrile and 0.1% TFA) spotted onto the MALDI target plate. A manual, internal calibration was performed using the 8421.51- and 2211.11-Da peaks generated by the autolysis of trypsin. Peaks were assigned manually using the GRAMS software package (Perseptive Biosystems, Framingham MA, USA). Spectra were acquired in reflectron mode using a mass range of 800-3898 Da.

The peptide masses were then used to search the SWISS-PROT and TrEMBL databases using the program PeptIdent. Peptide mass fingerprint database results were searched by APAF.

2.14.3 The Proteomics International Procedure

In-gel protein digests. Gel pieces that had been stained with Coomassie G250 were washed for 45 minutes at 37°C with 50µl of 25 mM ammonium bicarbonate in 50% acetonitrile. After washing, the gel pieces were dried under vacuum to allow maximum adsorption of the trypsin digest solution. Ten µl of trypsin digest solution (12.5 µg/ml of trypsin solution containing 25 mM ammonium bicarbonate) was added to each gel piece and the material incubated overnight at 37 °C. Ten µl of 100% acetonitrile containing 1% TFA was added to the digest solution and this was incubated at 25 °C for 20 min. The extraction was repeated and pooled extracts were dried under vacuum to concentrate the samples to 1-2µl final volume.

Mass spectrometry. One µl of each sample was diluted 1/10 with matrix HCCA (10 mg/ml in 50% acetonitrile and 1% TFA) and 1µl spotted onto the MALDI target plate. Samples were analysed with MALDI-TOF mass spectrometry (Voyager-DE PRO, Applied Biosystems Inc, Chicago, IL, USA). The instrument was calibrated with SIGMA peptide standards (bradykinin fragment 1-7, angiotensin II (human), $P_{14}R$ (synthetic peptide) and adrenocorticotropic hormone (ACTH) fragment 18-39 (human) using HCCA as matrix. The following settings* were used.

Mode of operation: reflector
Extraction mode: delayed
Polarity: positive
Accelerating voltage: 20000 V
Grid voltage: 76%
Mirror voltage ratio: 1.12
Grid wire: 0.01%
Extraction delay time: 100 nsec
Mass range: 600-5000 Da
Number of laser shots: 50/spectrum

Laser intensity: 2200

*Note these settings were unknown for the APAF data because it was lost due to their computer failure.

The peptide masses were then used to search the NCBI and SWISSPROT databases using the MS-FIT program.

2.15 Protein identification

There are a number of peptide mass fingerprint on-line programs that may be utilised to search databases. Each on-line program provides access to a range of different databases and parameters affecting the search result. In the present study, several on-line programs were examined to determine which program should be used. The main parameters of interest were: 1) the type and number of accessible databases; 2) the ability to specify a particular species search, in particular *H. pylori*; 3) the ability to modify the M_r and pI ranges to be searched and; 4) the ability to specify particular protein modifications (including post-translational modifications). Species filtering is a procedure that restricts database searches to a user-designated species or collection of species. In the present study only proteins specific for *H. pylori* were initially of interest.

Eventually the peptide mass fingerprints were analysed by using the on-line Protein Prospector program MS-FIT developed by the UCSF Mass Spectrometry Faculty (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm). Both the NCBI and SWISS-PROT databases (http://www.ncbi.nih.gov/Entrez/ and http://us.expasy.org/sprot/sprot-top.html, respectively) were searched for protein identification.

Monoisotopic peptide masses were used to search the database, allowing an observed molecular mass range of $M_r \pm 15\%$, a peptide mass accuracy of 50 ppm, and one partial tryptic cleavage. The pI range selected was 3-11. Pyroglutamic acid modification of N-terminal glutamine, oxidation of methionine and acrylamide modification of cysteine were considered. In order to obtain strong matches, this

study searched for protein identities with six or more matching peptides (default is four). Protein identification was confirmed by analysing spots obtained from the corresponding positions of at least two isolates.

If proteins were not identified using the above parameters, the parameters of partial cleavage and/or peptide mass accuracy were extended (see Table 2.2 page 96 for parameters for each spot identified). When the extension of these parameters did not result in protein identification, species filtering was removed. Peptide mass fingerprints that resulted in a *no hit* answer with peptide mass accuracies of 50-100 ppm were reanalysed to confirm the result.

Possible post-translational modification of proteins was searched using MS-FIT and FindMod at http://us.expasy.org/tools/findmod/.

3.1 First dimension protein separation (isoelectric focusing)

Two previously published studies of *H. pylori* employed broad-range IPG strips (pH 3-10) (Lock et al, 2001; Jungblut et al, 2000) and were subsequently able to resolve 381 proteins (Coomassie stained) and 1800 proteins (silver stained) respectively. In addition, both studies found that, in comparison to other organisms, such as *Mycobacterium tuberculosis*, basic proteins were dominant in the *H. pylori* 2-DE patterns. Jungblut et al (2000) stated that more than 70% of the predicted proteins had a pI ≥7. This figure was somewhat higher than the actual observed 2-DE pIs of basic proteins in the present study that found that only 47% of the proteins had a pI≥7.

Broad range strips do not have the spatial resolution of narrower range strips, which often results in two or more proteins comigrating. In order to improve the spatial resolution and thereby increase the number of observable proteins, proteins in the present study were separated in two overlapping pH ranges 4-7 and 6-11 (Pharmacia, Upsala).

3.2 Separation of proteins in the second dimension (SDS-PAGE)

3.2.1 Gradient gels

In the early stages of the project *H. pylori* proteins were separated in the second dimension using commercially pre-cast polyacrylamide gels [Bio-Rad Ready-cast™ gradient gels (10-16%)]. The electrophoresis method employed Tris-glycine electrode buffer. The proteins were poorly resolved by these gradient gels. Problems consisted of slow migration of proteins, distortion of protein patterns, and insensitivity to silver staining.

Slow migration of proteins

Homogeneous (single concentration) 10% polyacrylamide gels prepared in this laboratory have an appropriate electrophoresis time of about 5 hours at 25 mA (constant current) per gel. Under these conditions, tracker dye, bromophenol blue (BPB), migrates to the bottom of the gel, and protein migrates across the whole area of the gel. Bio-rad instruction with the precast gradient gels recommended 16 mA per gel for 30 min, then 24 mA per gel for 4 hrs 40 min. When these conditions were

used, the BPB tracker reached the bottom of the slab as expected. However, when the gel was stained with Coomassie G250, the proteins had migrated only about half way down the slab and the bottom half was blank. During the gradient gel run, voltages of about 500V were observed for a setting of 25 mA per gel, whereas 10% polyacrylamide gels have voltages of approximately 380-400V at the same setting.

Distortion of protein patterns

In order to increase the distance of protein migration on gradient gels, the electrophoresis time was lengthened. After 8 hours at 25 mA per gel, the protein pattern was still well resolved but still only covered about two thirds of the gel (the BPB tracker had migrated off the gel). When the electrophoresis time was extended for longer than 8 hours the protein patterns became highly distorted and not only was there considerable streaking but also "reverse smiling" (see Figure 3.1), an effect caused because there is a faster migration rate of proteins at the edges of the gel than in the centre.

Insensitivity to silver staining

Under the standard electrophoresis conditions used in this laboratory, silver is about 20 times more sensitive than Coomassie G250 and produces low background staining. When the precast gradient gels were utilised, good results were still obtained with G250, but the silver staining sensitivity was reduced. In fact, silver and G250 had comparable sensitivity (Figure 3.1). In order to improve the number of protein spots visualised the development time was extended. This did not improve the total number of proteins visualised and resulted in extremely high background staining. Silver staining, however, is not a suitable method for comparisons of different proteomes because it produces variable proteomes. This variability is the result of an operator determined stopping point in the development of the stain. The proteome needs to be reproducible in order to obtain accurate quantitative and qualitative comparisons, therefore no further silver stain methods were examined.

3.2.2 Crosslinkers

Two different crosslinkers were used in this study. The polyacrylamide gels electrophoresed with a Tris-glycine buffer system utilised piperazine diacrylamide

(PDA) as a crosslinker whereas gels electrophoresed with a Tris-tricine buffer system utilised N N'methylenebisacrylamide (Bis). The Tris-tricine buffer system gels were run according to a modified Fountoulakis et al (1998) system using crosslinker at a concentration of 6%. However, PDA is a more expensive crosslinker compared to Bis, which made it more economical to use Bis under these running conditions. However, when Bis was utilised at lower crosslinker percentages (10%*T* 2.5%*C*) this resulted in a greater swelling of the gels. When the gels were swollen this made it impossible to fit the whole gel onto the densitometer scanning plate. Therefore PDA was utilised in the resolution of proteins separated on the 10%*T* 2.5%*C* gel.

3.2.3 Running Conditions

Electrophoretic Conditions

Fast separations of proteins on large gels, within a few hours instead of overnight, resulted in spots that were less diffuse compared to those from overnight runs. This was particularly noticeable for Coomassie stained gels. Diffuse spots reduce the enzyme efficiency during in-gel digestion (Westermeier and Naven, 2002). For this reason Coomassie stained gels, in particular those to be utilised for later mass spectrometry analysis, were never run overnight. Instead, they were electrophoresed for 5 1/2 hours for Tris-glycine gels and 7 1/2 hours for Tris-tricine gels. The current applied did not exceed 25 mA for Tris-glycine gels and 45 mA for Tris-tricine gels. If these settings were exceeded this resulted in vertically distorted protein spots. This may also have been due to uneven increases in the gel temperature generated by the high current.

3.2.4 Tris-glycine electrode buffer gels

The Tris-glycine buffer system is the conventional method for the separation of proteins in the second dimension. It provides excellent resolution for proteins with $M_r \geq 15$ kDa.

The method by Jungblut et al (2000) and Enroth et al (2000) recommended a separation gel consisting of a 13% *T*, 2.5% *C* for the resolution of the entire proteome of *H. pylori*, (M_r 5-150 kDa), while a study by Lock (2001) indicated excellent resolution of proteins ≥ 20 kDa on a 10% *T* 2.5% *C*. These studies

separated the proteins using the Tris-glycine electrode buffer system. In the current study, the separation of *H. pylori* proteins was examined according to both of these methods, the first for the resolution of low molecular mass (M_r) proteins and the second for the resolution of proteins ≥ 20 kDa.

Electrophoresis of proteins on 13% polyacrylamide gels using the Tris-glycine method (Jungblut et al, 2000; Enroth et al; 2000) achieved poor spot resolution of proteins with M_r below 20 kDa (see Figure 3.2). These protein spots were poorly reproducible which would present problems in comparative proteome analysis, especially in the image analysis portion of the project. Proteins with $M_r \leq 20$ kDa displayed distinctive smearing and streaking. In addition, staining with Coomassie G250 resulted in poor staining of these proteins. The difficulties in the resolution and visualisation of low M_r proteins in this study was overcome by replacing the Tris-glycine buffer system with a Tris-tricine buffer system (see Results section 3.2.5).

The Tris-glycine method utilising 10% *T* and 2.5% *C* provided excellent resolution of proteins with M_r between 20-150 kDa and improved the spatial resolution of high M_r proteins ≥ 62 kDa (Figure 3.3). The M_r region between 62-150 kDa displayed 146 proteins and resolved an additional 66 protein spots that could not be efficiently separated by the Tris-tricine buffer system. See section 3.3.1 for spot detection of proteins within the 62-150 kDa region.

3.2.5 Tris-tricine electrode buffer gel

Proteins with M_rs below 20 kDa are poorly resolved using a Tris-glycine Cathode Buffer because they co-migrate with the SDS front and this leads to streaking, smearing and poor staining by Coomassie stains (Westermeir and Naven, 2002). A number of methods have been designed to examine low M_r proteins/peptides. The one used in this study was a method proposed by Fountoulakis et al (1998).

Gel composition

The method of Fountoulakis et al (1998) recommended a resolving gel consisting of a 10.4% *T*, 6% *C* containing either 3.1 M urea or 6.2 M urea. This improved the

resolution of proteins with molecular masses between 10-20 kDa and molecular masses < 10 kDa respectively. See Table 3.2 and Figure 2.1 for a summary of all the different separating gels and stacking gels %T%C compositions and buffer systems utilised in this study.

This study initially separated proteins according to this method. The gels were electrophoresed for 8 1/2 hours at 25 mA per gel. Proteins separated at *T* 10.4% only migrated 2/3 of the way down the length of the gel. Although the low M_r proteins were well resolved, those with M_r > 20 kDa had poor spatial resolution and were compacted and streaked. In the next series of experiments, the *T* value was reduced to 9.5% and the *C* value kept constant. While this *T* concentration efficiently resolved the *H. pylori* proteins with M_r 7-62 kDa, there was approximately an 8 cm gap at the bottom. A *T* value of 8.5% was then selected and this provided the best separation of proteins in the y-axis.

The next parameter examined was the concentration of urea in the polyacrylamide solution. As previously noted, Fountoulakis et al (1998) recommended two urea concentrations, 3.1 M (for proteins M_r 10-20 kDa) and 6.2 M (for proteins M_r <10 kDa). In addition, the present study examined the effect of using a polyacrylamide solution without urea. The gels with the lower urea concentration yielded an improved resolution of protein spots in the M_r 8-20 kDa region. Gels containing 6.2 M urea showed no improvement in resolution and, in some instances, displayed an increase in streaking. Gels without urea displayed some problems in resolving of proteins with M_r < 14 kDa. These problems included streaking and some poorer visualisation of low M_r proteins when using CBB G250. The smallest M_r proteins that could be visualised when using the Tris-tricine technique, 3.1 M or 6.2 M urea concentration, had an observable M_r of about 7 kDa.

Stacking gel

A number of 2-DE studies that had previously examined methods to improve resolution of low molecular mass proteins in proteome analysis utilised a stacking gel (Fountoulakis et al, 1998; Kruft et al, 2001). A stacking gel, which has a lower *T* concentration and crosslinking *C* concentration than the resolving gel, is employed to

stack proteins together and prevent their aggregation prior to entry into the resolving gel. However, for 2-DE (in spite of previous studies [Fountoulakis et al, 1998; Kruft et al, 2001]) the stacking gel is not required, because the proteins are already pre-separated by isoelectric focusing and therefore do not aggregate prior to entry into the resolving gel.

Omitting the stacking gel resolved a few technical, issues these being: 1) problems with reproducibility, the stacking gel did not consistently present a straight polymerisation line for each gel; 2) proteins get caught between the stacking gel and resolving gel and; 3) reduction in additional time-consuming steps in the preparation of two different gel types. In addition, removal of the stacking gel increased the area for resolving proteins. Therefore the final homogeneous Tris-tricine gel formulation of 8.5% *T* 6%*C* that was utilised in this study contained 3.1 M urea without the stacking gel. This improved the resolution of proteins in the second dimension.

3.3 Image analysis

The comparative analysis software utilised in this study was PDQUEST™ version 6.2. For most isolates there were four experimental protocols (see page 80), the exception was strain 26695 for which there was no Tris-glycine pH 6-11 protocol. Strain 26695 was a late addition to this study and due to insufficient freeze-dried sample only three protocols could be completed.

The protocols chosen for strain 26695 were selected based on where the majority of proteins were found to be located in other isolates. The majority of high M_r proteins were located in the pH 4-7 region and for this reason improved resolution and separation of proteins in this region was essential and required separation by the Tris-glycine buffer system (See results section 3.2.4). For resolution of high M_r proteins (> 65 kDa) in the pH 6-11 region there was little difference in the protein patterns between the two buffer systems, therefore the Tris-tricine buffer system was utilised. The Tris-tricine buffer system meant that the proteins in the range 7-150 kDa could be compared between all isolates. See Appendix figures A/1-A.22 for examples of Tris-Tricine gels utilised for comparison of *H. pylori* strains.

Therefore, there were 44 gel images included in the comparative image analysis for each protocol, with the exception of Tris-glycine pH 6-11 which consisted of 40 gel images.

3.3.1 Spot detection

The number of proteins detected within a region or the entirety of each gel was determined by PDQUEST, some manual detection was required. Through this method the number of proteins within the M_r 62 – 150kDa for both Tris-Tricine and Tris-Glycine gels could be determined. After spot detection and spot editing an average of 1168 spots were detected (Table 3.1, Figure 3.4) in the patterns for each of the eleven isolates (stained by Coomassie G250) included in this study. Figure 3.4 provides a visual display of an increase in the number of spots by silver nitrate. An additional 350-400 proteins were visualised by silver nitrate. Forty seven percent of the proteins were basic (pIs range 7-11) and the majority of high molecular mass proteins (≥ 70 kDa) were found to be situated in the pH 4-7 region.

3.3.2 Matchset strategies

Image analysis can be conducted at two levels, level 1 matchset (comparing original images from the disease groups) and the level 2 matchset (comparison of Standards from disease groups). Both of these strategies were examined to identify the better procedure for comparative analyses of gel images.

Level one matchset analysis involved the generation of four large matchsets each containing all the gels for a specific protocol (Figure 2.2A Methods section). Level two strategy involved the construction of high-level matchsets each of which contained four level one matchsets (one for each disease group) (Figure 2.2B Methods section). The major factors that differed between the strategies were: 1) the types of analyses that could be conducted; 2) the number of Boolean sets that were required in order to compare two disease groups; 3) formation of replicate groups; 4) use of original spot image files and; 5) reproducibility of spot matching between gel members.

Boolean sets construction

The advantage of a level two matchset strategy was that it reduced the amount of time required to match and construct analytical Boolean sets for comparison of disease groups. Each level two matchset consisted of four submatchsets (one submatchset for each disease group). Each submatchset functioned as a grouping method. The disadvantage of using the level two strategy for the PDQUEST program involved the type of information exported from one matchset standard to the higher level standard. The matchset strategy utilised needed to be able to perform both quantitative and qualitative analysis between all the gels in this study. The level two matchset strategy was only able to compare both quantitative and qualitative information between strains within a disease group. All information that involved spot volume (quantitative) for each spot in each disease group standard was not exported to the *H. pylori* master standard, this meant that only qualitative analysis could be performed between disease groups. Therefore, both quantitative and qualitative analyses within a matchset required the formation of a level one matchset strategy.

Types of analyses that could be conducted

The second matchset strategy made quantitative comparisons between the isolates of a particular disease outcome more easily handled. Quantitative comparisons between isolates of a particular disease outcome in the first matching strategy involved the complicated construction of quantitative analysis sets coupled with Boolean sets in order to compare observable quantitative differences between these isolates (see Methods section 2.12.8).

Spot matching reproducibility

The major advantage of a level one matchset was the reproducibility of spot matching between member gels. A level two matchset strategy compared the Standards (gel spot files) of each disease group whereas a level one matchset strategy compared the original image files to a computer generated Standard. Any variability in the positioning of the spot centres in the Standards may result in incorrect matching of spots at the second level of the level two matchset. Because level one

matchsets utilised original image files it was visually easier to determine computer errors in matching.

The construction of high-level matchsets had a number of difficulties that could not be resolved by the PDQUEST algorithms currently available. Therefore, it was determined that the better method for comparative analysis between all the isolates was the construction of a large level one matchset (i.e. first matching strategy) which contained all the original images for each protocol.

3.3.3 Proteome reproducibility within replicate groups

A replicate group was formed for each isolate within each protocol. Each replicate group consisted of four gels. In order for a spot to be accepted as valid, this study required that it should be observable in at least three out of four of the gel images for each replicate group. Within the region 9-70 kDa and within pIs 4.25-6.8, variations in the patterns between the replicate group gel images for each isolate were only slight. They often involved very low abundance proteins that were poorly resolved (diffuse appearance). Outside of these ranges there was greater variation observed between replicate gel images, largely because of effects on the gel edges.

Differences at the edges of gels included:

1) distortions at the right edge for pI 4-7 and left edge for pI 6-11 (see Figure 3.3);
2) variability of proteins visible on the bottom edge of the gel. Some proteins just visible on the bottom edge of some gels had migrated off the bottom of other gels and;
3) overcrowding of very high M_rs, 70 – 150 kDa, proteins towards the top of gels in some replicate groups (see Figure 3.3).

Excluding spots in these categories, the average number of proteins that were observed in only a single gel image of a replicate group was generally about 3/1168 (or approximately ≤ 0.2%) of the total visualised proteome for each isolate.

After normalisation of the gel images, no significant variation in spot volume was observed between the replicate images for specific isolates. Protein spots were considered to have a significant difference if the spot volume varied by a factor of more than three.

The determination that a spot was significantly different if the spot volume varied by a factor of more than three was user determined. Less than 124 spots varied by a factor of greater than five, of these only twelve were differently expressed between disease groups. A factor of greater than two generated over 242 differently expressed protein spots, therefore a factor of greater than three was selected as being significantly different.

3.3.4 Gel pattern comparisons between and within groups

Differences between the isolates within a particular disease group were established by comparing one isolate's replicate group to that of another both qualitatively and quantitatively.

Location of the majority of protein differences

Comparative analyses between disease groups and isolates revealed that the majority of protein spot differences were located within the pI range 4-7 and the M_r range 20-62 kDa. Protein spots in the pH range 6-11 were observed to have a greater percentage of positional conservation than within pH 4-7.

Qualitative comparison of proteomes

Qualitative analysis indicated that there was a greater percentage of matching spots between isolates within each particular disease group than there was between isolates from different disease groups. Matches between all three isolates within a disease group were 68% (GU), 68% (NUD) and 67% (DU) (see Table 3.2). Gastritis, which only consisted of two isolates for the disease group, obtained a 78% match between the two expression maps. For any two isolates within a group their proteomes matched by 76-77% (Table 3.3). Comparison of all the isolates in the study showed that those from other groups gave lower match percentages. For example, GU1 and GU2 were 77% similar whereas if DU1/DU2 DU3 patterns were matched to GU1

patterns the matches were only 67-69%. This varied depending on the isolate. Some isolates were more closely related to the GU isolates than were others (see Table 3.3).

Quantitative comparisons of proteomes

A quantitative comparison was also performed for all four groups. Quantitative analysis indicated that there was a greater percentage of matching spots between isolates within a particular disease group when compared to those isolates from other disease groups. Matches between all three isolates within a disease group were 48% (GU), 48% (NUD), and 46% (DU). However, this was not indicated from the pattern comparisons of the gastritis isolates, which had a greater percentage match of its proteome with other gastric isolates. Matching of spots by quantitative analysis between any two isolates within a disease group was lower than that by qualitative analysis by 20%. This was due to the exclusion of the conserved portion of the proteome that did not vary by a factor greater than three or more. Comparative quantitative analysis showed that 48% of the proteome of isolates within a disease outcome could be matched.

3.3.5 Pattern expression comparisons

A comparison of the eleven different patterns revealed high proteomic variability. Whereas 471 spots (40.6% of the proteome) migrated on gels at the same position for all isolates, there were at least 14 positional pI shifts (~1.2% of the proteome), 12-14% differentially unique and 15-17% absent.

The protein spots within the gel patterns for each isolate could be categorized into four subdivisions:

1) strain-specific;
2) positionally conserved spots which were either,
    i) quantitatively conserved (volume varied by less than a factor of three)
    ii) differentially expressed (whose volume varied by a factor of three or more);
3) unique to a specific disease group; and

4) absent from a specific disease group.

Strain-specific Spots

These formed a very small subset of the *H. pylori* proteome. Prior to the inclusion of the strain 26695 in the study, the strain-specific proteins comprised approximately 2% of the total. However, after the addition of strain 26695 the strain-specific spots comprised less than 2% of the total protein spots.

The numbers of strain specific proteins for each isolate were GU1 24, GU2 32, GU3 30, GI1 30, GI2 22, NUD1 29, NUD2 19, NUD3 22, DU1 33, DU2 26 and DU3 22.

Positionally Conserved Protein Patterns

Comparison of the protein expression maps of the eleven isolates revealed that 471 spots (40.6%) were found in corresponding positions on gels (Table 3.3). Comparative quantitative analyses between all the isolates revealed differences that could be further categorised into three subgroups

1) quantitatively conserved spots whose volume varied by less than a factor of three. These were 25% of the *H. pylori* proteome (291 spots);
2) protein spots which quantitatively differed by a factor of more than three: were 12% of the proteome (142 spots). While certain proteins were specific to particular disease groups others were shared with another disease group and;
3) Protein spots that differed quantitatively but were not specific to any disease group were 3 % of the proteome (35 spots).

Unique and Absent Proteins

The present study was concerned with determining which proteins were specific for a disease outcome. Each disease outcome group consisted of three isolates, with the exception of GI (two isolates). The initial approach was to examine all possible three-isolate subsets (C^{11}_4 =164) from the total 11 isolates. For proteins to be classified as unique to a particular disease group they should only be found within

the particular three-isolate subset forming that disease group, and should be absent from all others. Furthermore, it was predicted that the three three-isolate subsets having the largest number of uniquely shared spots should correspond to the three disease groups. One hundred and sixty one of the three-isolate subsets included five or fewer such spots (see Table 3.4). The three-isolate subsets that contained the largest number of unique spots were indeed found to be those that corresponded to the three disease specific groups. It was found that the GU subset had 58 unique proteins the NUD 54 and the DU 55 (Table 3.6).

GI was the only disease group that consisted of two isolates, both of which had been examined by previous researchers (GI1 [Lock et al, 2001] and GI2 [Jungblut et al, 2000]. Therefore, to determine if there were any possible unique GI proteins, the two-isolate subsets were also examined (C^{11}_2 = 55 subsets). Fifty four of the two-isolate subsets contained six or fewer spots (summarised in Table 3.5). It was found that the largest number of unique proteins in a two-isolate subset was found in the GI group (54 proteins) (summarised in Table 3.6).

The present study was also concerned with determining which proteins were specifically absent for only one disease outcome. In this case, the initial approach was to examine all possible eight-isolate subsets. For proteins to be classified as specifically absent from a particular three-isolate isolate subset (C^{11}_4=164 such subsets) the proteins should not only be absent from the subsets but also be present in all the remaining isolates. It was predicted that the eight-isolate subsets, which had the largest number of absent spots would correspond to the three disease groups. One hundred and sixty one of the eight-isolate subsets contained four spots or fewer (see Table 3.7). The eight-isolate subsets with the largest number of absent spots were indeed found to be those that belonged to the three disease specific groups. These were GU (39 proteins), NUD (44 proteins), and DU (38 proteins) (summarised in Table 3.6).

For analysis of GI proteins, nine-isolate subsets were examined. It was found that the largest number of specifically absent proteins in a nine-isolate subset was that

found in the Gastritis group(C^{11}_9= 55 subsets and 34 proteins) (see Tables 3.8 and 3.7).

Standard image utilised to represent significant differences

The standard image is a representation of all the spots located in all the images within a disease group. A reference gel was used to generate the standard image and therefore all initial spots had the appearance of the reference image. The reference gel did not contain all the spots within the disease group, therefore spots from another strain's representative image needed to be added to the standard if it was not located on the reference strain's image. All additional spots added from other images in a matchset were also similar to its corresponding spot.

For a spot to be considered unique to a disease group it had to be in images of all three strains within the disease group. The spots in the standard image also included those that were located only in one or two strains of the disease group. This meant that visual comparisons of each disease group's standard images might reveal a false unique spot.

In order for a protein spot to be quantitatively significant it had to have a spot intensity greater than three fold for all three strains in a matchset. Since all initial spots in the standard image was based on the reference image this did not necessarily mean that the other images included in the matchset had the same appearance or spot intensity. See the appendix for examples of gels utilised in matchsets. Reference images were;

- GU group – Figure A.1 and Figure A.12,
- NUD group – Figure A.4 and Figure A.15,
- DU group – Figure A.7 and FigureA.18,
- GI group – Figure A.10 and Figure A.21.

This study generated two types of standard images the first image standard based on the level one strategy and the second based on the level two strategy. The level two strategy standards were used to display significant differences between disease

groups. The level one strategy standard is representative of all spots for all eleven strains and was utilised for analysis. If the level one strategy standard image was used it would have been impossible to observe unique and absent protein spots for each disease groups because all spots would have been represented. Therefore the level two matchset strategy standard images were utilised to display the differences between disease groups and the actual comparisons were performed using the level one matchset strategy.

3.4 Cluster analysis of protein patterns

Cluster analysis was based on PDQUEST pattern match results (including normalised spot volume, M_r and pI). Dendrograms were derived using two methods (see Figure 3.5). The first method used the quantitative data (normalised spot volume) obtained from image analysis, while the second used the qualitative information (normalised spots absent or present). In dendrogram A, which was derived from quantitative data, it was observed that isolates could be sorted into two major clusters i.e. gastric (including GU, NUD, and GI isolates) and duodenal (DU isolates) (Figure 3.5, Table 3.9). In dendrogram B, in which clusters were based on qualitative data (proteins absent or present), isolates were divided into four subclusters: GU, GI, NUD, and DU. The first three subclusters belonged to the gastric major cluster and the final subcluster of DU isolates formed the duodenal major cluster (Table 3.9). Therefore, both dendrograms indicated that isolates fell within two major clusters: gastric and duodenal.

3.5 Protein Identification

The Coomassie spots generated by 2-DE using the IPG strip of pH 4.0–7.0 and 6.0–11.0 were numbered, excised, and destained, followed by in-gel digestion using trypsin for peptide fingerprinting. The mass of the resulting peptide mixtures was measured by MALDI-TOF-MS. Mass spectrometric analysis of proteins spots was largely conducted on specifically unique/specifically absent proteins and those spots that differed in volume by a factor greater than three. Specifically absent in this case is determined by the presence of the protein at a particular x/y position in a gel for

eight/nine isolates but not for the other two/three isolates. The spots that were excised from at least two different isolates within a disease group from the same position on gels. For example if a specifically absent GU protein was examined this would require excision of the spot from the same site from GI1, NUD3 and DU1. Spectrometric analysis on specifically unique proteins was conducted on spots excised from the same site from two different isolates within the same disease group. For quantitative differences (by a factor greater than three) spots were excised from at least two isolates, each isolate had to belong to a different disease outcome. These spots were found in the same 2-DE gel location for all isolates. Two hundred and forty spots were analysed in this project (see Table 2.2 for spots excised). From these 240 spots, 181 protein spots were identified in this study and nine spots included in this study were also identified in a previous study conducted in this laboratory. Fifty spots returned a "no match" in the database searches.

The present study found a similar degree of high proteomic diversity to that indicated by past *H. pylori* comparative studies (Jungblut et al, 2000; Enroth et al, 2000) with only approximately 40% of the proteome being positionally conserved on 2-D gels. The 190 proteins identified by the comparative study fall into the following categories 1) conserved between all isolates (96 protein spots); 2) proteins that differed quantitatively by a factor of three between groups (24 protein spots); 3) proteins unique to and specifically absent from particular groups (42 protein spots) (see Table 3.10 for a list of tables and figures), 4) Flagellin proteins (8 protein spots); and 5) protein spots that may have positional pI shifts that varied from a pI of 0.5-1 were also identified (20 protein spots). Positional pI shifts refers to spots that shift on the x-axis (i.e. pI shift) with no detectable shift on the y-axis (M_r conserved). Four protein spots that had pI shifts were previously identified by Jungblut et al (2000). These were ribosomal protein L7/L12, GroEL, superoxide dismutase (Sod) and alkyl hydroperoxide reductase. Another three protein spots were identified in this study that also appeared to have positional pI shifts. These proteins were fructose bi-phosphatase, DNA-directed RNA α subunit, and neutrophil activating protein (Figure 3.13 and Table 3.16). Several proteins initially believed to be disease specific proteins were selected for comparison and analysed by MALDI-TOF. On subsequent identification, these spots were found either to have a positional shift or

to be isoforms in a serial charge train. Serial charge trains are protein isoforms that appear as a train of spots, usually horizontal, that differ in pI and/or M_r. These may indicate post-translational modifications such as glycosylation or phosphorylation.

Protein differences between the four disease groups, GU, DU, NUD and GI belonged to five categories (classifications in NCBI database). These categories are proteins that are involved in:

- metabolism;
- cellular processes and signalling;
- information processing;
- pathogenesis and;
- poorly characterised or unknown proteins

Forty eight percent of the spots identified in this study were located in the poorly characterised and unknown proteins. The other 52% of proteins had been categorised. However, many of the functions of these proteins are still poorly understood.

Many of the differences between isolates were located in the categories of metabolism and cellular processes and signalling. The differences in the proteins between the GU group and the other groups was mainly in the category of metabolism. These differences included thioredoxin reductase (GU absent spot number 121 [Table 3.14]), thioredoxin (GU up-regulated spot numbers 157 and 158 [Table 3.15]), an iron (III) ABC transporter (NUD absent spot number 131 [Table 3.14]), catalase (KatA) (DU down-regulated spot numbers 140-142 [Table 3.15]) and non-haeme iron containing ferritin (Pfr) (DU and NUD down-regulated spot numbers 145 and 146 [Table 3.15]).

The TrxA/TrxB, KatA and Pfr proteins are essential elements in protection against hydrogen peroxide. Other proteins, such as homoserine dehydrogenase, S-adenosylmethionine synthetase 2, HP0431, and JHP0295 (see spot numbers 151 and

144 [Table 3.15]; 136 and 130 [Table 3.14] respectively]) represent differences in carbohydrate transport and metabolism.

The FindMod program and The MS-FIT post-translational modification search parameter were used to identify probable post-translational modifications. Nearly 30% of all proteins identified were post-translationally modified, many with more than one modification (Table 3.17). Many of the spots of interest were identified as multiple isoforms, probably resulting from post-translational modifications to the primary gene products.

3.6 Comparison of theoretical and observed data

Figures 3.14A and 3.14B show scatter diagrams comparing the observed properties of the 190 identified proteins on 2-DE gels with the predicted properties of the corresponding theoretical gene products. In Figure 3.14A, observed M_r is plotted against theoretical M_r. When total data is related to an equivalence plot (slope 1.0, passing through the origin), the correlation coefficient is 0.983 which is significant at the $P<0.01$ level. From the data utilised in this study the majority of proteins showed M_rs closely to their actual M_rs (±15%) with eleven outliers. In Figure 3.14B, observed pI values for the 190 identified proteins are plotted against theoretical pI values. The equivalence plot correlation coefficient is 0.96 which was significant at the $P<0.01$ level (Note this figure was probably lower but due to the limits set by the SPSS program lower figures were not displayed). The majority of proteins analysed correlated closely to the actual pI (±1.0) however, there were 14 possible outliers. The set of proteins predicted for *H pylori* and the set actually observed both displayed a bi-modal distribution with respect to pI (see Figure 3.15). In each case, there were few proteins with pI near 7.0.

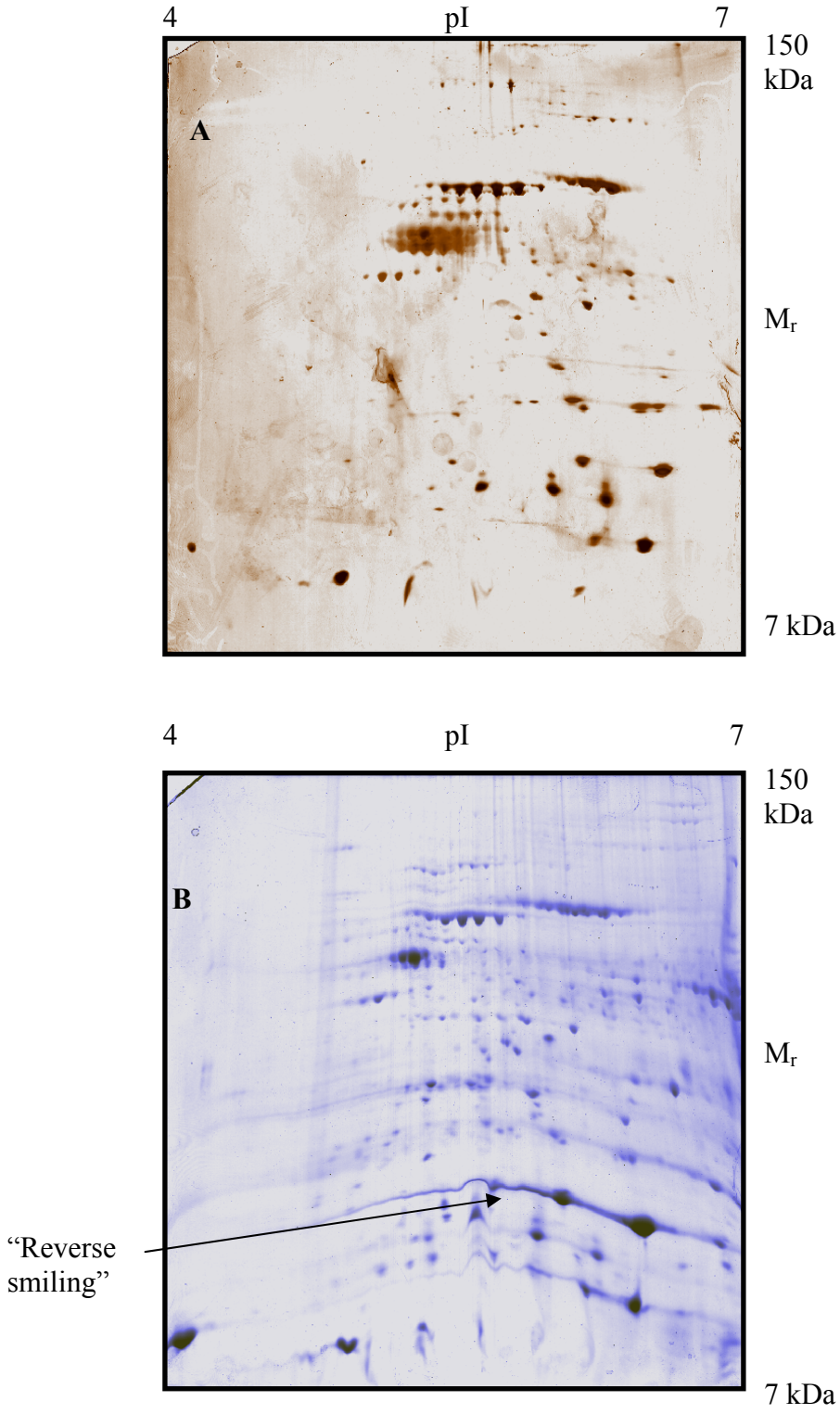


Figure 3.1 Separation of Proteins by Gradient Gels.

H. pylori proteins separated on 10-16% gradient polyacrylamide gels (Bio-Rad Ready-Cast gels) with Tris-glycine electrode buffer for 9 hours at 25 mA per gel. Proteins were stained in A) by silver nitrate and in B) by Coomassie Brilliant Blue G250. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

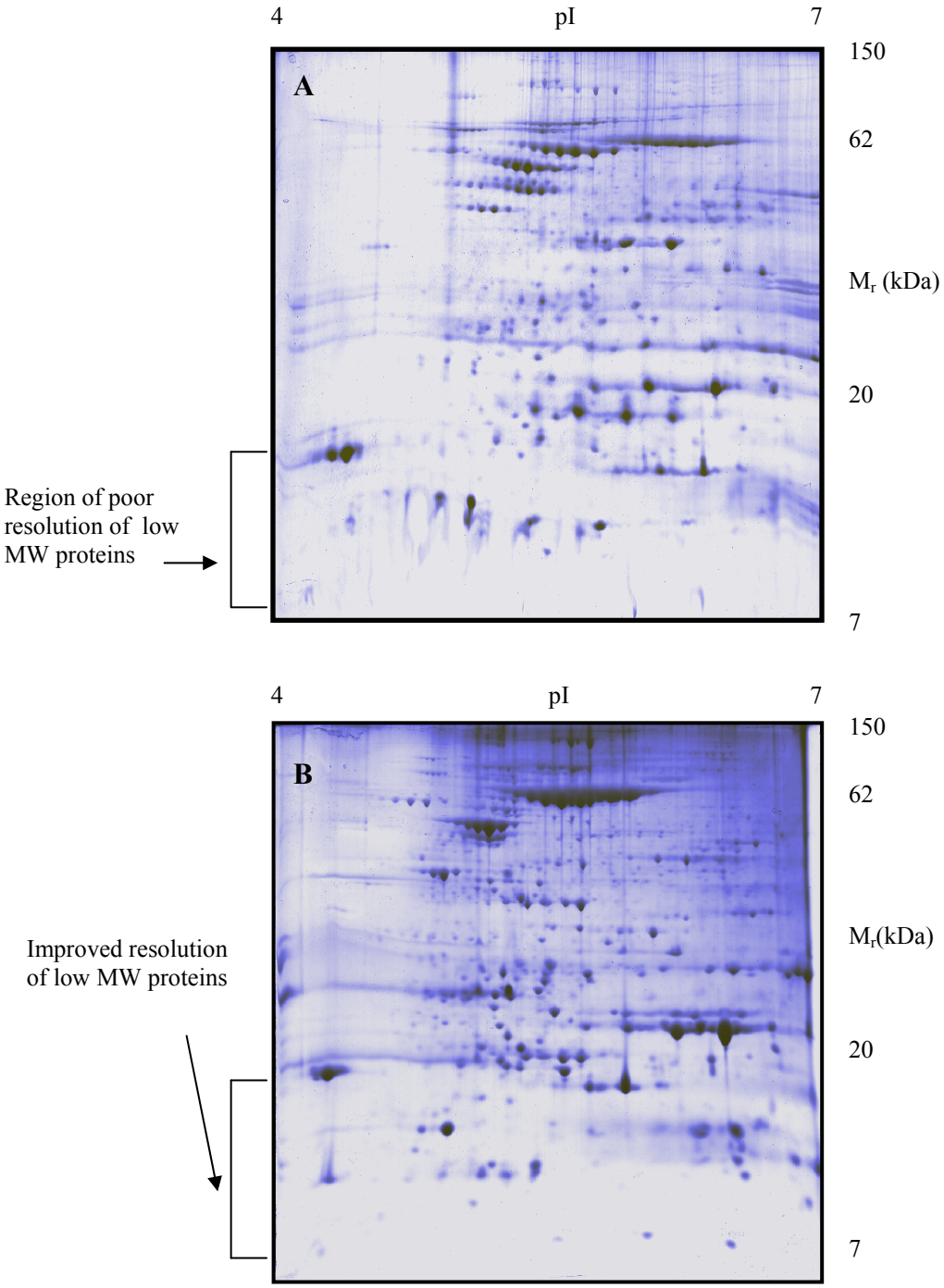


Figure 3.2 Comparison of Low Molecular Mass Separation Techniques.

A) H. pylori proteins separated in the second dimension on a 13% polyacrylamide slab gel using 2.5% PDA crosslinker, Tris-glycine electrode buffer and electrophoresed at 25 mA per gel for 6 hours. B) *H. pylori* proteins separated in the second dimension on an 8.5% polyacrylamide slab gel using 6% Bis crosslinker, Tris-tricine electrode buffer and electrophoresed at 40 mA per gel for 7 1/2 hours. Both gels were stained with Coomassie Brilliant Blue G250. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

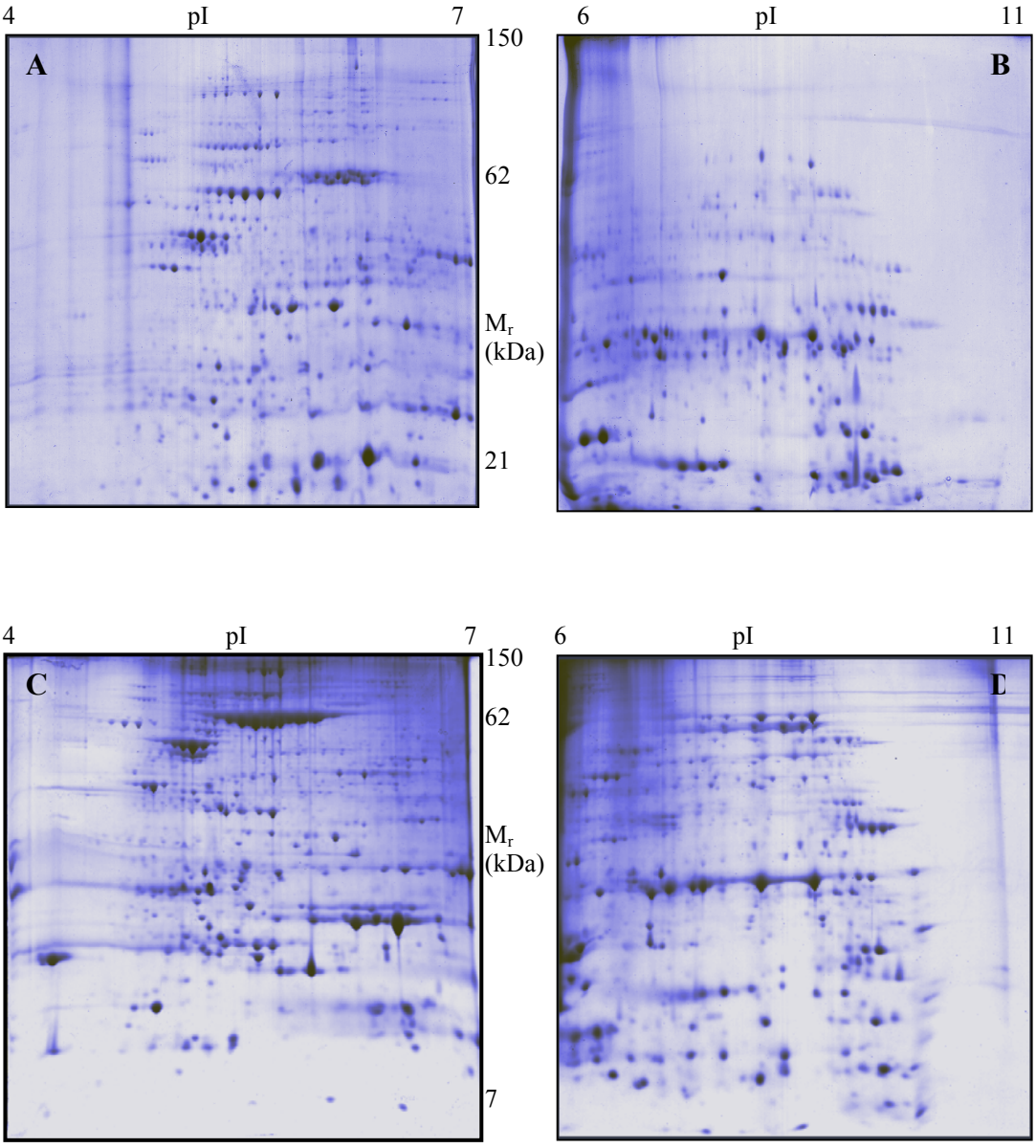


Figure 3.3 Tris-glycine and Tris-tricine gels

H. pylori proteins separated by 10% polyacrylamide gels using Tris-glycine buffer system at 20 mA per gel (Gels A and B). Proteins separated by 8.5% polyacrylamide gels using Tris-tricine buffer systems (Gels C and D). Gels A and C were separated in the first dimension by pH 4-7 IPG strips and Gels B and D were separated by pH 6-11 IPG strips. These gels were stained by the modified Neuhoff method. Gels separated by the Tris-glycine system show improved separation of proteins between 60-150 kDa. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

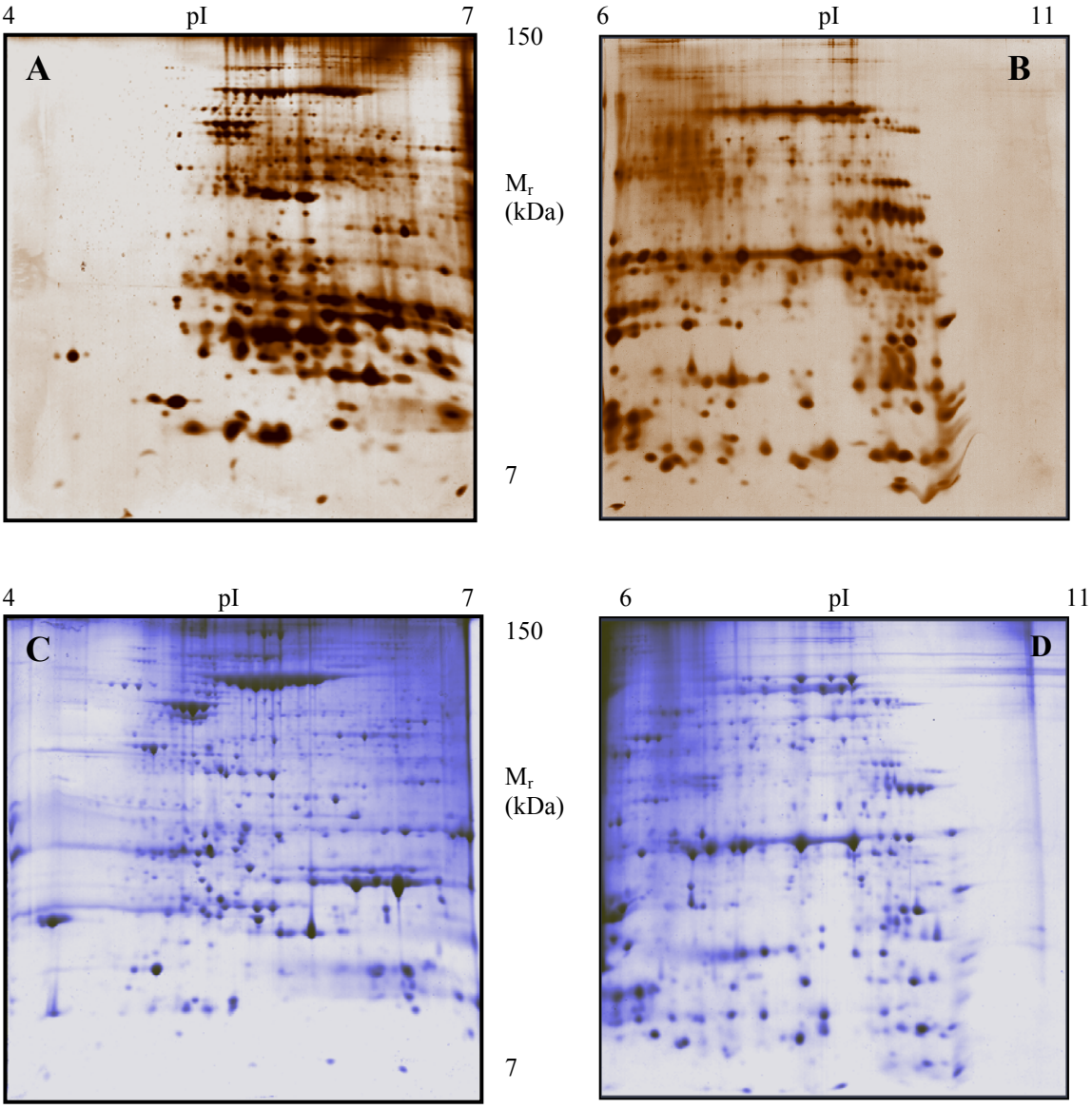


Figure 3.4 Silver and Coomassie stained gels

H. pylori GU1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer at 40 mA per gel for 7 1/2 hours. Proteins in Gel A (pH 4-7 IPG strip) and B (pH 6-11 IPG strip) were two overlapping gels stained with silver nitrate. Proteins in Gels C (pH 4-7 IPG strip) and D (pH 6-11 IPG strip) were stained with Coomassie Brilliant Blue G250. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

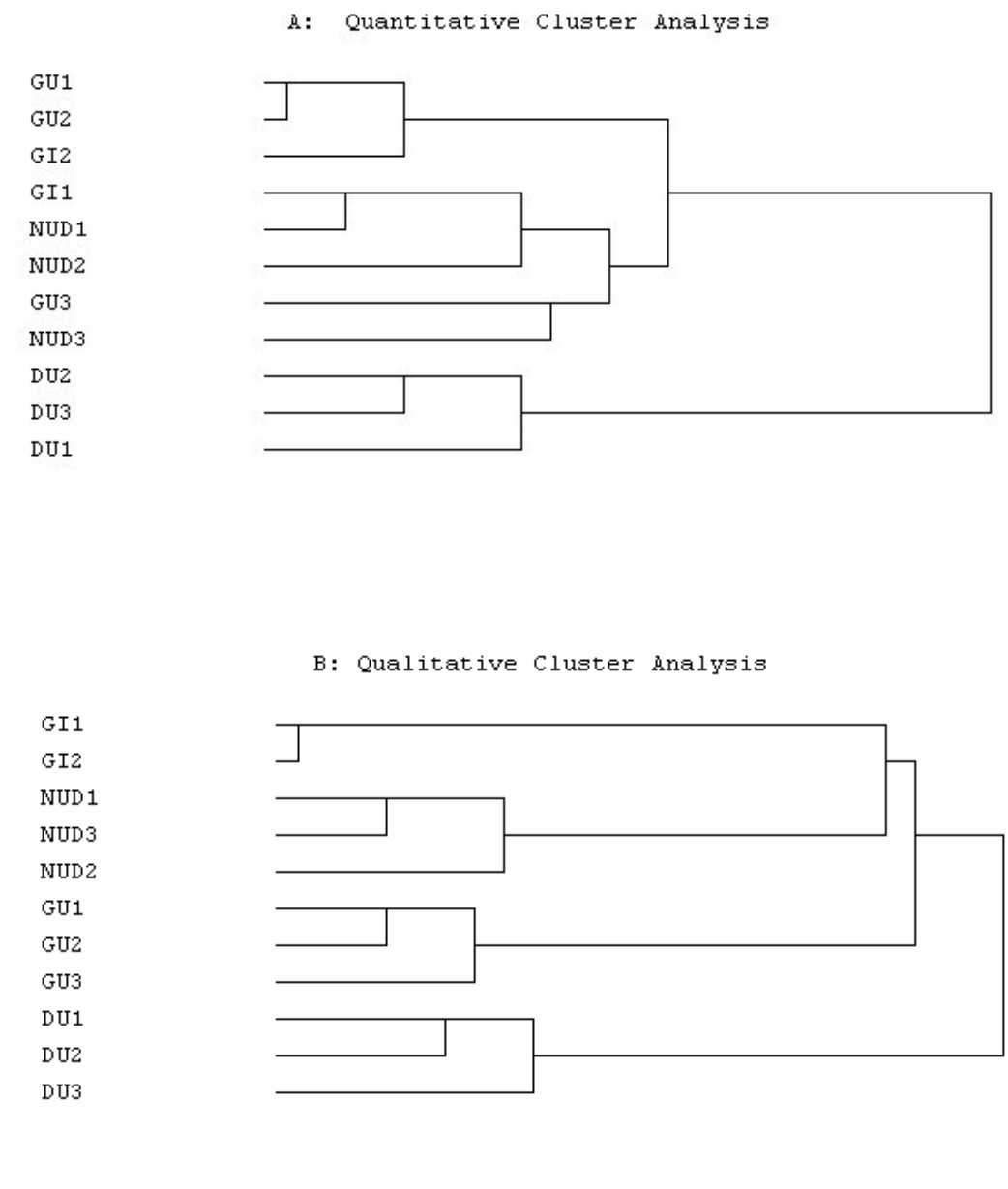


Figure 3.5 Cluster analysis dendrograms

Two dendrograms showing the relationships among eleven *H. pylori* isolates. Dendrogram A was produced by the method of average linkage using quantitative data, and B was produced by the method of average linkage using binary data (absent or present information). On the left side are the isolate identities

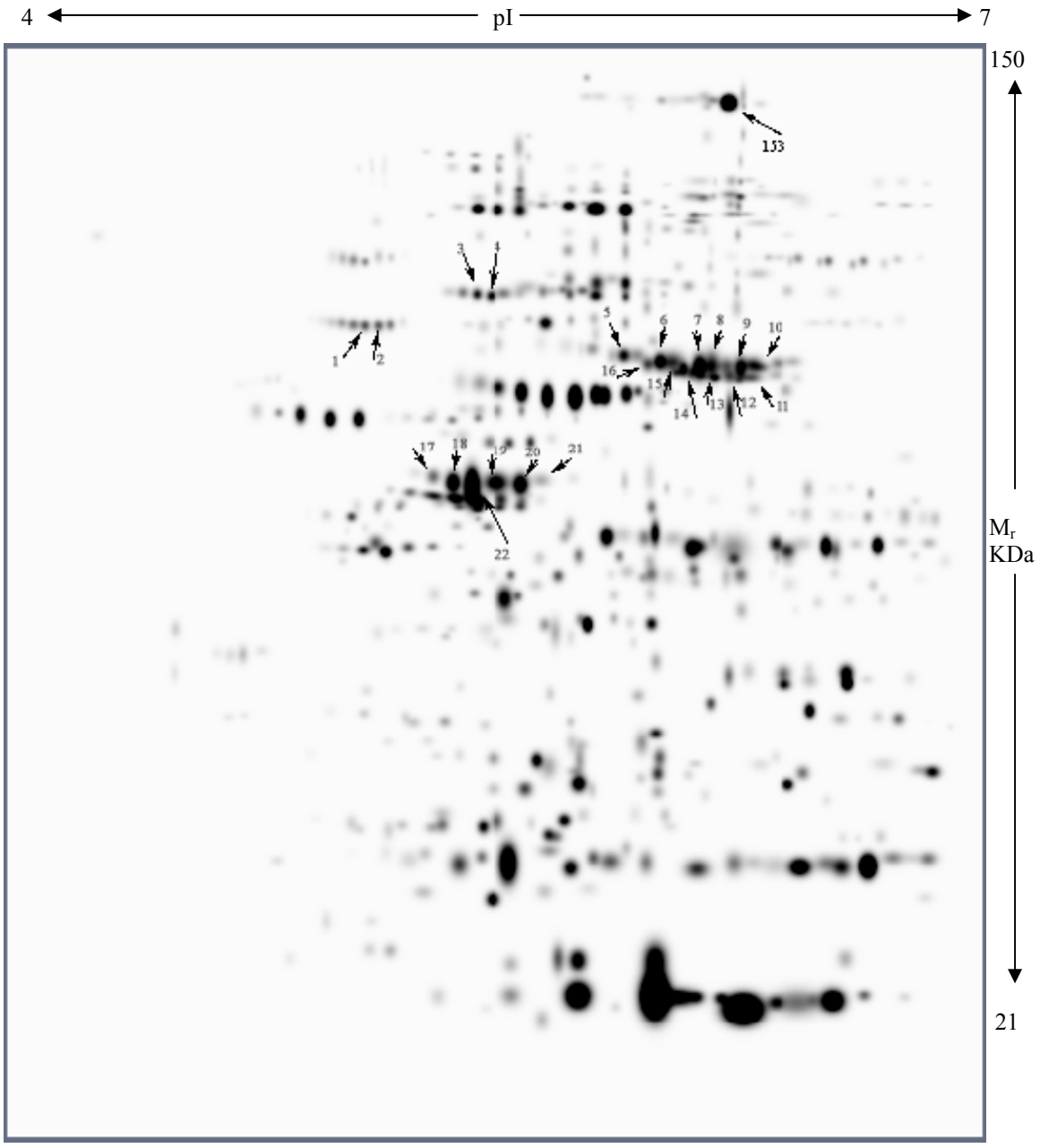


Figure 3.6 Positional and quantitatively conserved proteins 46-150 kDa

A digitised representation of the protein standard map of GU *H. pylori* isolates within pH 4-11 processed by PDQUEST software. The numbered features are those that were harvested for peptide mapping. Spots 1–22 represent the positional and quantitatively conserved proteins that were identified. Spot 153 (see Table 3.15) is a differentially expressed protein its inclusion into this figure is due to the improved separation area and the fact that this spot is seen in all the isolates. Data for these numbered spots are collated in Table 3.12.

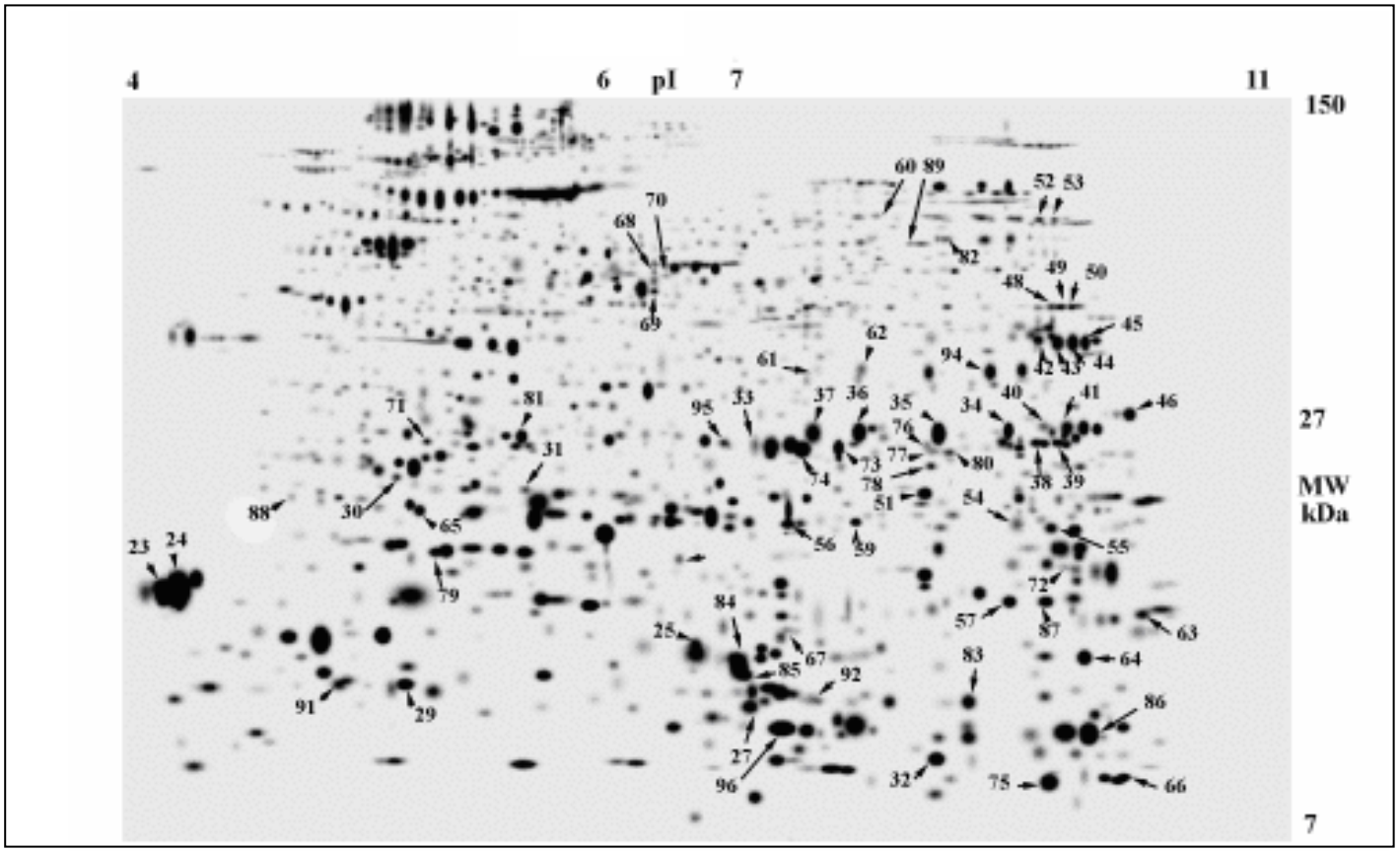


Figure 3.7 Positional and quantitatively conserved proteins 7-46 kDa

A digitised representation of the GI *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 23–96 represent the positional and quantitatively conserved proteins that were identified. Data for these numbered spots are collated in Table 3.12.

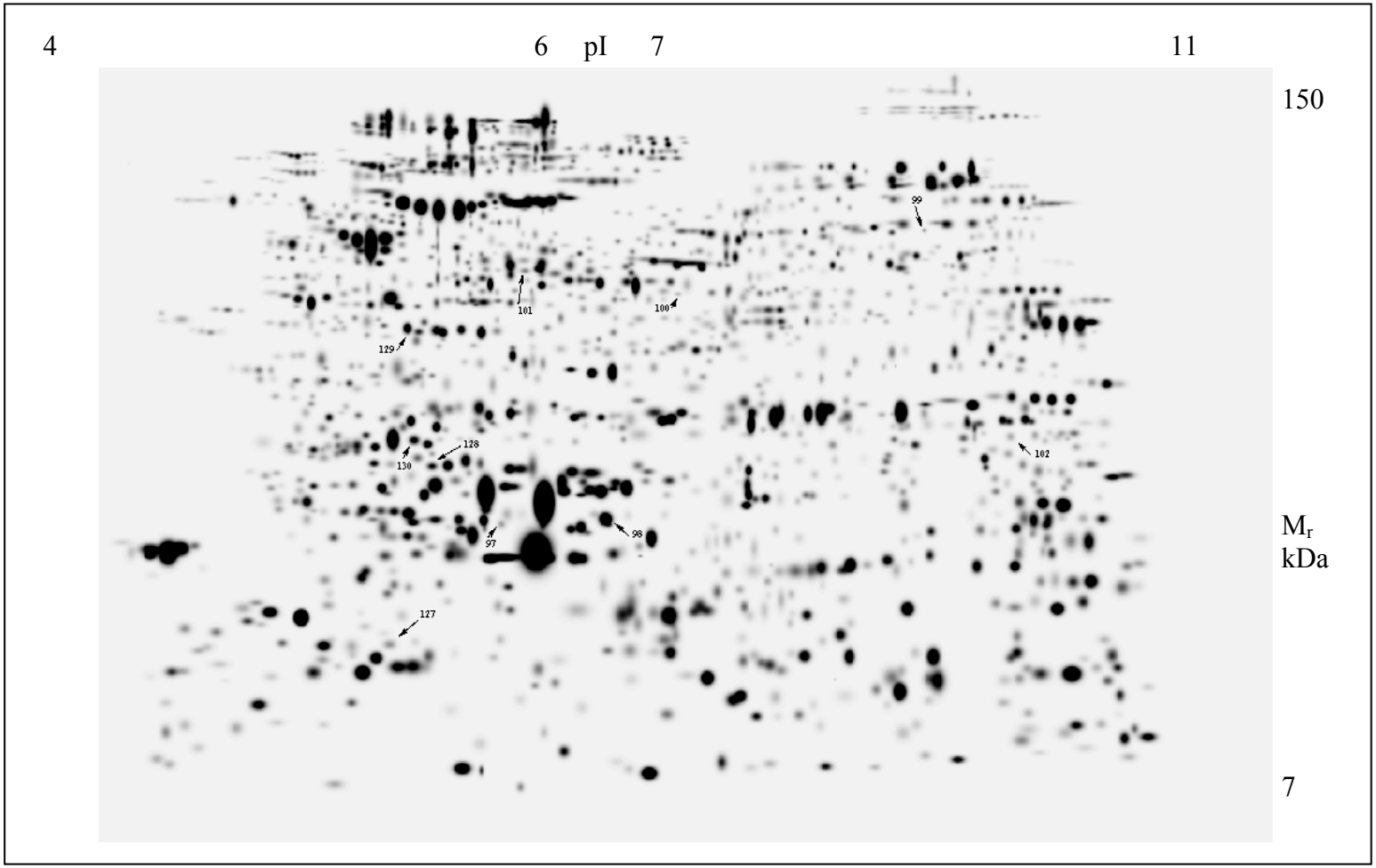


Figure 3.8 Unique (GU) and Absent (DU) spots

A digitised representation of the GU *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 97–102 represent the GU specifically unique proteins that were identified. Spots 127-130 represent the DU absent proteins that were identified. Data for these numbered spots are collated in Table 3.13 and 3.14.

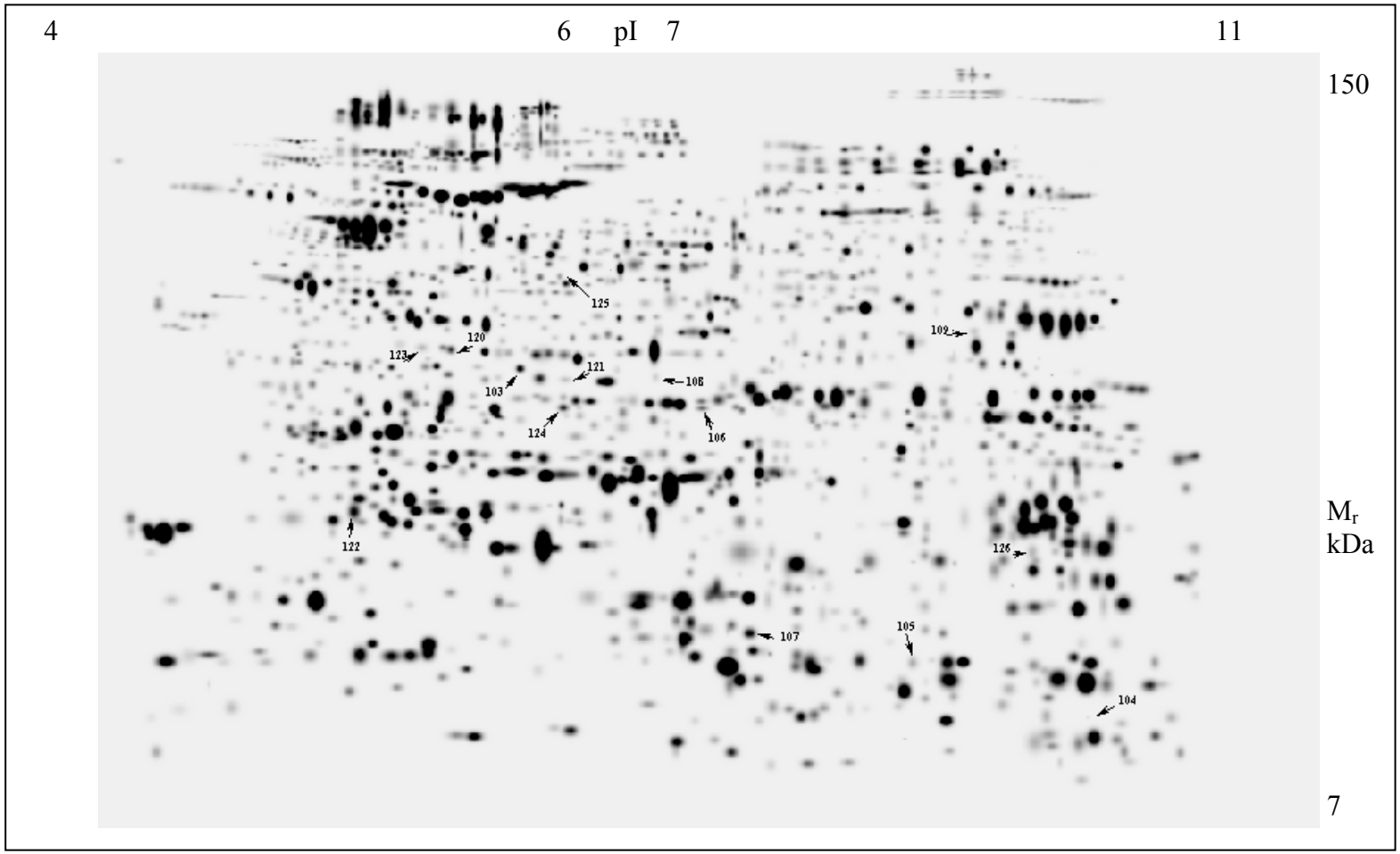


Figure 3.9 Unique (DU) and Absent (GU) spots

A digitised representation of the DU *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 103–109 represent the DU specifically unique proteins that were identified. Spots 120-126 represent the GU absent proteins that were identified. Data for these numbered spots are collated in Table 3.13 and 3.14.

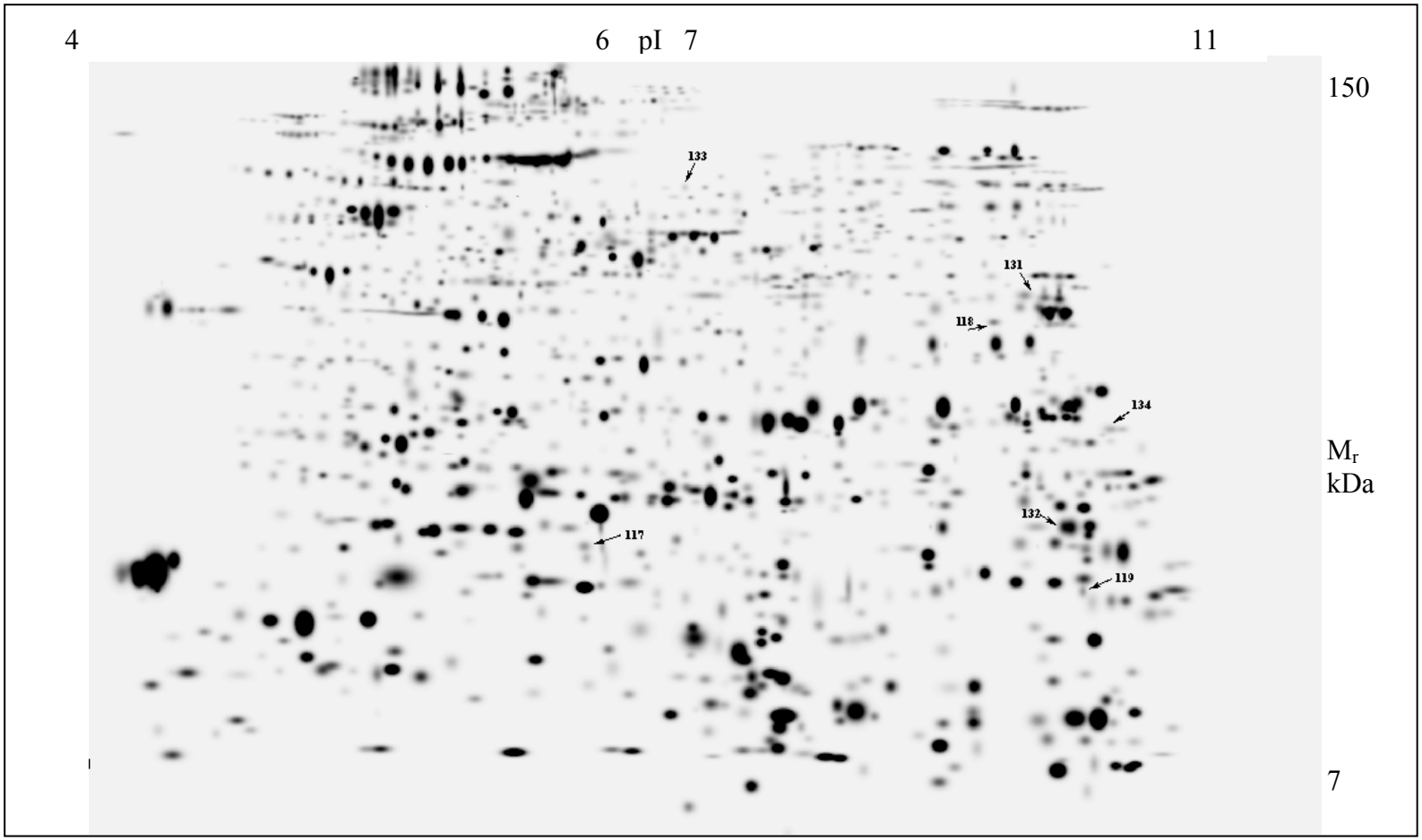


Figure 3.10 Unique (GI) and Absent (NUD) spots

A digitised representation of the GI *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 117–119 represent the GI specifically unique proteins that were identified. Spots 131-134 represent the NUD absent proteins that were identified. Data for these numbered spots are collated in Table 3.13 and 3.14

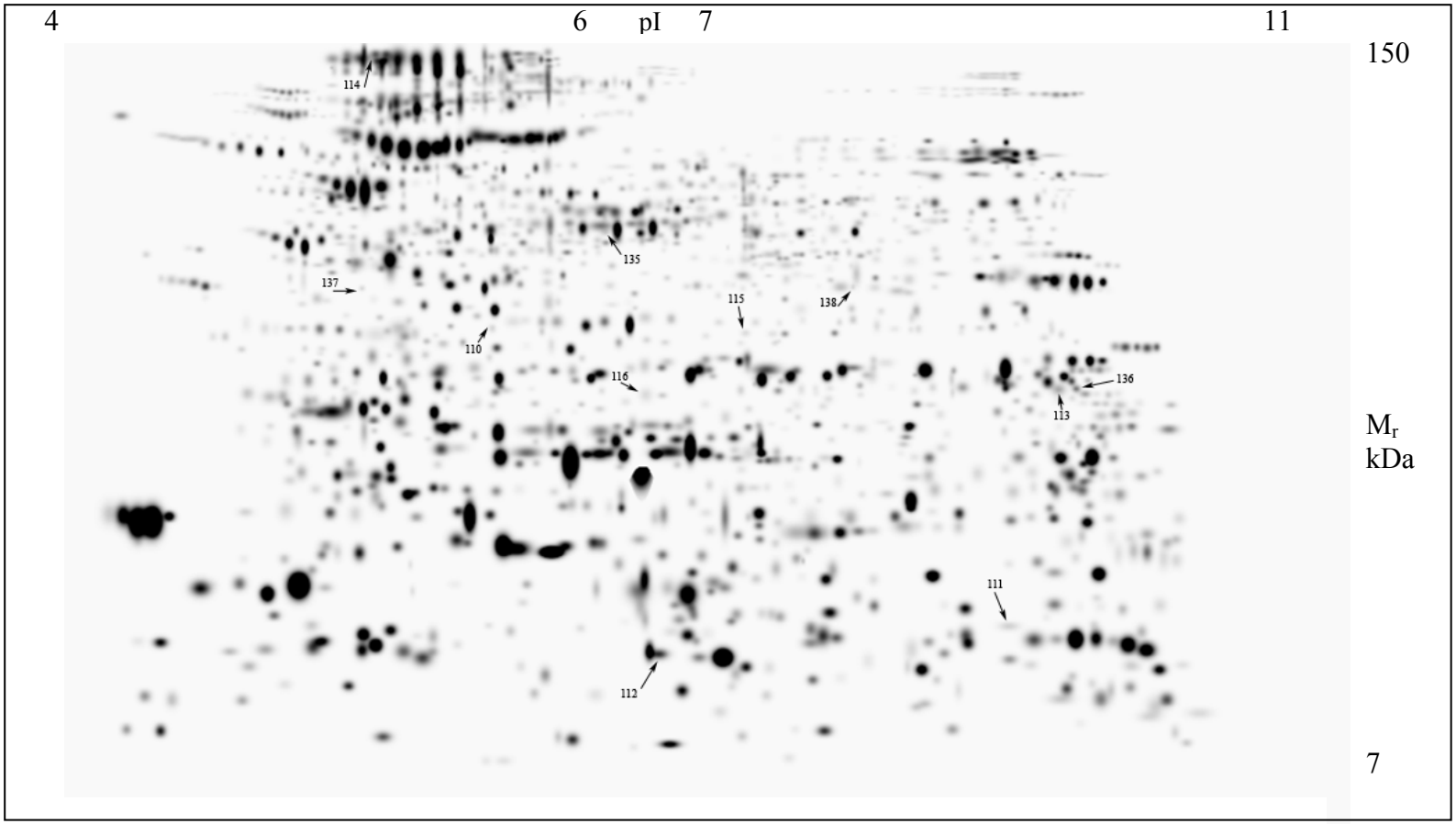


Figure 3.11 Unique (NUD) and absent GI spots

A digitised representation of the NUD *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 110-116 represent the NUD specifically unique proteins that were identified. Spots 135-138 represent the GI absent proteins that were identified. Data for these numbered spots are collated in Table 3.13 and 3.14.

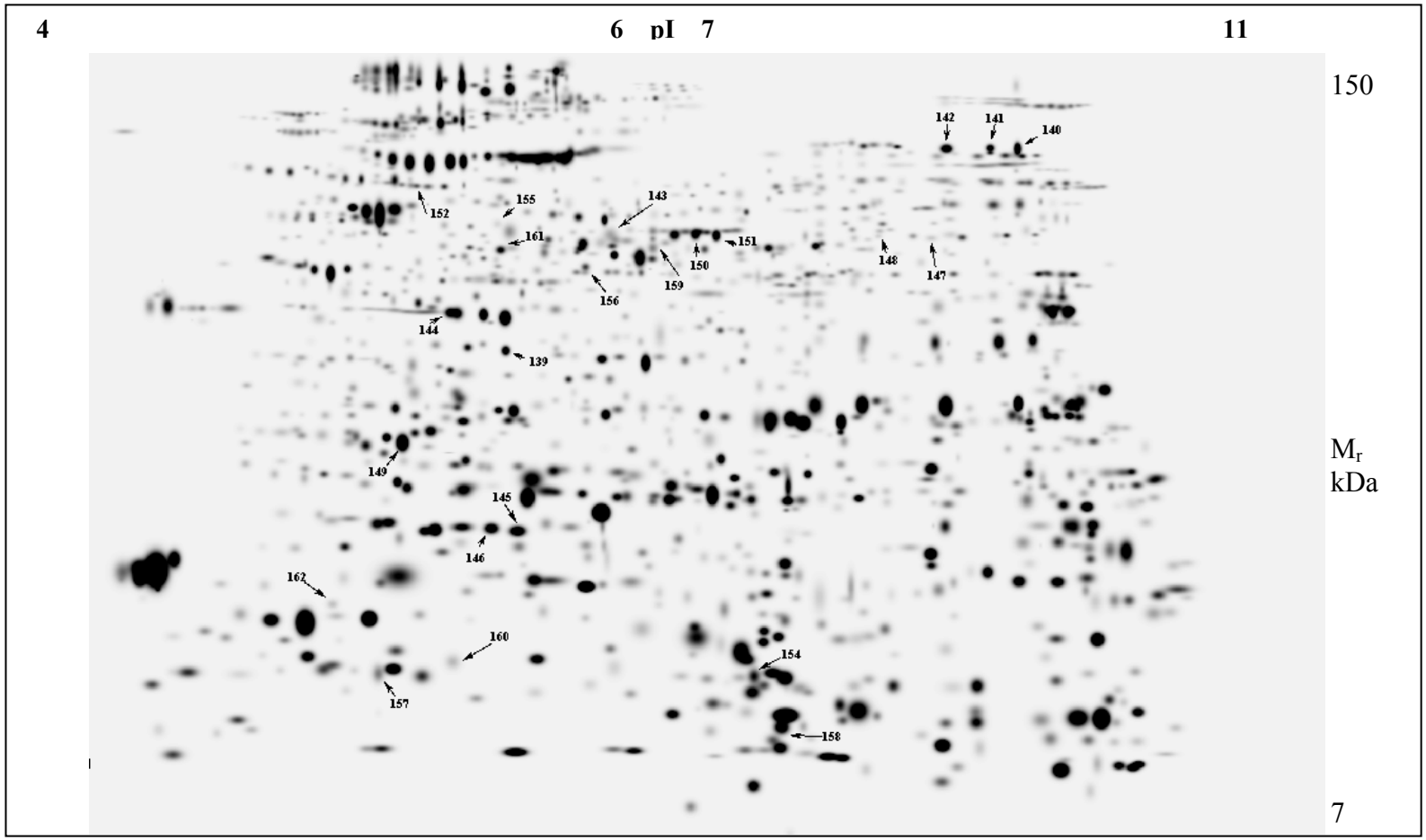


Figure 3.12 Proteins that are differentially expressed

A digitised representation of the GI *H. pylori* isolates standard map within pH 4-11 processed by PDQUEST software. The features that have been highlighted are those that were harvested for peptide mapping. Spots 139–162 represent the differentially expressed proteins (positionally conserved) that were identified. Spot 153 is located on Figure 3.6. Changes in spot intensities (factor greater than three) are given using GI isolates as the reference. Data for these numbered spots are collated in Table 3.14

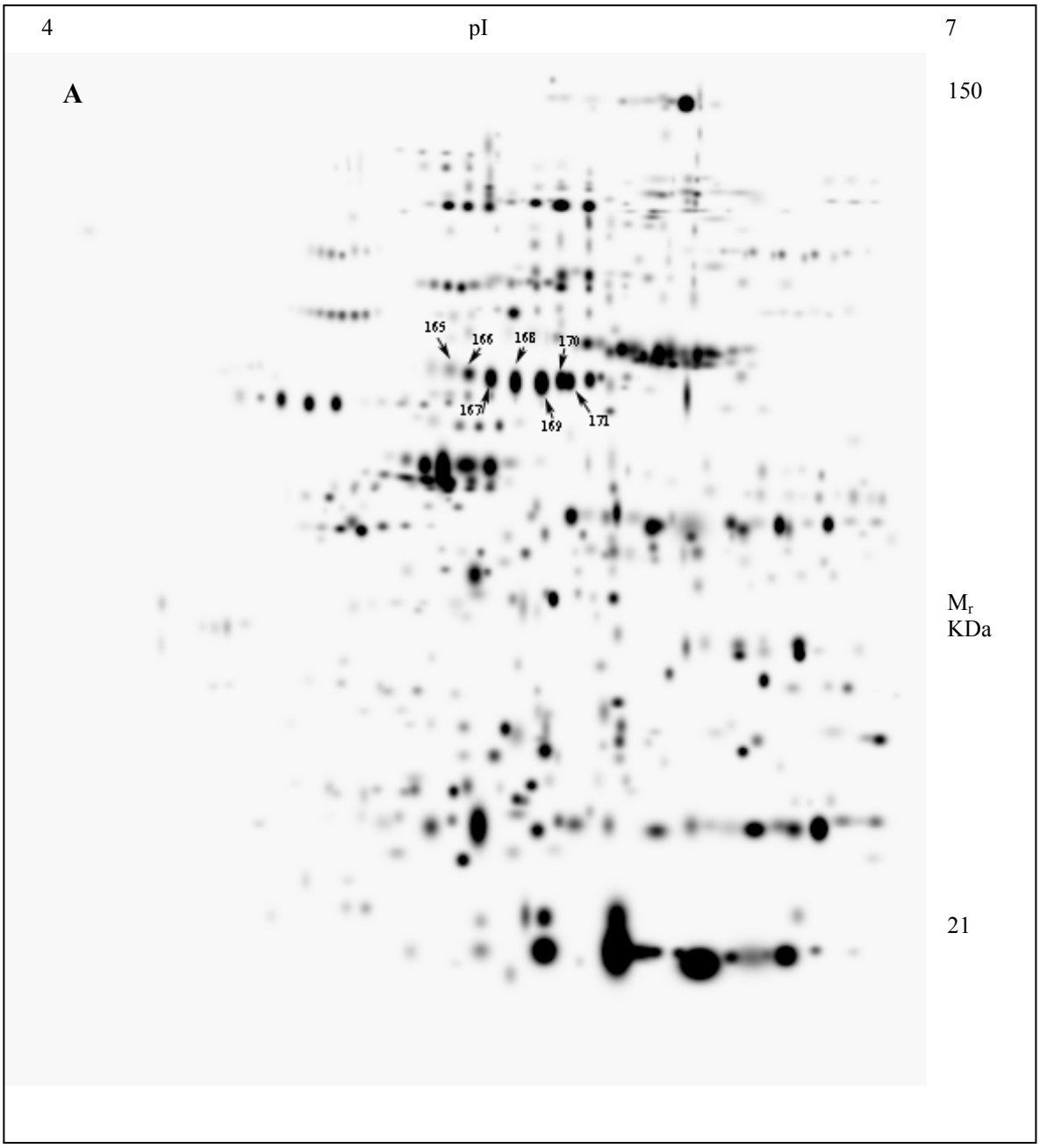


Figure 3.13 Proteins with pI positional shifts

Figure 3.13A is a digitised representation of the two-dimensional gels of the GI *H. pylori* isolates that represents spots 164, 165 and 173 – 183 within the pH 4-7/7-50 kDa range. These spots represent proteins that appear to be positionally shifted. The 11 features that have been highlighted are those that were identified by peptide mapping. The GI master map was the reference map used to compare which are the protein spots that may have shifted between the proteome maps of the isolates. Protein spots detected by PDQUEST are shown.

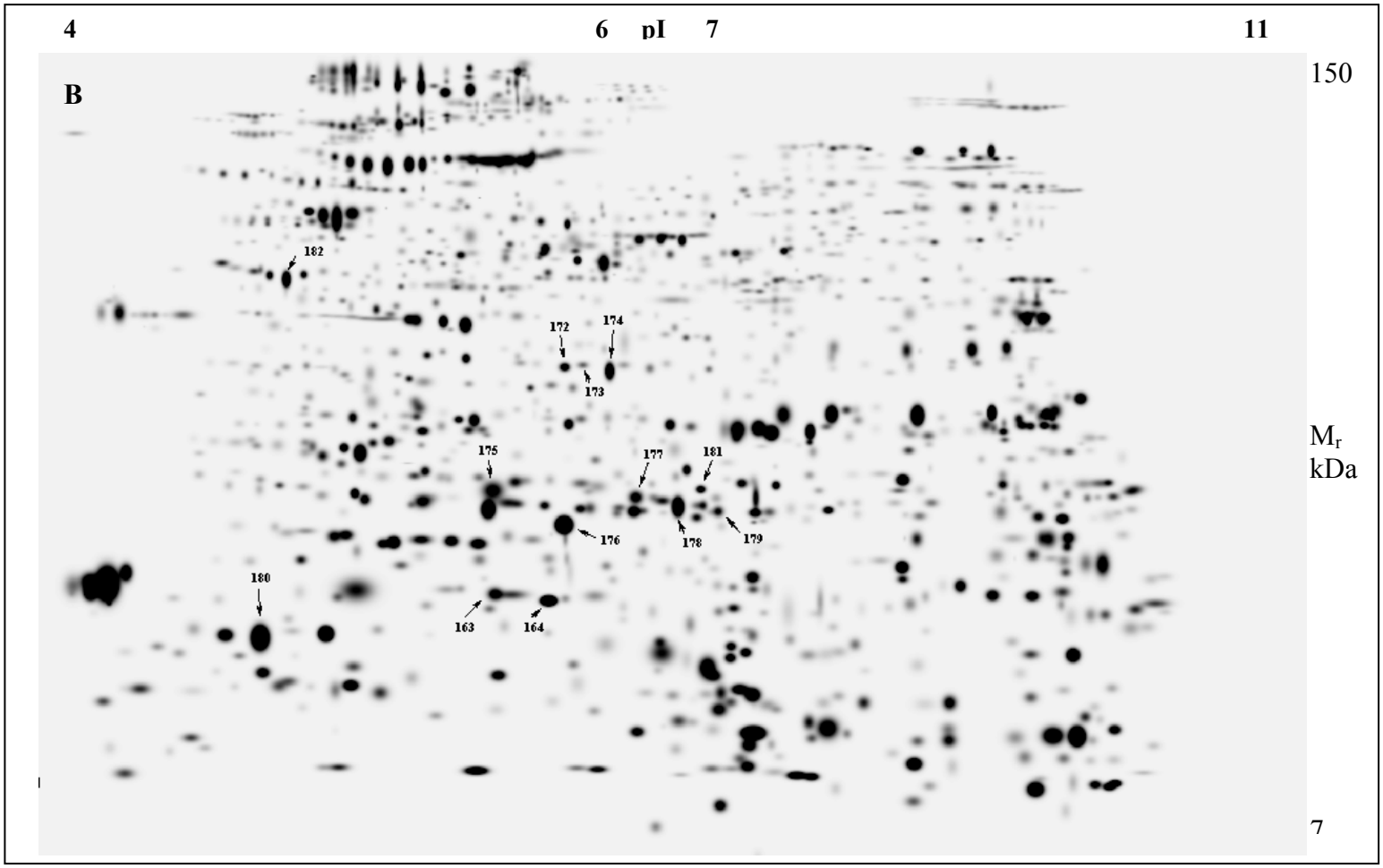
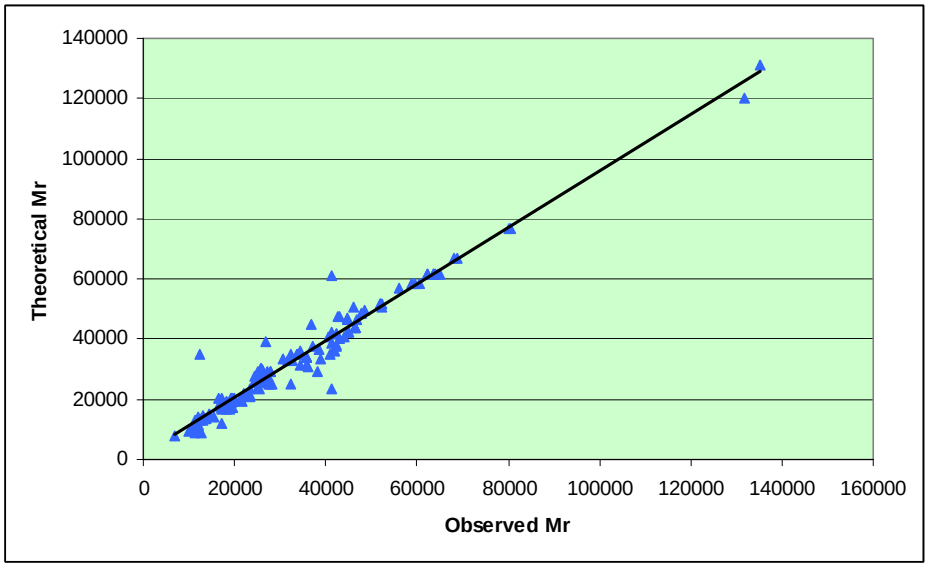


Figure 3.13B is a digitised representation of the two-dimensional gels of the GI *H. pylori* isolates that represent spots 166 – 172 within the pH 4-11/50-150 kDa range processed by PDQUEST software

14A



14B

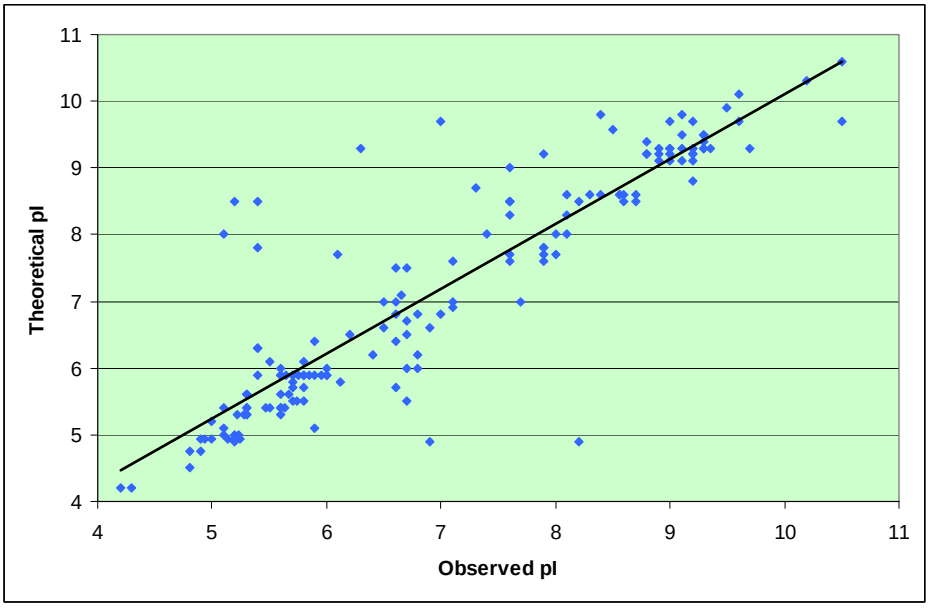


Figure 3.14 Scatter diagrams

Comparison of the observed pI and M_w of the 2-DE protein spots that were selected for identification by mass spectrometry is compared to the theoretical pI and M_w of the gene products (www.tigr.com). Observed pI and M_r of protein spots is determined by the application of 2-DE standards in conjunction with the protein sample. (a) Observed M_r versus theoretical M_w. The solid line is an equivalence plot. (b) Observed pI versus theoretical pI. Solid line, equivalence plot (slope 1.0).

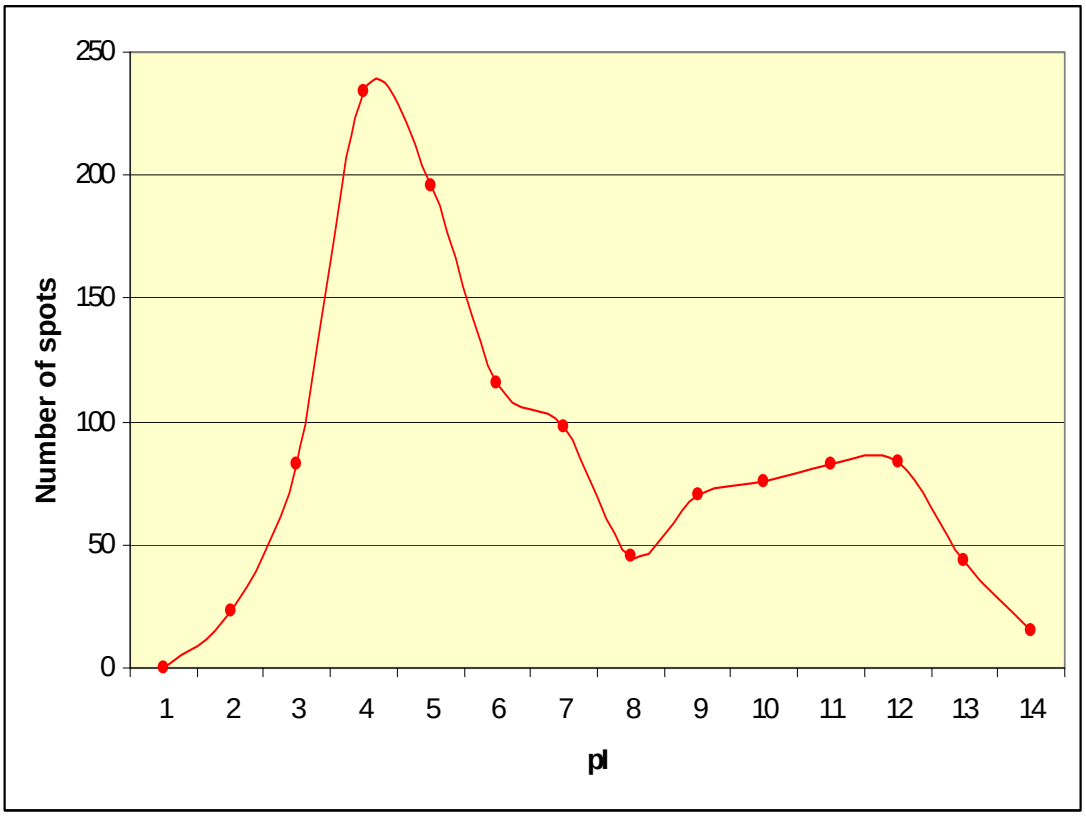


Figure 3.15 pI Distribution of *H. pylori* proteins

These figures represent the pI of the total observed proteome of *H. pylori*.

Table 3.1 Valid spots

The number of valid protein spots resolved for each isolate examined in this study (Valid spots are defined as those that are located in the same position, on three out of four gels in a replicate group).

Number of Proteins			
Strain	Tris-tricine gels		Tris-glycine gels
	CBB G250	Silver	CBB G250
GU1	1166	1586	1002
GU2	1167	1580	1010
GU3	1163	1560	1006
DU1	1161	1575	1011
DU2	1158	1567	1009
DU3	1167	1584	1014
NUD1	1152	1601	1007
NUD2	1156	1576	1012
NUD3	1156	1577	1016
GI1	1165	1568	1003
GI2	1150	NA	1001

Table 3.2 Summary of the %T%C and buffer systems

Buffer System	Separation gel		Stacking gels		Urea
	T%	*C%*	*T%*	*C%*	M
Tris-Glycine	10	2.5			0
	13	2.5			0
Tris-Tricine	9	6			3.1
	9	6			6
	8.5	6			3.1
	8.5	6			0
Tris-Tricine with Stacking gel	10.4	6	5.4	3	3.1 6

Table 3.3 Spot matches between disease groups

Spots common to two disease outcomes

	GU	NUD	DU	GI
GU group	768 (66%)	598 (53%)	596 (50%)	654 (56%)
NUD group		741 (64%)	569 (48%)	624 (54%)
DU group			757 (65%)	603 (52%)
GI group				894 (77%)

Spots common to three disease outcomes

GU and NUD and DU 502 (43%)
GU and NUD and GI 540 (47%)
DU and NUD and GI 522 (44.5%)

Spots common to all isolates 471 (40.6%)

.

Table 3.4 Spot matches between isolates

The number of protein spots that matched between any two isolates.

	GU1	GU2	GU3	NUD1	NUD2	NUD3	DU1	DU2	DU3	GI1	GI2
GU1	1137 (100%)	887 (78%)	872 (77%)	841 (74%)	849 (74%)	844 (74%)	778 (69%)	780 (69%)	782 (70%)	825 (73%)	821 (73%)
GU2		1138 (100%	872 (77%)	830 (73%)	838 (73%)	845 (74%)	767 (68%)	769 (68%)	776 (70%)	812 (72%)	803 (71%)
GU3			1127 (100%)	815 (72%)	837 (73%)	818 (72%)	775 (69%)	773 (69%)	776 (70%)	807 (72%)	800 (71%)
NUD1				1138 (100%)	884 (77%)	890 (78%)	779 (69%)	791 (70%)	787 (70%)	817 (72%)	803 (71%)
NUD2					1160 (100%)	876 (76%)	798 (70%)	810 (70%)	797 (70%)	809 (71%)	822 (72%)
NUD3						1146 (100%)	821 (72%)	803 (71%)	797 (71%)	804 (71%)	807 (71%)
DU1							1120 (100%)	882 (79%)	848 (77%)	795 (71%)	785 (70%)
DU2								1125 (100%)	855 (77%)	797 (71%)	787 (72%)
DU3									1097 (100%)	759 (73%)	821 (74%)
GI1										1120 (100%)	875 (78%)
GI2											1124 (100%)

Table 3.5 Three isolate comparisons

The proteome of each isolate was compared with the proteomes of each combination of two isolates. Proteins that were unique to each of the three isolates are shown in the table.

3-Strain Subsets	No of Proteins	3-Strain Subsets	No of Proteins
GU1/GU2/GU3	58	NUD2/NUD3/DU2	1
GU1/GU2/NUD1	1	NUD2/NUD3/DU3	0
GU1/GU2/NUD2	0	NUD2/NUD3/GI1	0
GU1/GU2/NUD3	2	NUD2/NUD3/GI2	0
GU1/GU2/DU1	3	DU1/DU2/DU3	55
GU1/GU2/DU2	2	DU1/DU2/GU1	0
GU1/GU2/DU3	2	DU1/DU2/GU2	0
GU1/GU2/GI1	0	DU1/DU2/GU3	0
GU1/GU2/GI2	1	DU1/DU2/NUD1	2
GU1/GU3/NUD1	2	DU1/DU2/NUD2	2
GU1/GU3/NUD2	1	DU1/DU2/NUD3	0
GU1/GU3/NUD3	0	DU1/DU2/GI1	2
GU1/GU3/DU1	0	DU1/DU2/GI2	2
GU1/GU3/DU2	2	DU1/DU3/GU1	0
GU1/GU3/DU3	0	DU1/DU3/GU2	2
GU1/GU3/GI1	1	DU1/DU3/GU3	0
GU1/GU3/GI2	1	DU1/DU3/NUD1	0
GU2/GU3/NUD1	1	DU1/DU3/NUD2	0
GU2/GU3/NUD2	1	DU1/DU3/NUD3	1
GU2/GU3/NUD3	0	DU1/DU3/GI1	0
GU2/GU3/DU1	1	DU1/DU3/GI2	1
GU2/GU3/DU2	1	DU2/DU3/GU1	0
GU2/GU3/DU3	1	DU2/DU3/GU2	3
GU2/GU3/GI1	1	DU2/DU3/GU3	0
GU2/GU3/GI2	0	DU2/DU3/NUD1	0
NUD1/NUD2/NUD3	54	DU2/DU3/NUD2	1
NUD1/NUD2/GU1	1	DU2/DU3/NUD3	1
NUD1/NUD2/GU2	2	DU2/DU3/GI1	0
NUD1/NUD2/GU3	1	DU2/DU3/GI2	0
NUD1/NUD2/DU1	1	GI1/GI2	55
NUD1/NUD2/DU2	0	GI1/GI2/GU1	5
NUD1/NUD2/DU3	1	GI1/GI2/GU2	5
NUD1/NUD2/GI1	1	GI1/GI2/GU3	3
NUD1/NUD2/GI2	0	GI1/GI2/NUD1	2
NUD1/NUD3/GU1	0	GI1/GI2/NUD2	0
NUD1/NUD3/GU2	2	GI1/GI2/NUD3	1
NUD1/NUD3/GU3	1	GI1/GI2/DU1	0
NUD1/NUD3/DU1	0	GI1/GI2/DU2	1
NUD1/NUD3/DU2	1	GI1/GI2/DU3	1
NUD1/NUD3/DU3	2	GU1/DU1/NUD1	1
NUD1/NUD3/GI1	0	GU1/DU1/NUD2	0
NUD1/NUD3/GI2	1	GU1/DU1/NUD3	0
NUD2/NUD3/GU1	0	GU1/DU2/NUD1	3

Table 3.5 Continued

3-Strain Subsets	No of Proteins	3-Strain Subsets	No of Proteins
NUD2/NUD3/GU2	0	GU1/DU2/NUD2	1
NUD2/NUD3/GU3	0	GU1/DU2/NUD3	1
NUD2/NUD3/DU1	0	GU1/DU3/NUD1	1
GU1/DU3/NUD2	0	GU2/DU3/GI2	0
GU1/DU3/NUD3	0	GU2/NUD1/GI1	0
GU2/DU1/NUD1	0	GU2/NUD1/GI2	0
GU2/DU1/NUD2	1	GU2/NUD2/GI1	0
GU2/DU1/NUD3	1	GU2/NUD2/GI2	1
GU2/DU2/NUD1	0	GU2/NUD3/GI1	0
GU2/DU2/NUD2	0	GU2/NUD3/GI2	0
GU2/DU2/NUD3	0	GU3/DU1/GI1	0
GU2/DU3/NUD1	0	GU3/DU1/GI2	0
GU2/DU3/NUD2	0	GU3/DU2/GI1	0
GU2/DU3/NUD3	0	GU3/DU2/GI2	0
GU3/DU1/NUD1	1	GU3/DU3/GI1	1
GU3/DU1/NUD2	0	GU3/DU3/GI2	0
GU3/DU1/NUD3	1	GU3/NUD1/GI1	1
GU3/DU2/NUD1	0	GU3/NUD1/GI2	0
GU3/DU2/NUD2	0	GU3/NUD2/GI1	0
GU3/DU2/NUD3	0	GU3/NUD2/GI2	1
GU3/DU3/NUD1	0	GU3/NUD3/GI1	0
GU3/DU3/NUD2	0	GU3/NUD3/GI2	0
GU3/DU3/NUD3	0	DU1/NUD1/GI1	0
GU1/DU1/GI1	0	DU1/NUD1/GI2	0
GU1/DU1/GI2	0	DU1/NUD2/GI1	0
GU1/DU2/GI1	0	DU1/NUD2/GI2	0
GU1/DU2/GI2	1	DU1/NUD3/GI1	0
GU1/DU3/GI1	0	DU1/NUD3/GI2	0
GU1/DU3/GI2	0	DU2/NUD1/GI1	0
GU1/NUD1/GI1	0	DU2/NUD1/GI2	1
GU1/NUD1/GI2	1	DU2/NUD2/GI1	0
GU1/NUD2/GI1	0	DU2/NUD2/GI2	0
GU1/NUD2/GI2	0	DU2/NUD3/GI1	0
GU1/NUD3/GI1	0	DU2/NUD3/GI2	0
GU1/NUD3/GI2	1	DU3/NUD1/GI1	2
GU2/DU1/GI1	0	DU3/NUD1/GI2	1
GU2/DU1/GI2	0	DU3/NUD2/GI1	0
GU2/DU2/GI1	0	DU3/NUD2/GI2	0
GU2/DU2/GI2	0	DU3/NUD3/GI1	0
GU2/DU3/GI1	0	DU3/NUD3/GI2	0

Table 3.6 Two isolate comparisons

The proteome of each isolate was compared with the proteome of each other isolate. Proteins that were unique to each pair of isolates are shown in the table.

Two-isolate Subset	Spots	Two-isolate Subset	Spots
GI1/GI2	49	GU2/GU3	6
GI1/GU1	2	GU2/NUD1	6
GI1/GU2	6	GU2/NUD2	5
GI1/GU3	1	GU2/NUD3	4
GI1/NUD1	5	GU2/DU1	1
GI1/NUD2	1	GU2/DU2	3
GI1/NUD3	3	GU2/DU3	3
GI1/DU1	2	GU3/NUD1	1
GI1/DU2	4	GU3/NUD2	5
GI1/DU3	3	GU3/NUD3	4
GI2/GU1	3	GU3/DU1	1
GI2/GU2	5	GU3/DU2	5
GI2/GU3	1	GU3/DU3	3
GI2/NUD1	3	NUD1/NUD2	3
GI2/NUD2	3	NUD1/NUD3	4
GI2/NUD3	1	NUD1/DU1	3
GI2/DU1	3	NUD1/DU2	1
GI2/DU2	3	NUD1/DU3	0
GI2/DU3	3	NUD2/NUD3	0
GU1/GU2	6	NUD2/DU1	5
GU1/GU3	4	NUD2/DU2	3
GU1/NUD1	4	NUD2/DU3	1
GU1/NUD2	2	NUD3/DU1	6
GU1/NUD3	4	NUD3/DU2	6
GU1/DU1	6	NUD3/DU3	1
GU1/DU2	1	DU1/DU2	6
GU1/DU3	3	DU1/DU3	5
		DU2/DU3	4

Table 3.7 Protein summary

Summary of the numbers of proteins that were specifically absent from, or unique to, a specific disease group.

GU Unique	58 (5.5%)
DU Unique	55 (5%)
GI Unique	54 (5%)
NUD Unique	55 (5%)

GU Absent	39 (3.5%)
DU Absent	38 (3.5%)
NUD Absent	44 (3%)
GI Absent	34 (3%)

Table 3.8 Eight and nine isolate comparisons

The proteomes of every combination of eight isolates (in the case of GI isolates) were compared for the absence of proteins.

8 and 9 Strain Subsets	No of Proteins
NUD1/NUD2/NUD3/DU1/DU2/DU3/GI1/GI2	39
GU1/NUD2/NUD3/DU1/DU2/DU3/GI1/GI2	1
GU1/NUD1/NUD2/NUD3/DU1/DU2/DU3/GI1	1
GU1/NUD1/NUD2/NUD3/DU1/DU2/DU3/GI2	1
GU1/NUD1/NUD2/NUD3/DU2/DU3/GI1/GI2	1
GU1/NUD1/NUD2/NUD3/DU1/DU3/GI1/GI2	2
GU1/NUD1/NUD2/NUD3/DU1/DU2/GI1/GI2	4
GU2/NUD1/NUD2/NUD3/DU1/DU2/DU3/GI1	2
GU2/NUD1/NUD2/NUD3/DU1/DU3/GI1/GI2	1
GU3/NUD1/NUD2/NUD3/DU1/DU2/DU3/GI2	1
GU3/NUD1/NUD2/NUD3/DU1/DU2/DU3/GI1	2
GU3/NUD2/NUD3/DU1/DU2/DU3/GI1/GI2	1
GU3/NUD1/NUD2/DU1/DU2/DU3/GI1/GI2	1
GU3/NUD1/NUD2/NUD3/DU1/DU3/GI1/GI2	1
GU3/NUD1/NUD2/NUD3/DU1/DU2/GI1/GI2	1
GU1/GU2/GU3/DU1/DU2/DU3/GI1/GI2	44
NUD1/GU1/GU2/GU3/DU1/DU2/DU3/GI1	1
NUD1/GU1/GU2/GU3/DU1/DU2/GI1/GI2	1
NUD2/GU1/GU2/GU3/DU2/DU3/GI1/GI2	1
NUD2/GU1/GU2/GU3/DU1/DU2/GI1/GI2	1
NUD3/GU1/GU2/GU3/DU1/DU2/DU3/GI1	3
NUD3/GU1/GU3/DU1/DU2/DU3/GI1/GI2	1
NUD3/GU1/GU2/GU3/DU2/DU3/GI1/GI2	2
GU1/GU2/GU3/NUD1/NUD2/NUD3/GI1/GI2	38
DU1/GU2/GU3/NUD1/NUD2/NUD3/GI1/GI2	1
DU1/GU1/GU2/NUD1/NUD2/NUD3/GI1/GI2	1
DU1/GU1/GU2/GU3/NUD1/NUD2/NUD3/GI1	2
DU1/GU1/GU2/GU3/NUD1/NUD2/GI1/GI2	1
DU2/GU1/GU3/NUD1/NUD2/NUD3/GI1/GI2	2
DU2/GU1/GU2/NUD1/NUD2/NUD3/GI1/GI2	2
DU2/GU1/GU2/GU3/NUD1/NUD2/NUD3/GI1	2
DU2/GU1/GU2/GU3/NUD1/NUD2/NUD3/GI2	1
DU2/GU1/GU2/GU3/NUD2/NUD3/GI1/GI2	1
DU2/GU1/GU2/GU3/NUD1/NUD3/GI1/GI2	3
DU2/GU1/GU2/GU3/NUD1/NUD2/GI1/GI2	2
DU3/GU1/GU3/NUD1/NUD2/NUD3/GI1/GI2	1
DU3/GU1/GU2/NUD1/NUD2/NUD3/GI1/GI2	1
DU3/GU1/GU2/GU3/NUD1/NUD3/GI1/GI2	1
DU3/GU1/GU2/GU3/NUD1/NUD2/GI1/GI2	1
GU1/GU2/GU3/NUD1/NUD2/NUD3/DU1/DU2/DU3	34

Table 3.9 Absent spots for two isolate subsets

Specifically absent proteins that can be categorised into two-isolate subsets

Isolates Absent	Spots Missing	Isolates Absent	Spots Missing
GI1/GI2	11	GU2/GU3	3
GI1/GU1	2	GU2/NUD1	2
GI1/GU2	1	GU2/NUD2	2
GI1/GU3	4	GU2/NUD3	1
GI1/NUD1	2	GU2/DU1	1
GI1/NUD2	0	GU2/DU2	0
GI1/NUD3	3	GU2/DU3	0
GI1/DU1	3	GU3/NUD1	1
GI1/DU2	2	GU3/NUD2	2
GI1/DU3	0	GU3/NUD3	0
GI2/GU1	2	GU3/DU1	1
GI2/GU2	1	GU3/DU2	0
GI2/GU3	0	GU3/DU3	0
GI2/NUD1	4	NUD1/NUD2	4
GI2/NUD2	3	NUD1/NUD3	4
GI2/NUD3	0	NUD1/DU1	2
GI2/DU1	2	NUD1/DU2	0
GI2/DU2	0	NUD1/DU3	4
GI2/DU3	2	NUD2/NUD3	1
GU1/GU2	3	NUD2/DU1	3
GU1/GU3	3	NUD2/DU2	0
GU1/NUD1	2	NUD2/DU3	1
GU1/NUD2	1	NUD3/DU1	4
GU1/NUD3	1	NUD3/DU2	0
GU1/DU1	0	NUD3/DU3	0
GU1/DU2	0	DU1/DU2	4
GU1/DU3	0	DU1/DU3	3
		DU2/DU3	4

Table 3.10 Strain cluster membership

Displays the qualitative and quantitative cluster membership of each isolate. Clusters refers to each subgrouping of the isolates.

Examination of two clusters for both quantitative and qualitative revealed that isolates belonged to two groups 1= gastric and 2= duodenal. At four clusters qualitative dendrograms isolates belonged to four different groups 1= GU, 2= GI, 3= NUD and 4= DU. For the quantitative the only isolates that belonged to the same group at the four cluster level were the DU isolates.

	Quantitative Membership			Qualitative Membership		
Case	2 Clusters	3 Clusters	4 Clusters	2 Clusters	3 Clusters	4 Clusters
GU1	1	1	1	1	1	1
GU2	1	1	1	1	1	1
GU3	1	2	2	1	1	1
GI1	1	1	3	1	2	2
GI2	1	1	1	1	2	2
NUD1	1	1	3	1	1	3
NUD2	1	1	3	1	1	3
NUD3	1	2	2	1	1	3
DU1	2	3	4	2	3	4
DU2	2	3	4	2	3	4
DU3	2	3	4	2	3	4

Table 3.11 List of Tables and Figures

A list of the tables and figures that correspond to positional and quantitatively conserved spots and spot differences.

Spot Number	Figure	Table	Group
Positional and Quantitatively Conserved		3.12	
1-22	3.6		All pH 4-7
23-96	3.7		All pH 4-11
Unique spots		3.13	
97-102	3.8		GU
103-109	3.9		DU
110-116	3.11		NUD
117-119	3.10		GI
Absent spots		3.14	
120-126	3.9		GU
127-130	3.8		DU
131-134	3.10		NUD
135-138	3.11		GI
Differentially regulated		3.15	
Down	3.12		
139, 152,161,162			GU
140-144, 146-148			DU
150, 151,153, 154, 156, 159, 145,146,			NUD
149-154, 156, 155,			GI
Up	3.12		
154, 157, 158			GU
160			DU

Table 3.12 Positional and quantitatively conserved spots

Spots number 1-22 excised from *H. pylori* within the pH 4-7 range and between 46-150 kDa (Figure 3.6). Spots numbered 23-96 excised from *H. pylori* within the pH 4-11 range and between 7-46 kDa (Figure 3.7)

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r 2-DE	pI Theoretical	pI 2-DE
1	HP0109	Chaperone and heat shock protein 70 †	67051	68700	4.76	4.8
2	HP0109	Chaperone and heat shock protein 70 †	67051	68000	4.76	4.9
3	HP1195	Translation elongation factor EF-G †	77020	80500	5	5.1
4	HP1195	Translation elongation factor EF-G †	77020	80100	5	5.2
5	HP0072	Urease β-subunit ‡	61683	65100	5.9	5.6
6	HP0072	Urease β-subunit ‡	61683	64680	5.9	5.6
7	HP0072	Urease β-subunit ‡	61683	64000	5.9	5.65
8	HP0072	Urease β-subunit ‡	61683	64000	5.9	5.7
9	HP0072	Urease β-subunit ‡	61683	63800	5.9	5.7
10	HP0072	Urease β-subunit ‡	61683	62200	5.9	5.75
11	HP0072	Urease β-subunit ‡	61683	63800	5.9	5.8
12	HP0072	Urease β-subunit	61683	62200	5.9	5.85
13	HP0072	Urease β-subunit	61683	63800	5.9	5.9
14	HP0072	Urease β-subunit	61683	63800	5.9	5.95
15	HP0072	Urease β-subunit	61683	63800	5.9	6
16	HP0072	Urease β-subunit	61683	62200	5.9	6
17	HP1205	Translation elongation factor EF-Tu ‡	43647	46610	4.93	4.9
18	HP1205	Translation elongation factor EF-Tu ‡	43647	46610	4.93	4.93
19	HP1205	Translation elongation factor EF-Tu ‡	43647	46610	4.93	5
20	HP1205	Translation elongation factor EF-Tu ‡	43647	46610	4.93	5.14
21	HP1205	Translation elongation factor EF-Tu ‡	43647	46510	4.93	5.2
22	HP1205	Translation elongation factor EF-Tu ‡	43647	46510	4.93	5.24
23	HP1161	Flavodoxin	17492	17100	4.2	4.2
24	HP1161	Flavodoxin	17492	17100	4.2	4.3
25	JHP0381	Putative transcriptional regulator	25500	27000	5.4	5.1
26	HP0011	10 kDa protein	14500	12990	6.6	6.5
27	H0614	Hypothetical protein	12983	12990	7	6.5

†Lock et al (2001), ‡ identified by current study and Lock et al (2001)

Table 3.12 Continued

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r 2-DE	pI Theoretical	pI 2-DE
28	HP0390	Adhesin-thiol peroxidase	18354	18200	5.5	6.7
29	HP1067	Chemotaxis protein	13915	12107	5	5.23
30	HP0618	Adenylate kinase ‡	21243	23384	5.3	5.22
31	HP0618	Adenylate kinase	21409	23161	5.9	5.8
32	HP0824	Thioredoxin	11584	11500	4.9	8.2
33	HP1380	Prephenate dehydrogenase	29429	27100	7	7.1
34	HP0073	Urease α-subunit ‡	26554	26600	8.5	8.2
35	HP0073	Urease α-subunit ‡	26540	26600	8.5	8.7
36	HP0073	Urease α-subunit ‡	26540	26600	8.5	7.6
37	HP0073	Urease α-subunit ‡	26540	26600	8.5	7.6
38	HP1564	Outer membrane protein	30151	25700	9.1	8.9
39	HP1564	Outer membrane protein	30151	25700	9.1	9
40	HP0231	Hypothetical protein	29490	27800	9.1	9.1
41	HP0231	Hypothetical protein	29490	27800	9.1	9.2
42	HP0175	Cell binding factor 2	34031	35900	9.3	8.9
43	HP0175	Cell binding factor 2	34031	35850	9.3	9
44	HP0175	Cell binding factor 2	34031	35800	9.3	9.2
45	HP0175	Cell binding factor 2	34031	35800	9.3	9.3
46	HP1201	Ribosomal protein L1	25266	28100	10.3	10.2
47	HP0356	Hypothetical protein	27793	24300	9.5	9.1
48	HP0955	Integrase/recombinase	42086	41200	9.7	9
49	HP1118	γ- glutamyltranspeptidase* ‡	61099	41200	9.3	9
50	HP1576	ABC transporter, ATP-binding protein	23484	41200	9.7	9.2
51	JHP1177	Ribosome releasing factor	20917	23300	8.6	8.1
52	JHP0405	Protease Do	51714	52000	9.2	8.8
53	JHP0405	Protease Do	51714	52000	9.2	8.9
54	HP1286	Conserved hypothetical secreted protein	20603	19400	9.2	8.8
55	HP1286	Conserved hypothetical secreted protein	20615	19400	9.2	9
56	JHP1076	Hypothetical protein	20395	19100	6.9	7.1
57	HP0721	Hypothetical protein	17520	16502	9.57	8.5
58	HP1285	Conserved hypothetical secreted protein	26298	24500	9.3	9.35
59	HP1203	Transcription termination factor	20261	17100	7	7.7
60	HP0829	Inosine-5'-monophosphate dehydrogenase	51802	52300	7.7	7.6

***γ-** glutamyltranspeptidase – Gamma-glutamyltranspeptidase
† Lock et al (2001), ‡ identified by current study and Lock et al (2001)

Table 3.12 Continued

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r 2-DE	pI Theoretical	pI 2-DE
61	HP1111	Pyruvate ferredoxin oxidoreductase, β-subunit	34956	33700	8.3	7.6
62	HP1111	Pyruvate ferredoxin oxidoreductase, β-subunit	34956	33700	8.3	8.1
63	HP1302	Ribosomal protein S5	16538	18100	10.6	10.5
64	HP1320	Ribosomal protein S10	11910	17300	9.9	9.5
65	HP0110	Co-chaperone and heat shock protein	22040	22000	5.3	5.3
66	HP1311	Ribosomal protein L29	7683	6900	9.7	10.5
67	HP0198	Nucleoside diphosphate kinase	15286	14607	8	7.4
68	JHP0086	Alpha-1,2-fucosyltransferase	35159	40800	6.7	6.7
69	JHP0537	Subunit of 2-oxoglutarate oxidoreductase	41573	42300	6	6.7
70	JHP0537	Subunit of 2-oxoglutarate oxidoreductase	41573	42300	6	6.8
71	JHP0178	Fumerate reductase‡	27625	25700	5.3	5.3
72	JHP1349	Membrane-associated lipoprotein	19122	18312	9.5	9.3
73	HP0073	Urease a-subunit ‡	26540	26600	8.5	7.6
74	JHP0939	Hypothetical protein	39251	26951	8.7	7.3
75	HP0827	ss-DNA binding protein 12RNP2 precursor	9385	10142	9.5	9.3
76	JHP0714	Uridylate kinase†	26172	27152	7.7	7.9
77	HP0561	3-ketoacyl-acyl carrier protein reductase †	26668	26039	7.8	7.9
78	HP0194	Triosephosphate isomerase †	26707	25016	7.8	7.9
79	HP0371	Biotin carboxyl carrier protein	17132	19498	5.4	5.3
80	JHP1489	Pyridoxal phosphate synthetase †	29720	26164	8	8.1
81	HP0900	Hydrogenase expression/ formation protein †	27310	27687	5.4	5.63
82	HP0026	Citrate synthase ‡	48350	47670	7.7	8
83	HP0742	Phosphoribosylpyrophosphate synthetase	34824	12445	9.8	8.4
84	JHP1024	JHP1024 like protein	13669	13845	7.5	6.6
85	JHP1024	JHP1024 like protein	13669	13360	7.5	6.7
86	HP0835	histone-like DNA-binding protein HU	10383	11760	9.7	9.6
87	HP1173	Hypothetical protein	20585	16492	9.2	9.2
88	HP1390	Hypothetical protein	19097	21629	4.5	4.8
89	HP0913	Outer membrane protein	57063	56236	9.2	7.9
90	HP0500	DNA polymerase III β-subunit	42185	44618	5.5	5.7

† Lock et al (2001), ‡ identified by current study and Lock et al (2001)

Table 3.12 Continued

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r 2-DE	pI Theoretical	pI 2-DE
91	HP1390	Hypothetical	13553	11500	5.3	5
92	HP0537	Hypothetical	13012	12333	9.7	7
93	JHP0050	Cag 18	26840	27100	5.7	6.6
94	HP1024	co-chaperone-curved DNA binding protein A	32908	32811	8.5	8.6
95	HP0602	Endonuclease III	25287	27747	6.2	6.8
96	HP0811	Hypothetical	12842	11300	6.6	6.9

† Lock et al (2001), ‡ identified by current study and Lock et al (2001)

Table 3.13 Specifically unique spots

Abundant protein spots that are specifically unique for a particular disease group.

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r Observed	pI Theoretical	pI Observed
GU Unique Spots						
97	JHP0421	Hypothetical protein	18653	19286	5.6	5.67
98	NA	Alkyl hydroperoxide reductase C22 protein	20266	19608	5.8	6.12
99	orf50	Unknown	46666	46944	8.6	8.3
100	HP0418	Hypothetical protein	38766	41258	7.1	6.65
101	JHP0516	Hypothetical protein	40604	44027	5.5	5.74
102	NA	Conserved hypothetical protein	24227	25259	9.8	9.1
DU Unique Spots						
103	HP0377	Hypothetical protein	25315	32345	5.7	5.7
104	JHP0376	Hypothetical protein	10316	10788	10.1	9.6
105	HP0332	Cell division topological specificity factor	8920	12584	8	8
106	HP1437	Hypothetical protein	27735	26841	6.5	6.7
107	JHP0051	Hypothetical protein	14130	13085	6.8	7
108	JHP0140	Hypothetical protein	32859	31592	7.7	6.1
109	HP0036	Hypothetical protein	38533	34584	10.2	8.7
NUD Unique Spots						
110	NA	Hsp12 variant C	36081	34456	6	5.6
111	HP1064	Hypothetical protein	10872	11898	9.4	8.8
112	JHP1163	Putative protein	9125	11304	6.2	6.4
113	HP0864	Hypothetical protein	25402	24812	8.8	9.2
114	JHP1044	Hypothetical protein	131029	135212	8.5	5.2
115	JHP1073	Hypothetical protein	33415	30722	7.6	7.1
116	HP0484	Hypothetical protein	29378	25177	6.5	6.2
GI Unique Spots						
117	NA	Alkyl hydroperoxide C22 protein	20266	19908	5.8	5.7
118	HP0538	Cag17 cagN; JHP0486	35275	34456	8.6	8.56
119	HP0842	Hypothetical protein	17755	18243	9.4	9.3

NA – Not available

Table 3.14 Specifically absent spots

Abundant protein spots that are specifically unique for a particular disease group.

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r 2-DE	pI Theoretical	pI 2-DE
GU Absent Spots						
120	HP0958	Putative protein	31126	34456	5.4	5.47
121	HP0825	Thioredoxin reductase	34986	32325	6.4	5.9
122	JHP0954	Hypothetical protein	16519	19056	5.6	5.3
123	HP1499	Hypothetical protein	30855	36272	8.5	5.4
124	JHP0955	hypothetical protein	25315	27564	5.7	5.8
125	JHP0788	Hypothetical protein	36036	42137	6.1	5.8
126	JHP0118	Hypothetical protein	16612	17300	9.3	9.1
DU Absent Spots						
127	HP0944	Conserved hypothetical protein	13374	12604	5.3	5.28
128	JHP1350	Hypothetical protein	23306	23282	7.8	5.4
129	JHP0295	Hypothetical protein	33643	38714	5.4	5.3
130	HP1078	Hypothetical protein	27122	25304	6.3	5.4
NUD Absent Spots						
131	HP1562	Iron(III) ABC transporter, periplasmic iron-binding protein	37552	36997	9.2	9.2
132	HP0305	Hypothetical protein	20385	21485	9.3	9.2
133	JHP0671	Hypothetical protein	49746	48345	6.8	6.6
134	HP1285	Conserved hypothetical secreted protein	26298	25517	9.3	9.7
GI Absent Spots						
135	JHP0934	Hypothetical protein	40138	42939	5.1	5.9
136	HP0431	Protein phosphatase 2C homolog	25964	25391	9.3	9.3
137	JHP0896	Hypothetical protein	29007	38251	5.1	5.1
138	HP0434	Hypothetical protein	45150	36800	9	7.6

Table 3.15 Differentially expressed proteins

NS = no significant difference D = spot intensity down by a factor of three, U = spot intensity up by a factor of three

Spot No	Spot Homologs Locus	Protein Identity	M_r Theoretical	M_r Observed	pI Theoretical	pI Observed	GI Group	GU Group	NUD Group	DU Group
139	HP0958	Putative protein	31126	35765	5.4	5.6	NS	D	NS	NS
140	HP0875	KatA catalase	58637	60700	8.6	8.7	NS	NS	NS	D
141	HP0875	KatA catalase	58637	59579	8.6	8.6	NS	NS	NS	D
142	HP0875	KatA catalase	58637	58979	8.6	8.4	NS	NS	NS	D
143	HP0197	S-adenosylmethionine synthetase 2	42363	45211	6	6	NS	NS	NS	D
144	HP0201	Fatty acid/phospholipid synthesis protein	36484	38450	6.1	5.5	NS	NS	NS	D
145	HP0653	Non-haeme iron containing ferritin	19346	19511	5.4	5.6	NS	NS	D	NS
146	HP0653	Non-haeme iron containing ferritin	19346	19584	5.4	5.5	NS	NS	D	D
147	HP0027	Isocitrate dehydrogenase	47532	42500	7.6	7.9	NS	NS	NS	D
148	HP0027	Isocitrate dehydrogenase	47532	42977	7.6	7.6	NS	NS	NS	D
149	HP0879	Hypothetical protein	23496	25432	5.6	5.3	NS	NS	D	NS
150	JHP0650	Hypothetical protein	47269	44860	6.4	6.6	NS	NS	D	D
151	JHP0761	Homoserine dehydrogenase	46427	44742	6.8	6.8	NS	NS	D	D
152	HP1520	Hypothetical protein	50574	52349	5.9	5.4	NS	D	D	NS
153	HP0453	Hypothetical protein	120115	131734	8	5.1	NS	NS	D	D
154	HP0614	Hypothetical protein	12983	12263	7	6.6	NS	U	D	D
155	HP1143	H. pylori predicted coding region	50645	46126	5.3	5.6	NS	U	U	U
156	JHP0052	Hypothetical protein	38221	42313	5.5	5.8	NS	NS	D	D
157	HP0824	Thioredoxin	11584	11900	4.9	5.2	NS	U	NS	NS
158	HP0824	Thioredoxin	11584	12000	4.9	6.9	NS	U	NS	NS
159	JHP0907	Putative protein	40755	40633	9.3	6.3	NS	NS	NS	D
160	JHP0154	Hypothetical protein	10390	12140	6.3	5.4	NS	NS	NS	U
161	HP0513	Recombinase RecA	37687	42425	5.6	5.6	NS	D	NS	NS
162	JHP0065	Hypothetical protein	14061	15551	5.2	5	NS	D	NS	NS

Table 3.16 Positional 2-DE protein differences for 11 *H. pylori* isolates

Symbols"←", "→", and "=" show positional shifts to the left, right or approximate similarity versus the reference isolate.
Note: The reference isolate for the comparison of protein positional shifts was GI1.

Spot No	Spot Homologs Locus	Protein Identity	M_r		pI		GI1	GI2	GU1	GU2	GU3	NUD1	NUD2	NUD3	DU1	DU2	DU3
			Theoretical	Observed	Theoretical	Observed											
163	HP0243	Neutrophil activating protein	16900	17044	5.6	5.8	=	←	=	=	→	=	=	=	=	=	=
164	HP0243	Neutrophil activating protein	16900	16639	5.6	5.6	=	←	=	=	→	=	=	=	=	=	=
165	HP0010	Chaperonin (groEL)	58200	60296	5.2	5.2	=	=	→	=	=	=	=	=	=	=	=
166	HP0010	Chaperonin (groEL)	58200	58319	5.2	5.3	=	=	→	=	=	=	=	=	=	=	=
167	HP0010	Chaperonin (groEL)	58200	58002	5.2	5.4	=	=	→	=	=	=	=	=	=	=	=
168	HP0010	Chaperonin (groEL)	58200	57000	5.2	5.45	=	=	→	=	=	=	=	=	=	=	=
169	HP0010	Chaperonin (groEL)	58200	56500	5.2	5.5	=	=	→	=	=	=	=	=	=	=	=
170	HP0010	Chaperonin (groEL)	58200	57122	5.2	5.6	=	=	→	=	=	=	=	=	=	=	=
171	HP0010	Chaperonin (groEL)	58200	57981	5.2	5.6	=	=	→	=	=	=	=	=	=	=	=
172	HP0176	Fructose bisphosphate aldolase	33800	35200	5.9	5.9	=	=	=	=	=	=	=	←	←	←	→
173	HP0176	Fructose bisphosphate aldolase	33800	34563	5.9	6.05	=	=	=	=	=	=	=	←	←	←	→
174	HP0176	Fructose bisphosphate aldolase	33800	34703	5.9	6.2	=	=	=	=	=	=	=	←	←	←	→
175	HP1563	26 kDa protein	22200	21885	5.6	5.4	=	→	=	=	=	→	→	→	→	→	=
176	HP1563	26 kDa protein	22200	21391	5.6	5.6	=	→	=	=	=	→	→	→	→	→	=
177	HP1563	26 kDa protein	22200	20730	5.6	5.8	=	→	=	=	=	→	→	→	→	→	=
178	HP1563	26 kDa protein	22200	21454	5.6	6.1	=	→	=	=	=	→	→	→	→	→	=
179	HP1563	26 kDa protein	22200	21676	5.6	6.3	=	→	=	=	=	→	→	→	→	→	=
180	HP1199	Ribosomal protein L7/L12	13746	12950	4.9	5.0	=	→	=	=	=	=	=	=	→	=	=
181	HP0389	Superoxide dismutase	24500	23196	6	6.5	=	→	←	=	=	=	←	=	=	→	→
182	HP1293	DNA-directed RNA polymerase α- subunit	39400	40920	5.0	5.05	=	←	←	←	=	←	=	←	←	←	←

Table 3.17 Potential post-translational modifications

Spot No	Protein Identity	Possible Post-translational modification
99	orf5O	Methylation, Acetylation, Trimethylation, Dimethylation, Bromination, Glucosylation, Deamidation
100	Hypothetical HP0418	Trimethylation, Dimethylation, Bromination, Glucosylation, Acetylation, Deamidation
101	Putative protein JHP0516	Phosphorylation, acetylation, bromination, deamidation, glucosylation, dimethylation
104	Putative protein JHP0376	Phosphorylation, palmitoylation, glutathionylation
105	Cell division topological specificity factor	Glucosylation, methylation
106	Hypothetical protein HP1437	Dimethylation, methylation, acetylation, pyridoxal phosphate
107	Putative protein JHP0051	Methylation, dimethylation,
108	Putative protein JHP0140	Methylation, dimethylation, trimethylation, acetylation, FAD
109,118	Cag17	Methylation, dimethylation, trimethylation, acetylation, deamidation, Glucosylation, palmitoylation
111	Hypothetical protein HP1064	Methylation, trimethylation, bromination, palmitoylation, FAD
112	Putative protein JHP1163	Methylation, dimethylation
113	Hypothetical protein HP0864	Dimethylation, trimethylation, methylation, bromination, glucosylation, deamidation, acetylation,
114	Putative protein JHP1044	Methylation, acetylation, trimethylation, dimethylation, bromination, glucosylation, palmitoylation, deamidation, pyridoxal phosphate, phosphorylation
115	Putative protein JHP1073	Dimethylation, trimethylation, glucosylation, palmitoylation, phosphorylation
116	Hypothetical protein HP0484	Acetylation, alkylation, methylation
120	Putative protein JHP0892	Methylation
121	Thioredoxin reductase	Methylation, palmitoylation, acetylation, glucosylation, phosphorylation
122	Hypothetical protein JHP0954	Trimethylation
123	Hypothetical protein HP1499	Acetylation, palmitoylation, deamidation,
124	Hypothetical protein JHP0955	Methylation, acetylation, glucosylation
125	Hypothetical protein JHP0788	Methylation, trimethylation, acetylation, glucosylation
126	Hypothetical protein JHP0118	Methylation, dimethylation, trimethylation, acetylation, glucosylation
127	Conserved hypothetical protein HP0944	Phosphorylation
128	Hypothetical protein JHP1350	Deamidation
129	Hypothetical protein JHP0295	Trimethylation, phosphorylation
130	Hypothetical protein HP1078	Trimethylation, dimethylation, methylation, acetylation, bromination,
131	Iron(III) ABC transporter, periplasmic iron-binding protein	Methylation, dimethylation, acetylation, palmitoylation

Table 3.17 Continued

Spot No	Protein Identity	Possible Post-translational modification
132	Hypothetical protein HP0305	Trimethylation, deamidation, glucosylation
133	Hypothetical protein JHP0671	Methylation, trimethylation, dimethylation, acetylation, glucosylation, palmitoylation
134	Conserved hypothetical secreted protein HP1285	Phosphorylation, bromination, dimethylation, acetylation, deamidation, methylation
135	Hypothetical protein JHP0934	Glucosylation, methylation, acetylation, dimethylation, deamidation
136	Protein phosphatase 2C homolog HP0431	Phosphorylation, acetylation, methylation, dimethylation, trimethylation
137	Hypothetical protein JHP0896	Methylation, dimethylation, trimethylation, glucosylation, alkylation,
138	Hypothetical protein HP0434	Methylation, dimethylation, trimethylation, acetylation, glucosylation, phosphorylation
140, 141, 142	KatA catalase	Methylation, dimethylation, trimethylation, glucosylation, acetylation, phosphorylation
143	S-adenosylmethionine synthetase 2	Palmitoylation, trimethylation, methylation, deamidation, acetylation
144	Fatty acid/phospholipid synthesis protein	Acetylation, methylation, dimethylation, trimethylation, glucosylation, phosphorylation
145, 146	Non-heme iron containing ferritin	Methylation
147, 148	Isocitrate dehydrogenase	Methylation, deamidation, phosphorylation
149	Hypothetical protein HP0879	Methylation, trimethylation, acetylation, glucosylation, palmitoylation
150	Hypothetical protein JHP0650	Methylation, trimethylation, glucosylation, palmitoylation
151	Homoserine dehydrogenase	Methylation, dimethylation, trimethylation, bromination, phosphorylation, palmitoylation
153	Hypothetical protein HP0453	Methylation, dimethylation, trimethylation, bromination, glucosylation, phosphorylation, palmitoylation, alkylation, acetylation
152	Hypothetical protein HP1520	Methylation, dimethylation, trimethylation, phosphorylation, acetylation, palmitoylation, bromination
162	Hypothetical protein HP0065	Alkylation, methylation
154	Hypothetical protein HP0614	Methylation, dimethylation, glucosylation
155	H. pylori predicted coding region HP1143	Methylation, dimethylation, acetylation, glutathionylation
156	Hypothetical protein JHP0052	Methylation, dimethylation, trimethylation, phosphorylation, deamidation, acetylation, glucosylation
159	Putative protein JHP0907	Methylation, trimethylation, phosphorylation
161	Recombinase RecA	Methylation, dimethylation, phosphorylation

4.1 Proteome Analysis of Microorganisms

The complete genome of at least 131 microorganisms has been sequenced, including those of at least 40 pathogenic bacteria. The gene sequence data is available at http://www.tigr.org/tdb/mdb/mdbcomplete.html. Comprehensive studies at the proteome level (i.e. with more than 200 proteins identified) have been published for *Mycobacterium tuberculosis* (Jungblut et al, 1999; He et al, 2003), *H. pylori* (Jungblut et al, 2000; Lock et al, 2001), *Haemophilus influenzae* (Langen et al, 2000), *Mycoplasma pneumoniae* (Regula et al, 2000), *Pseudomonas aeruginosa* (Nouwens et al, 2002), *Staphylococcus aureus* (Cordwell et al, 2002) *and Escherichia coli* (Tonella et al, 2001). The majority of comparative proteome studies between strains within these species have included only two strains. *H. pylori* has been the only bacterial species to date (December 2005) in which proteomic variability has been compared between more than ten isolates (Enroth et al 2000).

4.2 Comparisons between the *H. pylori* studies

4.2.1 Isolates included in the study

The Jungblut (Jungblut et al, 2000) study compared only one strain from each disease group, which would have reduced the number of proteins that could be matched between these isolates. This is because, if the proteome variability between isolates was greater than that seen between groups then it would be expected that a study which compared only one strain from each group would find fewer matches between the proteomes. The Jungblut (Jungblut et al, 2000) study was preliminary and only gave a possible indication of the degree of proteomic diversity between isolates; a high degree of proteomic diversity has since been further confirmed by both Enroth's group (Enroth et al, 2000) and the present study. The Jungblut study did not explore any similarities within groups and/or identify possible disease markers.

Enroth (Enroth et al, 2000) compared three disease groups, two of which are also included in the present study (DU and GI). In the current study, with the exception of two isolates (GI1 and GI2), all isolates are from recent biopsies (see Table 2.1

page 69). The Enroth study stated that the isolates were obtained from "five patients with gastric adenocarcinoma (Ca), seven with duodenal ulcers (Du), and four with gastritis (GI), all endoscopically and histologically verified" but did not mention the year of each biopsy. Therefore, it is unknown whether all the samples in the Enroth study were from recent biopsies or whether they had been maintained in the laboratory for a considerable time.

Isolates maintained in the laboratory may evolve to adapt to their new environment during repeated laboratory subculturing. This may, for instance, effect the expression of virulence factors. In the present study only GI1 and GI2 had been maintained in the laboratory for any length of time.

4.2.2 Spot resolution

Enroth's (Enroth et al, 2000) study had the smallest gel area (11 x 11 cm = 121 cm^2) in comparison to Jungblut ((Jungblut et al, 2000) (23 x 30 cm = 690 cm^2) and the current study (18.5 x 18.5 cm = 342.5 cm^2). The small resolution area of Enroth's study would most likely have resulted in many protein spots comigrating. Such overlapping of spots would have significantly affected the number of spots visualised, detection of low abundance proteins, spot matching, comparative analysis, and any subsequent cluster analysis. For example, Enroth (Enroth et al, 2000) stated that only 48% of matches could be obtained between two isolates within a disease group. In comparison to the Enroth (Enroth et al, 2000) study the application of large format gels in the current study improves the spatial resolution and hence increases the number of protein spots visualised. Therefore, the application of large format gels increases the number of possible matches between isolates within any particular disease groups.

Jungblut's (Jungblut et al, 2000) group had the greatest resolution area of 23 x 30 cm. Their study increased the length of the M_r axis spatial resolution by 11.5 cm in comparison to the gels utilised in the current study. This increase would enhance detection of any differences in the M_r of proteins. Proteins that appear to have a similar M_r in a gel with an 18.5 cm length might now appear as two individual spots in a larger gel. However, the current study improved the pI resolution through the

use of two pH ranges on 18 cm IPG strips (pH 4-7 and 6-11 with an overlapping region of 3.6 cms =1 pH unit) giving an effective pI separation length of 32.4 cm. Jungblut (Jungblut et al, 2000) utilised a single 24 cm IPG strip of pH 3-10. While Jungblut may have had better y-axis (M_r) resolution of proteins, the current study has the advantage of improved x-axis (pI) resolution. This means that spots that may have co-migrated on the x-axis in Jungblut's (Jungblut et al, 2000) study might appear as separate protein spots in the current study.

The current study improved on the previous two studies in two ways: 1) on Jungblut's resolution by increasing the pI separation length; and 2) on Enroth's resolution by increasing the size of the gels in both dimensions. In addition, a pH range (pH 10-11) was examined which had previously not been utilised in a *H. pylori* comparative study.

4.2.3 Staining technique

Gels may be stained by a variety of dyes. Both Jungblut (Jungblut et al, 2000) and Enroth (Enroth et al, 2000) used silver nitrate staining techniques for the detection of proteins whereas the present study utilised Coomassie G250 colloidal dye for 2-DE comparative proteome analysis. Silver staining techniques were not utilised for protein identification by Jungblut and Enroth studies, instead the alternative Coomassie was used. The staining procedures utilised for protein identification for the Jungblut study was Coomassie Blue G-250. It could not be determined from Enroth's article which type of Coomassie was utilised.

While silver stains are more sensitive than Coomassie stains it is difficult to obtain gel-to-gel reproducibility and it is also difficult to perform quantitative comparisons. Coomassie G250 does not have these problems, the current study is the first *H. pylori* comparative study that effectively examines the proteomic differences both quantitatively and qualitatively.

4.3 The *H. pylori* proteome

Eleven bacterial isolates of *H. pylori* were chosen for this study, comprising nine exclusive to this study and two that have previously been examined (GI1 and GI2). This comparison was undertaken primarily to detect protein markers whose presence might be correlated specifically with the development of one of four clinical consequences of *H. pylori* infection: DU, GU, NUD and GI. Comparative 2-D proteome analysis revealed that the *H. pylori* proteome was very diverse with only 471 spots (40.6%), out of ~1168 per isolate, located at the same 2-D position for all isolates (Table 3.2). In contrast, the comparative analyses of the 2-DE patterns of strains within each of the species of *S. aureus* (Cordwell et al, 2002), *M. tuberculosis (*Jungblut et al, 1999; He et al, 2003), *Haemophilus ducreyi* (Scheffler et al, 2003), *P. aeruginosa* (Nouwens et al, 2000) and *E. coli* (Champion et al, 2000) were found to be highly conserved. The Cordwell (2002) study which compared *S. aureus* strains 8325 (methicillin sensitive) and COL (methicillin resistant), showed that they shared the same pattern of spots except for 12 in COL and 11 in 8325, 23 altogether. This was less than 4% of the total proteome for these strains. This is different from the results obtained with *H. pylori* included in this study in which comparison of any two isolates associated with a disease state displayed at least 280 protein spot differences for each isolate (24% of the protein spots). Additionally this study found that isolates from different disease states displayed at least 373 protein spot differences (32% of the protein spots) (See Table 3.3). However, another factor in the degree of dissimilarity observed between isolates in the present study is the number of isolates included. The majority of comparative proteome analysis studies only included two or three isolates. Except for the Enroth (2000) study, the present study is the only 2-DE proteome analysis to include more than four isolates.

The *H. pylori* proteome displayed a bi-modal pI distribution with a greater percentage (57%) of the spots being located in the acidic range (Figure 3.15). There were few protein spots located at the pH 7-7.5 range. Spots were well resolved throughout the 2-DE gels, however, there were only a few spots (30) in the high M_r basic range, pH 7-11. Figures 3.3B and 3.3D demonstrate how few protein spots were located in the alkaline M_r range 62-150.

4.4 Basic proteins of *H. pylori*

In contrast to other microorganisms whose proteomes have been studied, basic proteins form a large proportion of all 2-DE patterns of *H. pylori* isolates investigated (Jungblut et al, 2000; Enroth et al, 2000). This corresponds with the high pI values calculated from the protein sequences deduced from the genes of *H. pylori*. For example, from the genomic sequence of *H. pylori* strain 26695 it was calculated that more than 70% of the predicted proteins have theoretical pIs greater than 7.0, compared to about 40% in *H. influenzae* and *E. coli* (Tomb et al, 1997). However, results from the current study indicated that only 47% of the *H. pylori* proteome (averaged over 11 isolates) had protein spots with an observable pI > 7.0 this was less than might be expected from the predicted proteins (70%). From figures 3.3 and 3.4, it can be seen that there were more protein spots in pI range 4-7 than is seen in the pI range 6-11. This might suggest that at least 23% of the basic proteins of *H. pylori* may not be represented in the present study. The proteome map of *E. coli* also showed that predicted, basic proteins were not fully represented in the observed proteome map (<35% basic proteins) (Tonella et al, 2001). There are many possible reasons for the basic proteins not being fully represented in the observed proteome. These include extraction procedures, low copy numbers per cell, limitations of the technology (i.e. staining techniques) and predicted versus observed basic proteome (i.e. effects of post-translational modifications). Additionally, when *H. pylori* is removed from its acidic environment many of these basic proteins that may have been required for colonisation in a hostile *in vivo* environment may have ceased expression *in vitro*.

The protein extraction procedure in this study utilised Tris base 40 mmol (high pH ~ 11.00). Acidic proteins tend to extract more effectively at a high pH. This is because when proteins are a long way from their pI their maximum charge tends to increase their solubility in an aqueous solution. However, basic proteins in an alkaline solution are closer to their pI and may precipitate and be lost to solution. Therefore, an alkaline extraction procedure will reduce the number of basic proteins in the sample. Low pH extraction procedures were not explored in this study, which only compared *H. pylori* whole cell proteins extracted by one particular procedure. Further studies are required to examine different acidic extraction

procedures to determine which would provide the best resolution and increase in the number of alkaline proteins. Additional comparative studies can then be performed providing an additional protocol for the comparison of *H. pylori* disease groups.

It was noted in this study that there were only 30 proteins with a M_r of 70-150 kDa within the pH 7-11 region whereas, the pH 4-7 region had greater than 120 high M_r proteins (Figure 3.3). There are three possible explanations for why the observed high M_r proteins in the basic region are poorly represented; these are the M_r, hydrophobicity of the proteins and the method of extraction (previously mentioned). Proteins with M_r greater than 150 kDa are not generally seen on standard 2-DE gels as these proteins are often lost during the isoelectric focusing step. One possible reason for the low numbers of high M_r proteins in the pH 7-11 region is that large molecular mass proteins do not enter the IPG strip readily. Membrane proteins are very hydrophobic with many having pIs greater than 8.0. These very hydrophobic proteins need to remain soluble not only during the extraction procedure but also, during the IEF process, especially at their pI, which is the minimum solubility for each protein. This is not always possible to achieve in spite of advances in detergents and organic solvents for dissolving proteins.

A recent study examined the separation of basic proteins in *H. pylori* (Bae et al, 2003) and attempted to overcome the poor representation of basic proteins in proteome maps. Utilisation of the techniques in this study might have provided a slight improvement in the representation of the basic *H. pylori* proteins. Initially the Bae group utilised the standard 2-DE separation for strain 26695 and obtained good resolution across the entire pH 6-11 range, however there were few resolved spots in the high M_r region of the gel beyond pI 9.5. The current study also found that there were few resolved high M_r spots (70-150 kDa) in the pH 6-11 range.

Subsequently in order to provide an improved separation of the highly basic proteins the Bae et al (2003) study utilised a pH 9-12 IPG constructed with an N-acryloylaminoethoxyethanol (acrylamido) matrix which was combined with improved porosity for the separation of high mass proteins. This however, did not alter the results observed for high M_r proteins and only increased the number of

observable proteins beyond pI 9.5 by 15. The Bae et al (2003) study also attempted to prefractionate the basic protein by a Gradiflow™ and identify these proteins with liquid chromatography/mass spectrometry (LC/MS). The Gradiflow is a versatile, liquid-based protein separation device that separates proteins according to pI (or charge) in a liquid phase (Locke et al, 2002). Analysis of the predicted *H. pylori* proteome indicated that 68 proteins possess a pI of greater than 10. However, Bae et al (2003) were only able to identify 5 of these (7.35%) after pI prefractionation. It was not possible to explore the Bae et al (2003) and Locke et al (2002) procedures in this study for two reasons. These procedures were explored after the experimental procedures in this study were completed. Additionally specialised pH 9-12 IPG strips are not commercially available and need specific equipment to produce them. IPG strips produced within a specific laboratory are not always reproducible for each experiment.

Basic proteins remain difficult to resolve using 2-DE despite the advances in the alkaline range IPGs that have improved the quality and reproducibility of separations in the range of pH 6-11. The lack of better separation at more alkaline pH may be due to the lack of suitable acrylamido buffers, or the instability of the proteins themselves, at extremely alkaline pH.

4.5 Separation of Low Molecular Mass Proteins

The low M_r proteins of various organisms may include proteins with essential biological functions that cannot be visualised and identified using the Tris-glycine buffer systems because of insufficient spot resolution in the lower M_r range. Proteins with M_rs below 20 kDa are poorly resolved using a Tris-glycine Cathode Buffer because they co-migrate with the SDS and this leads to streaking, smearing and poor staining by Coomassie stains. In order to overcome the limitations of the Tris-glycine system this study separated *H. pylori* proteins smaller than 20 kDa using Tricine gels. When Tricine is used as a trailing ion, it moves in front of the SDS front, which improves the resolution of low M_r proteins (Westermeir and Naven, 2002).

Proteins were efficiently resolved using a Tricine gel without a stacking gel, and 3.1 M urea in the polyacrylamide gel. In addition, there was excellent resolution of proteins between M_r 20-62 kDa. Approximately 90 protein spots with $M_r \leq 20$ kDa, which previously could not be reproducibly resolved in this study by the Tris-glycine buffer system, were resolved in this region. On the Tris-tricine gels that were silver stained, approximately 120 proteins were visualised with M_r below 20 kDa (see Figure 3.4). Approximately 20% of these proteins had $M_r \leq 10$ kDa. The fact that only a few proteins with $M_r \leq 10$ kDa were visualised may by partially due to insufficient binding of the stain by the low M_r proteins. CBB stain complexes with basic amino acid residues, such as arginine, tyrosine, lysine and histidine (Link, 1999). Smaller proteins would be expected to contain fewer molar quantities of these residues in comparison to many high M_r proteins. As a result, this would affect the degree to which these low M_r proteins were stained. Another possible explanation is that the majority of small M_r proteins ran off the gels. Alternatively, the low M_r proteins were not expressed in sufficient amounts to be detected on the 2-D gels.

The Tricine method for the separation of low M_r proteins applied in the current study differed from the procedure developed by the Fountoulakis et al study (1998) for *H. influenzae*. Their study reported a streaked appearance and inefficient resolution of proteins greater than 25 kDa. When the *H. pylori* proteins were separated by a methodology that utilised their procedure this resulted in inefficient separation. Alterations in the Tricine method resulted in an improved resolution of proteins between 7-62 kDa. This differed from the Fountoulakis study that only obtained efficient resolution for proteins between 5-20 kDa. This may be due to three possible reasons: 1) their use of a stacking gel; 2) they used a 10.4% acrylamide solution whereas this study utilised a 8.5% acrylamide solution and; 3) differences between the species of bacteria.

The removal of a stacking gel increases the area in the y-axis for the separation of proteins and this may reduce the number of overlapped spots, thus improving the resolution of proteins between the M_r 20 and 62 kDa. The total acrylamide concentration (*T*) may increase the degree of streaking of proteins because when (*T*)

increases the pore size decreases. This study utilised a *T* concentration of 8.5% while Fountoulakis utilised 10.4% therefore the pore size would have been larger in this study which may have decreased streaking. The decrease in pore size would not affect the separation of low M_r proteins but would result in resistance for the movement of high M_r proteins movement. The majority of *H. influenzae* proteins separated by 2D PAGE are detected in the pH 4-7 region (Link et al, 1997; Langen et al, 2000) whereas in *H. pylori* proteins are bipolar with 55% in the pH 4-7 range. An increase in the number of proteins within a particular pH region may increase the proportion of overlapping proteins in the x-axis. The Tricine gel method utilised in this study helped to improve the separation of the small proteins, however a number of potentially small M_r proteins below 7 kDa were not observed implying that improvements for the detection of low M_r proteins of *H. pylori* may be required. These might include the use of dyes such as SYPRO™ Ruby and Liquid Chromatography-Mass Spectrometry.

4.6 Theoretical versus observed proteome

Alm et al (1999) predicted that the genome of the *H. pylori* isolates 26695 and J99 encode for 1552 and 1495 proteins, respectively. In the present study, the whole cell protein patterns of the *H. pylori* isolates revealed approximately 1160 proteins when stained by CBB G250 and about 1580 proteins stained by silver in the 2-DE gels (Table 3.1). Given that the entire genome of *H. pylori* encodes for approximately 1500 proteins it would seem that the majority of the proteome of this organism has been observed. However, the total number of protein spots observed by 2-DE may be greater than the number of theoretical predicted proteins, since a number of proteins can be processed by proteolytic enzymes or modified by processes such as methylation, phosphorylation, and acetylation. Post-translational modifications will increase the number of protein spots per gene product that are observed on the 2-DE gels. From this the global proteome of *H. pylori* would actually be greatly reduced from its original 1160 proteins.

The formation of dimers and truncation of proteins has been observed in many protein spots in 2D gels (Krah et al, 2003; Verdan et al, 2003; Pichoff et al, 1995;

Pichoff et al, 1997). Dimers of proteins will double the M_r of the protein and a truncation can give an observed M_r less than the theoretical M_r. There are also proteins that are processed into two or more subunits and appear to be fragmented.

Once each protein has been associated with a particular *H. pylori* gene, its apparent M_r and pI, as estimated by its mobility on a 2-DE gel, may be compared with the theoretical M_r and pI of the unmodified product of the gene. Figure 3.14 consists of scatter diagrams comparing the observed properties of proteins on 2-DE gels with the predicted properties of the corresponding theoretical gene products (www.tigr.com).

4.6.1 Posttranslational modifications

A large number of the 2-DE pI values highly corresponded to the theoretical pIs of protein spots in the pH 4.0 - 7.0 range (Figure 3.14b). However, the pI values for the protein spots in the pH 7.0–11.0 2-DE gels showed a larger variation from the theoretical pIs. Eleven of the pI data points in the plot were outliers in the pH range 7.0-11.0, these were spot numbers 74, 83, 89, 92, 108, 114, 123, 128, 138, 153 and 159. The remaining three outliers in the pH 4.0-7.0 range were spot numbers 28, 32 and 158. Spots were considered outliers when the pI variation was greater than 1.0. The post-translational processing and modification of proteins may explain why the 2-DE and theoretical pIs match less closely than the 2-DE M_r and theoretical M_r.

In this study, the expression products of many ORFs appeared in the form of a horizontal serial charge train within the 2-DE gel. This may have resulted from different isoforms of one protein with differently charged side groups caused by post-translational modifications. Examples of this phenomenon are the following identified proteins: translation elongation factor EF-Tu (HP1205) five spots, catalase (HP0875) three spots, alkyl hydroperoxide reductase (HP1563) eight spots, urease alpha-subunit (HP0073) six spots, and chaperone and heat shock protein GroEL (HP0010) 12 spots (Figure 3.6). The post-translational modifications of these proteins would have major effects on their charge properties while making little change to their size. The urease beta subunit located at 65 kDa and within pH 5.6-6.2 was the only protein identified where the post-translational modifications had an observable alteration in its M_r (Figure 3.6). Urease beta subunit was observed as two

rows of spots separated by a M_r change of 5 kDa. Protein charge trains have been observed in previous proteome studies of *H. pylori* (Fischer and Haas, 2004; Lock et al, 2001) and may be a common feature of eukaryotic and prokaryotic proteins.

4.6.2 Observed M_r versus Theoretical M_r

In general, the estimated M_r of the proteins in the 2-DE gels closely matched their theoretical values. In Figure 3.14a, the 2-DE molecular mass is plotted M_r against theoretical M_r. When the total data is related to an equivalence plot (slope 1.0, passing through the origin), the correlation coefficient is 0.983 which is significant at the $P<0.01$ level. Eleven of the M_r data points 49, 50, 64, 74, 83, 87, 103, 105, 112, 137 and 138 were outliers in the M_r plot, six of these proteins have no known function. The M_r of five of these outliers differed by ±33% (spot numbers 49, 50, 74 83 and 105), four of these proteins had a known function. The intact protein M_r, pre-filtering, is imperfect because sequences in protein databases often exist as dimers, fragmented, and subunit forms (MS-Fit information website). It is for this reason that when a protein match could not be found within the ±15% range that the range would be expanded to try and find a protein match. Outliers in this study are defined as those where the observed M_r differs by > ±15% from the theoretical M_r. These M_r outliers might be the result of dimers, truncations or can be subunits of the same protein.

Dimers and Truncation

Dimers are compounds formed by the joining of disulphide bridges among similar chains. Artificial dimerisation can occur during the run of the second dimension but this will cause smearing on the gel. Dimerisation has little effect on the pI and can occur in the first dimension if an insufficient amount of a reducing agent such as dithiothreitol is used. This reason for dimerisation is unlikely to occur if tributyl phosphate is used. Alternatively, dimers can occur *in vivo*, and the concentration of reducing agent in the sample buffer insufficient to break all the disulfide bonds. *H. pylori* proteins have been identified that form dimmers. A particular example is the 26 kDa protein, HP1563, (known as alkyl hydroperoxide reductase) a protein that consists of a number of isomers (Spot numbers 175-179) (Figure 3.13B, Table 3.16). Krah et al (2003) identified eight different alkyl hydroperoxide reductase spots, one

spot had an apparent M_r that was double the theoretical mass of the protein. The Krah et al (2003) study found evidence that this spot contained dimers of the protein because cysteine-containing peptides were not found in the dimer spot. It was believed in this instance that alkyl hydroperoxide reductase formed dimers *in vivo*. It has also been found that other members of the peroxiredoxin family form homodimers or even decamers (Wood et al, 2002). ABC-ATPases have been found to form head-to-tail dimers in *E. coli* and *Sulfolobus solfataricus* (Verdan et al, 2003). Given that alkyl hydroperoxide reductase has been found to form dimers in *H. pylori* it is possible that this has also occurred in the ABC transporter, ATP-binding protein (spot number 50, Table 3.12). Another protein that might form dimers in *H. pylori* is the cell division topological specificity factor (spot number 105). Studies in *E. coli* (Pichoff et al, 1997) and two-hybrid yeast analyses (Pichoff et al, 1995) did not find dimerisation of this protein. Further investigations would be required to determine if it occurs in *H. pylori*.

Truncation has been found to occur in *H. pylori* heat shock protein GroEL either as N-terminally truncated or C-terminally truncated spots (Krah et al, 2003). This truncation may have occurred in phosphoribosylpyrophosphate synthetase (spot number 83) such, truncation of both the N-terminal and the C-terminal might account for a significant 2/3 difference of the observed M_r. This truncation effect might also have occurred in hypothetical protein JHP0939 (spot number 74).

Subunits

Gamma-glutamyltranspeptidase (spot number 49) located at pI/ M_r 9.0/40 appears to be a truncated protein but might be one of two subunits for this protein. Figure 3.7b shows its location on the proteome map and Table 3.12 shows the difference between the 2-DE observed M_r and the theoretical M_r. This corresponds to the study of Krah et al (2003) in which two gamma-glutamyltranspeptidase proteins were detected, one at pI/ M_r 8.7/40 and the other at pI/ M_r 6.7/20. Gamma-glutamyltranspeptidase has been shown to consist of two subunits in *E. coli* K-12 (Hashimoto et al, 1995). The current study and the Hashimoto study did not identify a spot corresponding to the whole protein (theoretical M_r 61 kDa). Therefore, it is most likely that this apparent truncation is due to the existence of two subunits for this protein.

4.6.3 Protein Modal Distribution

When the current study analysed the protein isoelectric points of *H. pylori*, a bimodal pattern was observed, with few proteins in the plot having pI values near 7.5 (Figure 3.15). Recently Van Boegelen *et al* (1999) analysed several bacteria and identified a distinct bimodal pattern of pIs with peaks centred at approximately pI 5.5 and pI 9. Because the cytoplasm of most cells has a pH near neutrality (pH 7.2 - 7.4) it may be difficult to maintain protein structure, rigidity and solubility near their isoelectric points. Although the causes for and general bimodality were not established it was suggested that it may be a result of the relationship between intracellular pH values and protein pI values (Van Bogelen et al, 1999). The bimodal distribution of protein pIs with pIs greater than, or less than, pH 7.0 has been observed in *M. tuberculosis* (Brooke et al, 1998), *E. coli* (Blattner et al, 1997) and previous studies of *H. pylori* (Lock et al, 2001) and may be a general characteristic of prokaryotes. Studies of *H. pylori*, including the current study, have observed that there appears to be a greater number of basic proteins in both the observed and theoretical proteomes than has been found for other bacteria, such as *E. coli* (Tonella et al, 2001). These proteins may enable *H. pylori* to survive in its naturally acidic habitat.

4.6.4 In vitro and in vivo

The theoretical proteome can differ from the observed due to variations in an organism's environment. The *H. pylori* isolates 26695 and J99 consist of 1600 and 1520 gene products, respectively. Certain proteins are only stimulated under specific environmental conditions that may not be fully reproduced *in vitro*, for example pH, heat and stress, iron starvation and oxidative stress. An *in vivo* study would be required to investigate the combined effect of these factors on protein expression. In addition, the interactions of outer member proteins, Cag PAI proteins and vacuolating cytotoxin with the host may only be examined *in vivo*. As it was not possible to examine protein differences in the *in vivo* environment by 2-DE only a portion of the entire genome of the *H. pylori* isolates was observed.

A number of *in vitro* and animal studies have been conducted investigating virulence factors. However, it is evident that an organism may express different proteins when

grown under laboratory conditions compared with when it is grown in *vivo*. It has been observed in past studies that various isolates of *H. pylori* have exhibited a sudden and complete loss of proteins that are constitutively expressed under *in vivo* conditions when grown *in vitro* (Manos et al, 1998). One example of a protein where this can occur is catalase (spot numbers 140-142, Figure 3.12 Table 3.15). In the Manos et al (1998) study it was found that two isolates UNSW-RU1 and UNSW-N6 ceased expressing catalase after 35 and 30 passages respectively. The current study did not observe the loss of the expression of this protein most likely due to the low number of subcultures. This study only allowed four subcultures to reduce any morphological changes to the organism that can occur after repeated subcultures.

H. pylori catalase is a monomer lacking peroxidase activity and shares several properties with catalases of eukaryotic organisms. The protein catalase is expressed *in vivo* in response to oxidative stress in the gastric environment. Once the selective pressures are relieved by culturing *in vitro* the expression of catalase is no longer required and this results in catalase-negative isolates (Manos et al, 1998). However, when this bacterium is restored to the gastric environment the catalase expression returns to normal (Manos et al, 1998).

Growing *H. pylori in vitro* would therefore fail to examine the bacteria-host interactions under the multiple environmental conditions that occur in the gastric environment. However, the difficulty of examining proteins obtained from biopsies by 2-DE proteome analysis is the inability to obtain sufficient material for detection on gels. Although new detection techniques have been developed (e.g. SYPRO Ruby) these still have limitations in the amounts of protein that may be observed. Additional techniques would be required to enhance the observation of low copy proteins.

4.7 Cluster Analysis of *H. pylori*

Using match results obtained from PDQUEST, two subclustering analyses (qualitative and quantitative) were performed to determine if the isolates could be grouped according to their disease type. Hierarchical clustering was performed to

form dendrograms (Figure 3.5). The nodes of the tree represent the individual objects, while the branches of the cluster the nodes according to similarity or difference, which are represented by branch length. Strikingly, in both analyses isolates within a disease group clustered together into branches of the tree. Slight variations were observed between the different dendrograms in the order in which the isolates appeared within the groupings. The qualitative dendrogram suggested that isolates from the same patient group clustered together independently of the calculation method. In addition, the qualitative dendrogram formed two major clusters – gastric and duodenal. However, the quantitative dendrogram did not reveal that isolates from specific disease groups clustered together but suggests that isolates from the same environment (gastric or duodenal) formed clusters. The GU, NUD and GI isolates were more closely related than were the DU isolates to those within the other disease groups (Table 3.9). To some extent, these two dendrograms indicate that there might be disease-specific proteins. This corresponds to a cluster analysis of *H. pylori* which found that some isolates within each disease group clustered together (Enroth et al, 2000).

4.8 Variability among *H. pylori* isolates

Bacteria have varied genomes that reflect the phenotypic differences that facilitate adaptation to diverse environments. Bacteria that are subjected to rapidly changing environments display genomic plasticity that leads to the generation of sub-strains with various adaptations (Aras et al, 2003). This high degree of genetic variability, such as present in *H. pylori*, may also contribute to long-term host adaptation. Host selection of variants that are well adapted to particular environmental constraints is one mechanism for regulating host-microbial interactions (Vicente et al, 1999). The expression of bacterial genes can be regulated in response to environmental cues to allow rapid adaptation to changing conditions, for example Hsp12 (de Vries et al, 2003) and FrpB (Delany et al, 2001b). To optimize their chances for survival and growth, bacteria employ different strategies to adapt to their environment. These strategies include spontaneous point mutations, recombination with other bacterial cells, and intragenomic rearrangements involving mobile gene elements or repetitive DNA sequences; this genetic diversity may be spontaneous (Arber, 2000).

Initially the genomes of 26695 and J99 were believed have an overall genomic conservation and gene order (Alm et al, 1999). However, comparative proteome analyses of different isolates of *H. pylori* has since revealed that *H. pylori* is highly variable (Jungblut et al, 2000; Govorun et al, 2003). Nevertheless, considering the high sensitivity of isoelectric focusing to changes in single amino acids this proteome variability may reflect the exact chemical structure of the protein species and not the presence of proteins with different functions. The 2-DE approach analyses the genetic variability at the protein species level. Further the identification of protein species of different *H. pylori* isolates will detect differences in the proteins. This study identified nine proteins which showed variability in their migration on the gels and this could be due to 1) post-translational modifications GroEL, TsaA, SodF, Rpl7/12, Fba, RpoA and NapA (see Table 3.16), 2) mutations Flagellin A (FlaA) and 3) plasmids (orf5O) (see spot no 99).

4.8.1 Post translational modifications

Comparison of the 11 isolates included in this study revealed that there were seven proteins with different pI values (shifts) on the different proteome maps (Table 3.15). These differences were found for the housekeeping protein groEL, enzymes involved in free radical and hydrogen peroxide catabolism (SodF, NapA and TsaA), fructose-biphosphate aldolase (Fba) and DNA-directed RNA polymerase alpha subunit (RpoA) and ribosomal protein l7/12 (Rpl7/12). These changes may be attributed to possible posttranslational modifications. Two previous studies have also noted the existence of these pI shifts. Jungblut et al (2000) compared the amino acid composition of ten *H. pylori* proteins. Three proteins were found to have an identical composition and three proteins had amino acid exchanges without a net charge change which did not appear to affect their 2-DE gel positions (Jungblut et al, 2000). However, the other four proteins (GroEL, TsaA, SodF and Rpl7/12) had amino acid exchanges that resulted in a net charge change that altered the 2-DE gel pI values (Jungblut et al, 2000). The study by Govorun et al (2003) identified GroEL, SodF, TagD, Omp18 and Tig proteins with a 2-DE gel pI shift. The present study has increased the number of proteins with observed 2-DE pI shifts to include an additional three proteins, these being Fba, RpoA and NapA.

4.8.2 Variation in gene expression

H. pylori isolates over time have all evolved in relative isolation in the stomach of individual humans. The very special habitat of *H. pylori* at the surface of gastric epithelial cells or in the mucous layer covering the epithelium suggests that this bacterium has evolved specialized features for adaptation. There are variations in the virulence factors of *H. pylori* that are not produced by the effects of subculturing (e.g. FlaA). There are two possible explanations for these variations. These are mutations or deletions and adaptation to the host environment. Some of the variations may be the result of one or more possibilities.

The *H. pylori* isolates examined in this study showed considerable diversity in their expression of flagellin proteins. Examination of the eleven isolates in the current study showed that not all isolates expressed the same quantity of FlaA (see Appendix Figures A.23, A.24 and A.25). In some isolates the amounts of FlaA produced varied from undetectable to an abundance. In this study it was found that *in vitro* GU1 FlaA was barely visible and in Tomb 26695 it was not detected while another isolate NUD3 produced six times the spot intensity of the isolates DU1, GI1 and NUD1. These differences in spot intensities between the isolates was also noted for FlaB and FliM. Other proteins involved in flagellin expression that might have been co-expressed with this protein were not noted because this study only examined differences between disease groups (specifically absent/specifically unique) or positionally conserved/differentially regulated spots for all isolates. Further studies would need to be conducted in order to correlate other possible co-expressed proteins using FlaB and FliM as reference points.

GU1 was obtained from a recent biopsy in June 2001 and was only subcultured four times in the laboratory which means that its poor flagellin production is unlikely to be the result of repeated laboratory subculturing. In contrast GI2 (26695) had undergone repeated subculturing and from the protein pattern did not appear to produce FlaA, whereas GI1 (NCTC11637), which was originally isolated in 1982, produced abundant levels of FlaA. GI1 FlaA displayed equivalent spot intensity to isolates, obtained from three recent biopsies, while other isolates NUD2, NUD3 and

DU2 displayed an abundance of FlaA in comparisons to all the other isolates. Therefore if the variation of FlaA expression was due to spontaneous mutations from repeated subculturing then both isolates GI1 and GI2 should not have detectable levels of FlaA in their proteome. The most likely explanation is that another factor is influencing the variation of FlaA.

Mutations or deletions of specific genes may result directly or indirectly in the expression of other *H. pylori* gene products. Genetic mutations in specific genes have been shown in past studies (Figura et al, 2004; van Amsterdam and van der Ende, 2004). Bacterial flagella are essential for the motility of *H.* pylori to enable the colonisation of the gastric mucosa and to attain robust infection. CagA (Figura et al, 2004) and other genes such as *ylxH* (van Amsterdam and van der Ende 2004) may be essential for the assembly of flagella and the motility of *H. pylori*. In the Figura et al (2004) study it was found that the disruption of *cagA*, by gene deletion, that the lack of CagA expression reduces the expression of both flagellin A and flagellin B, two central components of flagella (Figura et al, 2004) The decreased expression was found to be sufficient to result in a non-motile strain. *H. pylori* YlxH (HP1034) is described as an ATP-binding protein with homology to the MinD family of ATPases that are involved in chromosome partitioning (NCBI database). However because *H. pylori* strains contain a MinD homologue (HP0331) it is believed that YlxH does not play a role in cell division (van Amsterdam and van der Ende, 2004). The deletion of the *ylxH* gene results in a non-motile strain in which the FlaA is down regulated and flagella are absent. Therefore it is believed that the *H. pylori* YlxH protein is essential for the assembly of flagella and hence for the motility of *H. pylori* (van Amsterdam and van der Ende, 2004).

4.8.3 Plasmids as a source of proteomic diversity

Approximately 50% of *H. pylori* strains carry cryptic plasmids ranging in size from 2 to about 100 Kb; however, the role of these plasmids is not well understood (Penfold et al, 1988). *H. pylori* can distribute a high number of diverse genetic modules within the species. Different combinations of modules might be created by recombination (deletion and insertion) and selected for by the needs of the bacteria in their individual hosts. Plasmids can acquire chromosomal genes and genes from

other plasmids by transposition and recombination. These plasmids with newly acquired genes will now produce different proteins and if the plasmids are acquired by other strains they will alter the proteome of the strains. If these plasmids are advantageous to their new host then these strains will have a selective advantage. Plasmids can be transferred between bacteria by transduction, transformation (Ando et al, 1999; Hofreuter and Haas, 2002) and conjugation, however plasmid conjugation has not yet been proven for *H. pylori*. Several genes that are involved in the transformation process have been identified (Ando et al, 1999; Hofreuter et al, 1998).

The GU isolates studied here produced a unique protein of unknown function orf5O, the best homologue to the hypothetical protein HP1334 (spot no 99 Table 3.13), an open reading frame which has been found to be located in/on plasmid pHel5 (Hofreuter and Haas, 2002). This suggests that the GU strains either have plasmids encoding orf5O, have acquired it from a plasmid of the pHeI5 type or were the source of the orf5O encoded by pHeI5. Hofreuter and Haas (2002) theorised that the pHe15 and pHe14 plasmids might provide a possible explanation for the macrodiversity of *H. pylori* strains. The carriage and/or integration of plasmids represent an important source of genetic material that may play an important role in genetic diversity and the diseases produced by a strain.

4.9 Conserved proteins

In the present study, 471 spots (40.6% of the proteome) were found to be positionally conserved. Positionally conserved spots were of two subsets: 1) spots which vary by less than a factor of three (quantitatively conserved); and 2) spots which vary in volume by a factor of three or more. Thirty percent of the proteome is quantitatively conserved. One hundred and twenty proteins of the 471 spots, which were positionally conserved, were identified (Table 3.12). These included 96 proteins that were quantitatively conserved (Figures 3.6, 3.7 and Table 3.12) and 24 proteins where expression differed by a factor greater than three (Figure 3.12 and Table 3.16). Fifty eight of the 120 positionally conserved proteins (48%) identified were protein isomers that may be due to post-translational modifications.

Fifty four of the 120 proteins were identified as being positionally conserved were found to belong to the following categories: electron transport proteins (5); translation factors (6); transcription factors (4); redox-oxidation proteins (2); cations (3); aspartate family (4); tricarboxylic acid cycle proteins (6); surface structures (2); urea related proteins (15); degradation of proteins (2); peptides and glycopeptides (3); salvage and interconversion of nucleosides and nucleotides (2); DNA restriction modification and repair (2); ribosomal proteins (3); synthesis and modification (1); fatty acid and phospholipid metabolism (2); detoxification (3); general cellular processes (5); central intermediary metabolism (4); molybdopterin (2); glutathionine (3); and gluconeogenesis (2).

Initially this study included ten isolates. A comparison of the ten isolates found that 48% (566 spots) of the proteome was conserved. After the addition of strain 26695, the conserved proteome was reduced to 40.6%. It is possible that if more isolates were included in the study that the percentage of conserved proteome might be even further reduced.

From the 2-D proteome map of the isolates included in this study it has been concluded that *H. pylori* is highly diverse with relatively few protein spots in common to all isolates. The conserved proteome covers a broad selection of proteins many of which are essential for maintenance of the cell. However, while many genes are essential, alterations, such as post-translational modifications can alter the proteome such that a comparison between the proteome of two similar isolates could appear dissimilar. An example of this is fructose bisphosphatase that is located in the same position in the 2-DE gel for only seven out of the eleven isolates (Table 3.16).

4.10 Proteomic Differences between Disease Groups

Sixty six significant protein spot differences, both specifically absent/unique and differentially expressed, were identified between the four disease groups, GU, DU, NUD and GI. Eleven of the 66 protein spots were isomers, KatA three isomers; Icd two isomers; TrxA two isomers; HP0958 – two isomers; and Pfr two isomers. These

proteins could be considered under five categories (according to classifications in the NCBI database). These categories are:

- metabolism – 14% (9 spots)
- cellular processes and signalling – 7 % (5 spots)
- information processing – 9% (6 spots)
- pathogenesis – 3% (2 spots)
- poorly characterised or unknown proteins – 67% (44 spots)

Spot differences between disease groups where the proteins had known functions were:

- one disease group versus three disease groups
    - DU - KatA, Icd, MetK, MinE, PlsX, JHP0295, JHP0140, YjgF, SpaP
    - NUD – MiaB, Hsp12 variant C and CeuE
    - GU – HP0958, TrxA/TrxB and RecA
    - GI – Ptc1, Predicted GTPases
- two disease groups versus two other disease groups
    - DU and NUD – Pfr, ComGF, Hom
    - GI - CagN

4.10.1 GU Unique and Absent Proteins

Thirteen disease specific proteins were identified for the GU isolates (six unique and seven absent). The specifically unique proteins (Figure 3.8 and Table 3.13) were either poorly characterised proteins or those with no known function.

Out of the seven absent proteins identified, three had a known or homologous function, the remainder had no known function or were poorly characterised (Figure 3.9 and Table 3.14).

4.10.1.1 Unique Proteins

Unknown protein Orf5O (spot number 99)

As discussed in section 4.7.3 the role of *H. pylori* plasmids is not well understood. Plasmids can be grouped into at least two separate classes, one with homology to

plasmids of the Gram-positive bacteria replicating via the "rollingcircle" mechanism (Kleanthous et al, 1991), the other belonging to into the group of iteron-containing plasmids and replicating via the theta mechanism (de Ungria et al, 1999). H. pylori contains five plasticity zones which have been speculated to have been obtained by horizontal gene transfer (Hofreuter and Haas, 2002). The plasmid pHe15 contains open reading frames with homology to products of genes clustered in plasticity zone three of *H. pylori* 26695 (Hofreuter and Haas, 2002). Such gene sequences can be exchanged between *H. pylori* plasmids and the *H. pylori* chromosome.

Further, proteins encoded by pHel5 and pHel4 have homology with products of genes widespread in the chromosomes of *H. pylori* 26695 and J99 Orf5O (best homologue to hypothetical protein HP1334 - 36.2%, HP0879 26.7%, JHP0812 33.5%) (Hofreuter and Haas, 2002). HP1334, HP0879 and JHP0812 are all poorly characterized proteins whose functions are unknown.

4.10.1.2 Absent Proteins

HP0958 (spot number 120)

This is a Zn-ribbon protein, possibly a nucleic acid-binding protein belonging to the information processing category of replication, recombination and repair.

This particular protein is part of a serial charge train. Another protein in the charge train was identified as a down-regulated protein of the proteins found in the isolates in other disease groups (see page 188). Therefore, the absence of this HP0958 in the GU isolates is most likely due to its decreased expression.

The zinc ribbon C(H/C)CC, metal (zinc) binding motifs are structurally diverse and are present among proteins that perform a broad range of functions in various cellular processes, such as DNA replication, repair, transcription and translation, and metabolism and signalling (Vallee and Falchuk, 1993). For example, the zinc ribbon, metal-binding motif is highly conserved amongst the TF (II) Bs of eukaryotes and the Archae and is found in several proteins that are involved in transcription such as TFIIS, RNA pol II subunits and TFIIE (Wang et al, 1998; Jeon et al, 1994; Frick and Richardson, 2001). Given the multiple functions of proteins having a Zn-ribbon,

it is not possible to determine the function of this zinc-ribbon protein in *H. pylori*. However, recent research has proposed that HP0958 is essential for normal motility and flagella production in *H. pylori* (Pereira and Hoover, 2005; Ryan et al, 2005).

Motility is an essential colonization factor for *H. pylori* in experimental infection models (Eaton et al, 1989; Eaton et al, 1996; Holger et al, 2003). Flagella are composed of two subunits Flagellin A (FlaA) (spot number 183 - 188), Flagellin B (FlaB) (spot number 189 – 193) (Kostrrzynska et al, 1991) and the flagellar hook component (FlgE) (O'Toole et al, 1994). The production of these flagellar proteins is regulated by at least three RNA polymerases; sigma factors σ^{80} (RpoD), σ^{54} (RpoN) and σ^{28} (FliA) (Alm et al, 1999; Tomb et al 1997) and by anti-sense σ^{28} factor (FlgM) (Colland et al, 2001; Josenhans et al 2002). The study by Pereira and Hoover (2005) found that cells defective in the HP0958 gene expressed very small amounts of FlaA and FlaB and had no observable FlgE. There was no alteration in the expression of RpoN and FliA whereas the expression of FlgM (spot number 194) was increased nine-fold in the HP0958 mutant in comparison to the wild-type (Periera and Hoover, 2005). Other important regulators whose expression was decreased in the strain which had a mutation in HP0958 were FlhA and FlhF. When Perier and Hoover compared the production of FlaB, FlgE and FlaA it was found that these proteins were significantly reduced in the mutant strain.

In the present study it was observed that in the isolates GU1, GU2 and GU3 the spot intensity for HP0958 was decreased more than three-fold in comparison to the other isolates. However, another notable characteristic of all eleven isolates was the production of FlaA, FlaB and FlgM. Spot intensities varied from non-existent (GI2) to highly abundant (NUD3 and DU2) but did not correspond to alterations in the production of HP0958. Isolates in the GU disease group had similar spot intensities for HP0958 but not for the proteins FlaA, FlaB and FlgM. The expression of these proteins in GU1 was almost non-existent. In the GU2 isolate it was found that the expression of these proteins was equivalent to GI1 and DU1 however, DU2 which had a strong spot intensity for HP0958, had a similar expression of FlaA, FlaB and FlgM to GU1. NUD3 was found to have abundant levels of these proteins but had a

reduced level of expression of HP0958 in comparison to DU2 but it was not a significant difference.

HP0825 (spot number 121)

Thioredoxin reductase (TrxB) belongs to the cellular processing and signalling category of proteins and the posttranslational modification, protein turnover, chaperones group. In the proteome of 26695 (Tomb et al, 1997) at least two isomers of TrxB were identified. One isomer had a pI of 5.93 and a M_r of 33.5 kDa and the other a pI of 6.42 and a M_r of 36 kDa.

In this study TrxB was not identified in GU isolates however, its counterpart, TrxA, in the thiol-dependent reduction-oxidation system was present as a highly abundant protein. TrxA was seen as a differentially up-regulated protein in the GU isolates in this study.

The redox protein thioredoxin (TrxA) and the associated enzyme thioredoxin reductase (TrxB) constitute a thiol-dependent, oxidation-reduction system that can catalyse the reduction of certain proteins by NADPH, usually with high selectivity (Holmgren, 1985). In anaerobic bacteria, the generation of low redox potential reductants, such as TrxA, can be used to assist electron flow to specific substrates. Thioredoxin reductase catalyses the reduction of the oxidised thioredoxin (Trx-S2) by NADPH, and the reduced thioredoxin (Trx-[SH]2) is the disulphide reductase (Swissprot database). There are two nearly identically subunits to this enzyme. Each subunit contains 1 mol of FAD per subunit and a NAD(P)H and FAD binding motif.

There are two explanations for the absence of TrxB in the GU isolates. The first is the differential expression of the thioredoxin system so that TrxB is down regulated in response to oxidative stress. Although TrxB was not observed this does not mean that it is not being expressed. It may be produced in insufficient quantities to be detected by Coomassie. Alternatively, a shift in the molecular mass may have occurred due to post-translational modification so that this protein spot appears to be absent. However, given that TrxA was down regulated in the isolates in the other disease groups the differential regulation of TrxB is the most likely explanation.

Windle (1999) has shown that the thioredoxin system in *H. pylori* behaves as a stress response element by secreting TrxA in response to chemical, biological and environmental stresses. The differential expression of TrxA and TrxB, and other redox active proteins in response to various stressors, has been observed both in eukaryotic cells and in *Staphylococcus aureus* (Berggren et al, 1996; Reddy et al 1999). Oxidative stress has been shown to result in the down-regulation of TrxB activity and an increase in the TrxA activity in cultured small intestine epithelial cells (Higashikubo et al, 1999). While the effects of oxidative stress on *H. pylori* expression of the thioredoxin system have not yet been fully reported, it is possible that a similar response occurs in this organism. Two other proteins that are differentially regulated in response to oxidative stress were also identified in this study, these are non-haeme iron (III) ferritin (NUD and DU isolates [see page 196]) and catalase (DU [see page 192]).

HP1499 (spot number 123)

This is predicted to belong to the HKD family of nucleases.

This HKD sequence motif has been found in a variety of enzymes having a very widespread range of activity, including phospholipase D, endonucleases, phosphatidylserine synthases (PSS), cardiolipin synthases, as well as proteins having no known function *in vivo* (Koonin, 1996; Morris et al 1996). The majority of investigations of enzymes with the HKD motif have been performed with phospholipase D (PLD). PLD is a membrane-associated protein in which the protein's interaction with the membrane may be mediated by two or more domains (Xie et al, 1998). The common denominator of proteins belonging to this PLD superfamily of enzymes is that they all perform catalysis involving a phosphorous atom. The PLD superfamily of enzymes catalyse the hydrolysis of phospholipids into phosphatidic acid and have a polar head group (Leiros et al, 2004).

In addition to this hydrolytic activity, PLD enzymes also catalyses a transphosphatidylation reaction *in vitro* to form new phospholipids (Waite, 1999). Even at very low levels (1%) of ethanol, transphosphatidylation can be preferred,

resulting in the almost exclusive production of phosphatidylethanol rather than phosphatidic acid and choline (Waite, 1999). This reaction is unique to PLD enzymes and thus provides an unambiguous indication of the presence of PLD activity (Leiros et al, 2004).

Mutation of the HKD motifs residing in region I or IV renders PLD proteins inactive but does not affect expression of the protein (Xie et al, 1998). If mutation in the PLD protein is due to a truncation this would result in a M_r shift, which in turn, would result in an apparently specifically unique protein in the GU group. Further investigation is required to ascertain the reason why the HP1499 protein was not detected in the GU isolates.

4.10.2 Unique and Absent Proteins in DU Isolates

In this study eight unique proteins (Figure 3.9 and Table 3.13) and four absent proteins (Figure 3.8 and Table 3.14) were identified in the DU isolates.

4.10.2.1 Unique Proteins

HP0332 (spot number 105)

The MinE (cell division topological specificity factor) protein belongs to the cellular processing and signalling category. This protein is involved in cell cycle control, mitosis and meiosis. Both MinE (spot number 105) and MetK (spot number 143) are two proteins that were identified in this study, which have possible roles in cell division. The role of MetK is described in section 4.9.6.

In *E. coli* cell division is controlled by the Min system. It consists of three proteins MinC, MinD, and MinE. These are necessary for site-specific inhibition of septation at the cell poles (Pichoff et al 1997: de Boer et al, 1992). MinC is the inhibitor of cell division and is activated by MinD. The MinD protein has two functions, as well as activating MinC it interacts with the MinE topological specificity factor (Pichoff et al 1997: de Boer et al, 1992). The MinE protein regulates MinC in a topological fashion and allows septation at the central region of the cell but prevents it at the cell poles. Currently it is not known how the MinE protein suppresses the MinC/MinD-mediated inhibition of division. The absence of the MinC or MinD protein or the

overexpression of *minE* results in the blocking of division being suppressed at the cell poles as well as at the mid cell, leading to minicell formation (de Boer et al, 1992). *H. pylori* does not contain a homologue to MinC but the homologue to the MinD and MinE have been identified (Doig et al, 1999).

JHP0140 (spot number 108)

This is predicted to be a periplasmic solute-binding protein belonging to the poorly characterised category whose general function is predicted only.

Periplasmic solute-binding proteins (SBPs [also called ABC transporters]) are part of a transport system that is composed of two cytoplasmic membrane domains and two peripheral membrane-associated ATP binding domains (Driessen, 2000). In Gram-negative bacteria these SBPs freely diffuse into the periplasm, however in gram-positive bacteria they are bound to the cell membrane and some are bound to the membrane domain (Obis D et al, 1999). The SBP transporters are required for the uptake of a number of small molecules that include amino acids, metal ions and sugars (Higgins, 1992). SBPs are high affinity, ATP driven, transporters that function in a unidirectional manner to drive the accumulation of solutes against high concentration gradients (>10, 000-fold) (Higgins, 1992). Currently there is no known function for this protein in *H. pylori*.

JHP0376 (spot number 104)

This has no known function but is predicted to be a nucleic-acid-binding protein involved in transcription termination and therefore would belong to the information processing category.

This protein is homologous to the following proteins: HP1049 in *H. pylori 26695*; CJ0135 in *Campylobacter jejuni NCTC 11168* and HH1012 in *H. hepaticus ATCC 51449*. These proteins also have no known function.

HP0377 (spot number 103)

This belongs to the poorly characterized category of proteins and has no known function.

HP0036 (spot number 109), JHP0051 (spot number 107) and HP1437 (spot number 106)

These proteins have no known function.

4.10.2.2 Absent Proteins

HP0944 (spot number 127)

This protein belongs to the YjgF/YER057c/UK114 family. This family belongs to the information processing category and is involved in transcription.

The YjgF/YER057c/UK114 family is a highly conserved class of proteins that is widely distributed in bacteria, Archaea, plants and other eukaryotes. A wide variety of biological roles has been attributed to these proteins depending on the organism examined. A number of the proteins have been described as putative translation inhibitors based on experiments with the human (hp145 or PSP1) and rat homologues (rp145 or L-PSP). Both of these proteins inhibit cell-free protein synthesis in the rabbit reticulocyte lysate system (Schmiedeknecht et al, 1996). It has also been characterised in a few bacterial species such as *Bacillus subtilis* (Rappu et al, 1999). In *B. subtilis*, the protein is required for adenine-mediated repression of purine synthesis (Sinha et al, 1999). In yeast, the homologue appears to play a dual role in mitochondrial maintenance and the regulation of isoleucine biosynthesis. However, to date (December 2005) the precise biochemical function of this protein in *H. pylori* species remains unknown (Schmitz and Downs, 2004).

In several organisms, strains lacking an YjgF homologue have a defect in branched-chain amino acid biosynthesis. In strains lacking *yjgF* the specific activity of transaminase B, which catalyses the last step in the synthesis of isoleucine, was reduced (Schmitz and Downs, 2004).

Therefore, based on homology with proteins belonging to other species, it is possible that this protein also functions as an inhibitor in *H. pylori*. In addition it was found from the comparison between the genome of the two isolates J99 (a DU isolate) and 26695 (a GI isolate) that this gene product has a different pI in each strain (Alm et al, 1999). The theoretical pIs of J99 and 26695 were 5.2 and 4.5 respectively (NCBI database). In this study the YjgF protein was found to have an observed pI of 4.5 in the GU, NUD and GI isolates studied. If all DU isolates have a similar pI for the YjgF as J99 then it would be expected that it would be absent at the pI of 4.5 observed for YjgF homologue in the GU, NUD and DU isolates. Another reason why this particular protein was not found in the DU isolates at the pI of 5.2 is because it may be a low copy protein. Low copy number proteins would not be detectable in a proteome map.

JHP1350 (spot number 128)

This protein is homologous to SpaP a member of the antigen I/II family of collagen-binding surface adhesion proteins. This is a poorly characterized category whose general function can only be predicted.

The antigen I/II proteins are a family of related surface polypeptides that have been largely examined in species of oral streptococci. These studies found that members of the antigen I/II family are highly conserved adhesions with approximately 65 to 70% primary sequence similarity (Demuth and Irvine, 2002). In oral bacteria these proteins initiate and develop a biofilm by mediating interactions of oral streptococci, other oral bacteria, cell matrix proteins (e.g. type I collagen) and salivary components (Demuth and Irvine, 2002). These proteins have also been shown to interact with eukaryotic collagen and human fibronectin and laminin (Baddour, 1994). The interactions of antigen I/II are thought to contribute to the formation of streptococcal abscesses. *H. pylori* has been shown to bind to connective tissues collagen I, II IV, laminin, fibronectin and vitronectin (Trust et al, 1991). *H. pylori* binds strongest to laminin and collagen type IV but only weakly to collagen type I, II, fibronectin and vitronectin (Trust et al, 1991). The ability of *H. pylori* to bind to components of the basement membrane would represent an important virulence characteristic for this organism. There have not been any studies to determine what

effect the absence of the collagen type I binding adhesion would have on the colonisation and virulence of the organism. Given that *H. pylori* binds weakly to collagen type I (Trust et al 1991) it is possible that loss of the JHP1350 a homologue to SpaP might not have any affect on the ability of the organism to colonise the gastric mucosa. Alternatively the extracellular matrix components of connective tissue, for example collagen type I, may play a role in the formation of duodenal ulcers. The expression of collagen binding adhesions might be a reflection of duodenal ulcer formation. For example, if DU isolates prefer binding to collagen type II or IV in the duodenum then adhesions for this type of connective tissue would be expressed while collagen type I binding adhesions would down-regulated or cease expression. Further examination would be required to determine the actual function of JHP1350 and its possible role in collagen type I binding.

JHP0295 (spot number 129)

This is predicted to be a xylanase/chitin deacetylase protein belonging to the metabolism category and is involved in carbohydrate transport and metabolism.

Xylan is a major hemicellulose component of the plant cell wall. Small amounts of plant fibres are degraded by microbes in humans. Microbial digestion of plant materials is significant and occurs mainly in the caecum and large intestine after prior host digestion in the stomach and small intestine (Hespell and Whitehead, 1990).

Xylanases are secreted into the media that contains pure xylan or xylan-rich residues. They are usually inducible enzymes (Balakrishnan et al, 1997) but some research studies have also observed constitutive production of xylanases (Hespell and Whitehead, 1990; Khasin et al. 1993; Lindner et al, 1994). Because xylan cannot enter the microorganism, the expression of xylanase is stimulated by low molecular mass xylan fragments. These low molecular mass fragments include xylose, xylobiose, xylooligosaccharides, heterodisaccharides of xylose and glucose and play a key role in the regulation of xylanase biosynthesis (Kulkarni et al, 1999). The fragments are released from xylan by a small amount of constitutively produced xylanase (Bastawde, 1992; Kulkarni et al, 1999). The component that stimulates

induction differs depending on the microorganism. For example, in *Trametes trogii* it is mostly xylan (Levin and Forschiassin, 1998), whereas in *Trichosporon cutaneum* (Liu et al, 1998), xylanase is induced by xylan and is repressed in the presence of glucose. In *T. trogii* (Levin and Forschiassin 1998) and *Staphylococcus* sp. SG-13 (Gupta et al, 1999) xylanase production is enhanced in the presence of amino acids. Gupta et al (2000) reported an improved xylanase production by *Staphylococcus* sp. SG-13 in a medium containing wheat bran. The regulation of xylanase secretion by microorganisms is still not completely understood. It would not be surprising that dietary components would induce xylanase in *H. pylori* but the significance of why only the DU isolates were found to be producing it is not known.

Spot number 130

This protein has no known function.

4.10.3 Unique and Absent Proteins in NUD Strains

In this study, eleven disease specific NUD proteins were identified. These consisted of seven unique (Figure 3.11 and Table 3.13) and four absent (Figure 3.10 and Table 3.14).

4.10.3.1 Unique Proteins

All the unique proteins, with the exception of fructose-biphosphate aldolase, that were identified in the NUD disease group have no known function or are poorly characterised.

HSP12 variant C

This protein has no known function.

The *hsp12* gene present in the *H. pylori* isolates NUD1, NUD2 and NUD3 is a member of the gene family 12 and its transcription is induced after growth under stress conditions of 42^{o}C, iron restriction and low pH (de Vries et al, 2003). Other proposed functions include penicillin binding, a weak β-lactamase activity and the maintenance of *H. pylori*'s spiral shape (Krishnamurthy et al 1999). However, there is currently no clear function for this family of proteins. Due to the stress activation

of *hsp12*, it was proposed that this gene plays a role in stress survival. One belief is that Fur, a regulator of iron uptake, may regulate the expression of *hsp12* under conditions of iron starvation (de Vries et al, 2003). However, there was no evidence that expression of *hsp12* affected growth under normal or stress conditions.

The *hsp12* locus was found to be not only strain specific but also *H. pylori* specific (de Vries et al, 2003). There is no precise homologue of the *hsp12* gene present in the genome sequences of *H. pylori* strains that have been sequenced. For example the 26695 and J99 strains contain *hsp12* variant A and B respectively (Alm et al, 1999; Tomb et al, 1997). Other variants found in strains include variant C and the alleles I, II, III, IV and V that contain unknown genes (Chanto et al, 2000). The high diversity of the *hsp12* gene is most likely due to recombination and may require the presence of *H. pylori* strains. It has been theorised that the diversity of the *hsp12* gene is the result of adaptation to the host environment and plays a role in the persistent colonisation of *H. pylori* to its specific host (de Vries et al, 2003). The de Vries et al (2003) study suggested that these genes might interact with the host environment by stress regulation and genetic variation. However, there is still no evidence that the expression of the *hsp12* gene is associated with any particular disease outcome.

HP1064, HP0864, JHP1044, and HP0484 (spot numbers 111, 113, 114, and 116).
These are proteins that have no known function.

JHP1163 and JHP1073 (spot numbers 112 and 115)
These proteins are poorly characterised.

4.10.3.2 Absent Proteins

HP1562 (spot number 131)
This is an Iron (III) ABC transporter, periplasmic iron-binding protein (CeuE) and is involved in inorganic ion transport and metabolism and belongs to the metabolism category.

There are two environmental conditions that may affect the expression of CeuE. These are iron starvation/iron excess and nickel excess.

The mechanisms by which *H. pylori* acquires Fe^{2+} and Fe^{3+} from the host are still poorly defined. There are a number of outer membrane proteins, which may serve as receptors for haeme (Worst et al, 1995). Human lactoferrin can also be used as a sole Fe source (Husson et al, 1993) and a putative lactoferrin binding protein has been identified (Dhaenens et al, 1997). In conditions of iron starvation Fur, an iron regulator, is able to prevent the expression of some iron transport proteins.

H. pylori possesses the FeoB transporter, which has an important role in the acquisition of Fe^{2+}. Other transport proteins involved in the transport of Fe across the membrane include CeuE and FrpB. CeuE, a homolog of the iron (III) ABC transporter, periplasmic iron-binding protein (Tomb et al, 1997), may function in the shuttling of Fe^{3+} across the periplasmic space (Velayudhan et al, 2000). The FrpB protein is a homolog of the Fe limitation-inducible outer membrane protein of *Neisseria meningitidis* (27.6% amino acid identity and 49.5% similarity) that belongs to the family of TonB-dependent receptors (Tomb et al, 1997).

Iron concentrations have both direct and indirect effects on the expression of the iron network genes (van Vliet et al, 2002) Proteins that respond to variations in the iron levels include KatA (Harris et al, 2002), Pfr (Delany et al, 2001a), iron transport proteins (e.g. FrpB and CeuE) (Delany et al, 2001b; van Vliet et al, 2001), Fur (Ernst et al, 2005) and Hsp12 variant C (de Vries et al, 2003). In iron abundant environments the expression of iron transport proteins of *H. pylori* is decreased. Conversely, under iron restricted conditions the expression of iron transport proteins including CeuE is increased (van Vliet et al, 2002). This differs from other bacteria where the iron transport proteins are down regulated.

A constant supply of nickel ions is required for the synthesis and activity of a number of metalloenzymes that play a crucial role in prokaryotic and eukaryotic cells (Eitinger and Mandrand-Berthelot, 2000). If nickel ions accumulate, they will then inhibit microbial cell growth and exhibit a toxic effect. Therefore, the level of

nickel ions needs to be tightly regulated by controlling its transport and/or by storing it in protein complexes. Transport proteins include the NixA (HP1077) and AbcCD (HP1576/1577). A nickel storage protein is believed to be HP1427 (Mobley, 1999). There are a number of transition metal pumps that are believed to play a role in the intracellular ion composition (Melchers et al, 1998). Nickel excess affects Pfr, Fur, FrpB4, ExbB/ExbD, CeuE, CheV, FlaA, FlaB, HrcA, GrpE, DnaK, Omp11, Omp31 and Omp32 by repressing these protein products (Contreras et al 2003). Nickel excess also upregulates UreA, UreB, (van Vliet et al., 2002), NixA and CopA2 (nickel transport proteins) and Hpn (nickel storage) (Contreras et al, 2003).

JHP0671 (spot number 133)

This is 2-methylthioadenine synthetase (MiaB) and is a translation protein belonging to the information processing category.

MiaB (a homologue of 2-methylthioadenine synthetase in *H. pylori*), in *E. coli,* is involved in the methylthiolation of the adenosine 37 residue modification of tRNAs specific for codons beginning with U, except $tRNA^{I,V}$ Ser (Esberg et al, 1999; Pierrel et al, 2002). MiaB, a Fe-S enzyme, contains a cysteine cluster reminiscent of iron-binding sites. MiaB is a monomeric iron-sulphur protein that is able to assemble into three possible clusters 4Fe-4S, 3Fe-4S and 2Fe-2S (Pierrel et al, 2002). This protein is believed to be a member of a superfamily of proteins that uses the combination of a Fe-S cluster and AdoMet to initiate radical catalysis whose iron centre may be essential for activity (Pierrel et al, 2002). On the other hand, the *miaB* gene has been shown to encode the thiotransferase required in tRNA modification, the first step of the pathway. This suggested that the MiaB protein is involved in C–S bond formation (Esberg et al, 1999). Deficiency in methylthiolation leads to a decreased efficiency of the corresponding tRNAs and an increased spontaneous mutation frequency (Connolly and Winkler, 1989). The increase in spontaneous mutation may be an advantage to the organism under conditions of environmental stress. Examinations of this protein have been conducted on organisms such as *E. coli* (Pierrel et al, 2002) and *Salmonella typhimurium* (Buck et al, 1982) but its precise function is still not known in *H. pylori*. Under conditions of iron starvation, this protein might be down regulated, resulting in it not being detected on 2-DE gels. If

spontaneous mutation has occurred this may give the NUD isolates an advantage under iron restricted conditions.

HP0305 and HP1285 (spot numbers 132 and 134 respectively)

These two absent proteins have no known function.

4.10.4 Unique and Absent proteins in GI Isolates

4.10.4.1 Unique Proteins

HP0538 (spot number 118)

This protein is Cag17/CagN expressed by the gene locus located in the cag pathogenicity island and is involved in pathogenesis.

The cag pathogenicity island (PAI) is a 40-kb insertion that is a complex structure of many different genes involved in virulence. The *cag17*/*cagN* gene is one of many genes located on the cag pathogenicity island; others include *cagA*, *cagS* and *cagT* (Censini et al, 1996; Tomb et al, 1997). Damage to any of a number of genes in the pathogenicity island has the potential to disrupt other virulence genes of *H. pylori* (Figura et al, 2004). It has been found in certain cases that disruption of the *cagA* gene affects the expression of both flagellin A and flagellin B (FlaA and FlaB) and may affect post-transcriptional events (Figura et al, 2004). These two elements of flagella are essential components for *H. pylori* motility which is necessary for it to colonise the gastric mucosa and are crucial for persistent infection (Ottemann and Lowenthal, 2002).

Variability among the genes contained on the cag pathogenicity island may determine the degree of virulence and disease outcome of a particular strain. For example, strains lacking both *cagE* and *cagT*, were more frequently found in patients with chronic gastritis (Ikenoue et al, 2001). The presence of CagN in the GI isolates, may be characteristic of GI isolates and may be related to the type of disease they cause. CagN may affect the expression of other proteins.

NA (spot number 117) and HP0842 (Spot number 119)

These two proteins have no known function.

4.10.4.2 Absent Proteins

There were four specifically absent proteins (Figure 3.11, Table 3.13) that were identified.

HP0431 (spot number 136)

HP0431 is an uncategorised protein. Its possible function may be a protein phosphatase 2C homologue (Ptc1). Reversible protein phosphorylation plays a central role in a variety of cellular mechanisms, including the control of metabolism, the cell cycle, cell proliferation and differentiation (Tatsuya et al, 1993). Tyrosine phosphorylation is affected by the activities of both protein tyrosine kinases and protein tyrosine phosphatases (PTPases). The PTPase family is extremely diverse and possibly consists of proteins with more than one role (Charbonneau and Tonks, 1992). PTP1 and PTP2, similar to PTPases in eukaryotes, are found in bacteria, either individually, or together, with no known function (Tatsuya et al, 1993). Mutation of the Ptc1 function in *Saccharomyces cerevisiae* was found to be nonessential, however, *ptc1* and *ptp2,* double mutants, showed a marked growth or germination defect at 37^{o}C (Tatsuya et al, 1993). The phosphate activity of Ptc1 is dependent on the presence of the divalent cations, Mg^{2+} or Mn^{2+}, and it hydrolyses a serine phosphorylated substrate in their presence. Therefore, the absence of Ptc1 would have no effect on the function of this protein. However, if there was an additional loss of the *ptp2* gene this would have an effect (Tatsuya et al, 1993). However, the absence of the *ptp2* gene product was not detected as an absent protein for the GI disease group. Therefore, it is possible that the lack of the Ptc1 protein might not have any affect on the cell cycle, cell proliferation and differentiation of *H. pylori*. The function of Ptc1 has not yet been determined for *H. pylori* and further studies would be needed to determine the effect of the protein's absence on the organism's growth. Additionally insufficient Mg^{2+} or Mn^{2+} might reduce the activity and expression of this protein. This might be possible if the expression levels were insufficient for detection by the techniques utilised in this study.

Predicted GTPases (dynamin-related) (spot number 138)

This is a poorly characterized protein where the general function is predicted only.

Numerous isoforms and homologues of dynamin related GTPases have been found in a variety of eukaryotic organisms from yeasts to humans (e.g. humans, *S. cerevisiae*, *Trypanosoma brucei* and Arabidopsis) (Obar et al, 1990; Rothman et al., 1990; Dombrowski and Raikhel, 1995; Gammie et al, 1995; Gu and Verma, 1996). A characteristic feature of the dynamin family is the conserved N-terminal GTPase domain (Bourne et al, 1990). However, the rest of the polypeptide sequence is less well conserved. It has a variety of different functions in each organism. In yeast, it is involved in the maintenance of mitochondrial morphology (Jin et al, 2003). In humans, dynamin is a large molecular mass (100 kDa), GTP-binding protein, consisting of three isoforms dynamin-1, dynamin-2 and dynamin-3 (Suzuki et al, 2001). Dynamin has also been shown to function in apical transport of intracellular vesicles (Kreitzer et al, 2000). Thus, dynamin is currently thought to catalyse many essential steps in vesicle formation and traffic (Kirchhausen, 1998). It was found that dynamin may play a role in intracellular vesicle formation and the transport of VacA (*H. pylori* vacuolating cytotoxin) (Rothman et al, 1990; Suzuki et al, 2001). The possible absence of GTPase (dynamin-related) may imply that VacA is not present in the GI isolates included in this study. It has been reported in previous studies that CagA and VacA are two important virulence proteins that may increase the severity of the disease outcome (Graham and Yamaoka, 2000). However, both proteins are not necessarily expressed in all isolates.

Other Absent Proteins

The two other absent proteins HP0934 (spot number 135) and JHP0896 (spot number 137) have no known function.

4.10.5 GU Differentially Regulated Proteins

Only four GU differentially regulated proteins (Figure 3.12, Table 3.15) were of sufficient abundance to enable identification, these were:

HP0958 (spot number 139)

A down-regulated Zn-ribbon protein, possibly nucleic acid-binding with no known function.

This study found that this protein is part of a serial-charge train (see page 173). The serial-charge train in NUD, GI and DU isolates consisted of two protein spots. The first protein spot in the charge train, number 120, was identified as an absent spot in the GU isolates. The second protein, spot number 139, had a greater than 3-fold reduction in the GU isolates. Therefore, given that this protein in the charge train has a decreased spot intensity the absence of protein spot number 120 is most likely because it has a decreased expression.

HP0824 (spot numbers 157 and 158)

An upregulated thioredoxin protein that belongs to the cellular processing and signalling category and is involved in posttranslational modification, protein turnover and chaperones.

In this study, TrxB (HP0825), with a 2-DE gel location of 33.5 kDa/pI 5.8, was identified as a specifically absent protein in the GU isolates. However, its counterpart TrxA (HP0824) in the thiol-dependent, reduction-oxidation system was present as an up-regulated protein for the GU isolates. While the two explanations mentioned on page 174 are both possible the most likely explanation for differences between the GU isolates and other isolates is a response to oxidative stress.

HP0153 (spot number 161)

This is a down regulated recombinase A (RecA) belonging to the information processing category and is involved in replication, recombination and repair.

H. pylori strains display a high rate of mutation, but the diversity of this organism is additionally increased by horizontal gene transfer and recombination between strains. However, the rate of infections with multiple strains is not very common. Since *H. pylori* is naturally competent for transformation, horizontal gene transfer is thought to occur mainly by this mechanism (Suerbaum et al, 1998). Most *H. pylori* strains

are naturally competent for DNA uptake therefore RecA recombination may be essential for DNA transformation. Homologous recombination between *H. pylori* strains not only plays a role in genomic variability through uptake of DNA from neighbouring cells but also is involved in maintaining genome integrity (Matic et al, 1995).

The ability to manage environmental stress is very important for the survival of *H. pylori*. It is exposed to many different types of stress when it infects the stomach e.g. low pH and oxidative stress. Some types of stress could result in DNA damage that might induce a repair response. The RecA protein plays a role in recombination and is a regulatory protein that responds to certain types of DNA damage and mediates derepression of the SOS regulon (Walker, 1984). The net affects of SOS induction are many and include increased DNA repair, an increased rate of mutation, inhibition of cell division, and prophage induction (Walker, 1984). The SOS response is important in the repair of DNA damage caused by exposure to reactive oxygen species produced by immune cells during bacterial infection. Thompson and Blaser (1996) found preliminary evidence, that *recA* mutants were more susceptible to low pH and antibiotics. This suggests that RecA may be important for survival of *H. pylori* during infection of the gastric mucosa.

HP1520 (spot number 152) (shared with NUD)

This protein has no known function and is down regulated.

JHP0065 (spot number 162)

This protein belongs to the cellular processes and signalling category and is involved in posttranslational modification, protein turnover, and is a chaperone protein.

4.10.6 DU Differentially Regulated Proteins

The DU disease group had the largest proportion of proteins that were down regulated in comparison to the other disease groups (see Figure 3.12 and Table 3.15). See section 4.9.7 for proteins that were down regulated in both DU and NUD isolates.

HP0197 (spot number 143)

This is S-adenosylmethionine Synthetase (MetK) belonging to the metabolism category and is involved in coenzyme transport and metabolism.

It is uncertain what the precise role of a decrease in MetK production might be in the physiology of DU isolates as its function has only been inferred from studies in other organisms. MetK is an enzyme which donates methyl groups to DNA and many other compounds (Wei and Newman, 2002; Cantoni, 1951). S-Adenosylmethionine (SAM) is synthesized in all cells from ATP and methionine and regulates cell growth, differentiation, and function (Mato et al, 1997).

A study of *E. coli* found that when *metK* expression was limited, methylation of DNA decreased and cell division was hampered (Wei and Newman, 2002). In section 4.10.2.1 another protein, HP0332 also has a role in cell division. In organism, cell division de Boer et al (1992) found that when the topological specificity factor MinE was overexpressed that cell division was suppressed. It is possible that MinE is not a specifically unique protein but instead it might have been overexpressed. The decrease in MetK in conjunction with a possible overexpression of MinE implies that DU isolates may have a lowered rate of cell division. Alternatively, it has been suggested that MetK plays a role in oxidative stress. For example, in eukaryotes *in vivo* study on the antioxidant properties of MetK in Fe^{2+} oxidation it was found that this protein exerts direct antioxidant action, mainly through iron chelation and inhibition of Fe^{2+} autoxidation (Caro and Cederbaum, 2004).

HP0201 (spot number 144)

This is a fatty acid/phospholipid biosynthesis enzyme (PlsX) involved in lipid transport and metabolism and was found to have a greater than three-fold decrease in DU isolates.

In *E. coli,* half of the fatty acid biosynthesis genes (*fab*) are clustered together. The *plsX* gene which is located upstream of *fabH* is considered to be part of this cluster due to its role in phospholipid biosynthesis (Larson et al, 1984; Zhang and Cronan,

1998). The *fab* cluster genes and *plsX* have been reported in *H. pylori* although these genes are arranged differently. While fatty acid biosynthesis has attracted some interest in other organisms such as *Bacillus subtilis* and *E. coli* (Zhang and Cronan, 1998; Larson et al, 1984) little has been done in *H. pylori.* The function of PlsX is still poorly understood but is believed to be one of two enzymes, which are involved in the pathway to produce phosphatidic acid, a key phospholipid synthesis intermediate (Rock and Cronan, 1982).

Studies involving the *E. coli plsX* gene and *fab* cluster have indicated that *rpmF* and *plsX* are cotranscribed and as such could play an important role in coordinating ribosome synthesis with cell-membrane synthesis (Podkovyrov and Larson, 1995). Since PlsX spot intensity in the DU isolates is decreased in comparison to the other isolates this may reflect a decreased synthesis of the protein homologous to RpmF in *H. pylori* as well. If the co-transcription of these proteins is important, then this decrease in production may reflect differences in cell membrane synthesis of DU isolates.

HP0875 (spot numbers 140-142)

This protein is a catalase (KatA) and is involved in detoxification.

The KatA can be altered in response to oxidative stress and an iron restricted environment. KatA is an enzyme that is responsible for the dismutation of hydrogen peroxide into water and molecular oxygen and protects *H. pylori* from the potentially damaging effects of hydrogen peroxide (Harris et al, 2002).

H. pylori damage of the gastric mucosa is the result of the secretion of the neutrophil activating protein and CagA which attract neutrophils and interleukins in response to infection. In activated neutrophils, NADPH oxidase in cell membranes becomes activated, and electron transfer takes place from NADPH within the cells to oxygen both inside and outside. The oxygen molecules that receive an electron become superoxide radicals that are rapidly converted to hydrogen peroxide (either by spontaneous dismutation or enzymatically by the action of superoxide dismutase)

and hydroxyl radicals that are formed non-enzymatically in the presence of ferrous ions as a secondary reaction (Cadenas, 1989; Farr et al, 1988).

KatA is expressed in high levels in response to oxidative stress *in vivo*. However, in an iron limited environment KatA expression is reduced, whereas normal iron levels promote KatA (Harris et al, 2002). The Harris et al (2002) study found that *H. pylori* cells with the knockout mutation in the *fur* gene, or grown on a low-iron medium, showed a reduction in catalase activity and were more sensitive to hydrogen peroxide. However, in the present study, all isolates were grown on the same medium that was not deficient in iron and none of the proteins that were selected for peptide mass fingerprinting were identified as the Fur protein. This indicates that some other mechanism in the DU isolates might be resulting in a down-regulation of KatA expression. However it is possible that Fur might have been one of the other differentially regulated proteins that were not identified therefore this protein might still be a factor in reduced catalase expression. Other proteins were identified that also respond to iron levels and oxidative stress i.e. CeuE, Pfr and KatA. Further examination and identification of the other unidentified differentially expressed proteins would need to be analysed to determine if other proteins such as Fur might also have been decreased/increased in conjunction with KatA. The down-regulation of KatA may be due to adaptation to a specific niche and may provide an advantage for survival under specific conditions e.g. oxidative stress and iron-starvation.

HP0027 (spot numbers 147 and 148)

This protein is isocitrate dehydrogenase (Icd). It is in the tricarboxylic acid (TCA) cycle and is involved in energy production and conversion.

The TCA cycle performs a dual role in cell metabolism providing biosynthetic starting compounds such as α-ketoglutarate, succinyl-CoA and oxaloacetate, which act as precursors for a wide spectrum of cell components, as well as a 'metabolic energy' source through the generation of reduced nucleotides, the reoxidation of which may be coupled to ATP synthesis. Initially it was believed that *H. pylori* had an incomplete TCA cycle (Tomb et al, 1997) but subsequent studies have shown that *H. pylori* has a complete TCA cycle (Pitson et al, 1999; Kather et al, 2000).

There are three forms of Icd in eukaryotics; mitochondrial NADP-dependent, mitochondrial NAD^+-dependent, and cytosolic NADP-dependent (ICDH) (Maeng, 2004). *H. pylori* only has the ICDH form, which is located primarily (96%) in the cytosolic fraction and was NADP specific with no detectable activity with NAD^+ (Pitson, 1999). The activity of ICDH in *H. pylori* was found to be inhibited at higher concentrations of both $NADP^+$ and isocitrate (Pitson, 1999).

The present study found that the isocitrate dehydrogenase Icd of the GU isolates displayed a greater than three-fold increase in spot volume in comparison to the DU isolates. Icd catalyses the oxidative decarboxylation of isocitrate to α-ketoglutarate. The Icd family of enzymes, either the NAD^+-dependent, or NADP-dependent form, exists in virtually all species and has a wide variety of functions, including the TCA cycle and is responsible for ATP production, and other biosynthesis pathways (Holms, 1987). One study of *H. pylori* found that in the presence of metronidazole, Icd was significantly decreased in activity (Hoffman et al, 1996). Heat shock may increase oxidative stress by increasing the production of reactive oxygen species and/or the promotion of cellular oxidation events. An examination into the effects of heat shock induced oxidative stress in *E. coli* found that ICDH was an important antioxidant enzyme protecting against heat shock (Choi et al, 2003).

In eukaryotic organisms, it has been reported that the overexpression of ICDH plays an essential role in protecting against oxidative damage by H_2O_2 (Maeng, 2004). In addition, it was found that it greatly reduced intracellular peroxide levels and cell death in response to H_2O_2. However, the overexpression of ICDH did not affect the expression levels of the other antioxidant enzymes (Maeng, 2004). In another study of the effects of temperature on the development of wild-type *Caenorhabditis elegans*, it was found that Icd was only detected at lower temperatures (Madi et al, 2003).

To date (December 2005) there are no published studies that have examined the effects of oxidative stress on Icd expression in *H. pylori*. However, given that other anti-oxidative stress proteins (e.g. catalase and non-haeme iron (III) ferritin) were

under-expressed in the DU isolates (by a factor greater than three) differences due to oxidative stress cannot be ruled out.

4.10.7 DU and NUD Differentially Regulated Proteins

Five proteins were differentially expressed by isolates from both NUD and DU disease groups. Two proteins had no known function, JHP0052 (spot number 156) and HP0453 (spot number 153). The remaining three proteins had a known function (Figure 3.12, Table 3.15) in both the NUD and DU disease groups.

When similar quantitative differences are observed for two disease groups the possible explanation for these differences might not be the same for both disease groups. Different factors or other proteins of unknown function may determine the most likely reason for the up-regulation and/or down-regulation of specific proteins. For example non-haeme iron containing ferritin (Pfr) (see page 196), which has been extensively studied for *H. pylori,* is involved in stress, iron and nickel excess and iron starvation. While down-regulation might be caused by oxidative stress in DU strains this might not be the case for NUD and there might be possible alternative explanation. Further investigation would be required in these cases to determine the most likely cause of the quantitative difference. See Section 4.9.10 for further discussion on possible explanations on the down-regulation of Pfr in DU and NUD strains.

JHP0761 (spot number 151)

This is homoserine dehydrogenase and it was down regulated in both the NUD and DU isolates (Hom). Hom belongs to the metabolism category and is involved in category amino acid transport.

The enzymes involved in the synthesis of essential amino acids have been extensively studied in plants and microorganisms. Hom is a component of the aspartate pathway that catalyses the NADPH-dependent reduction of aspartate to produce homoserine (Viola, 2000). This reaction occurs at a key branch point in the pathway, with the substrate, homoserine, serving as the precursor for the biosynthesis of lysine which leads to the synthesis of methionine, threonine, and isoleucine

(Viola, 2000). Regulation of the reaction catalysed by homoserine dehydrogenase (HDH) is a catalytic domain in a bifunctional enzyme with aspartokinase, and allosteric regulation of both activities is mediated through the aspartokinase domain (Truffa-Bachi et al, 1974).

Organisms that utilise the aspartate biosynthesis pathway contain several aspartokinases that catalyse the initial step in the pathway, the phosphorylation of L-aspartic acid (Viola, 2000). In many of these organisms, at least one of these enzymes is bifunctional, catalysing both the first, and surprisingly, the third reaction in this metabolic sequence (Cohen 1969). Many bacteria contain several isoforms of aspartokinases, some of which are bifunctional. These bifunctional enzymes are subject to differential regulation, both by feedback inhibition by the amino acids produced and by repression at the genetic level. In *E. coli,* the aspartokinase-HDH II enzyme has a feedback mechanism that is inhibited by L-threonine and possibly S-adenylmethionine levels (Viola, 2000). The aspartate pathway response to different conditions, such as oxidative stress and iron excess/restriction, has not yet been reported for *H. pylori*.

HP0653 (spot number 146)

This is a non-haeme iron containing ferritin (Pfr) and is down regulated in NUD and DU. It is involved in inorganic ion transport and metabolism. The protein Pfr is known to respond to various stimuli - oxidative stress, iron depletion and nickel excess.

While Fe^{2+} is an essential property required for most life forms, iron can be extremely toxic under aerobic conditions. Bacteria need to maintain a delicate balance in order to scavenge sufficient quantities of Fe^{2+} or Fe^{3+} without them becoming toxic. Bacteria manage iron balance by utilising five strategies (Andrews et al, 2003). The first involves the utilisation of high-affinity Fe^{3+} or Fe^{2+} transport proteins in order to scavenge iron, in various forms, from the surroundings (Delany et al, 2001b; van Vliet et al, 2001). Bacteria also control Fe^{2+} or Fe^{3+} consumption under iron-restricting conditions by down regulating the expression of iron-containing proteins. Bacteria also deposit intracellular Fe^{2+} stores to provide a

source of iron that can be drawn upon when external supplies are limited e.g. ferritins (Frazier et al, 1993; Evans et al, 1995). Also they employ redox stress-resistance systems such as superoxide dismutase (e.g. degradation of iron-induced reactive oxygen species and repair of redox stress-induced damage) to protect the microorganism from oxidative stress (Naito and Yoshikawa, 2002). Finally all bacteria utilise an iron-responsive regulatory system that co-ordinates the expression of the iron homeostatic machinery according to iron availability (Ernst et al, 2005).

The reduction products of oxygen superoxide and hydrogen peroxide are only mildly reactive physiologically. However, iron interacts with these molecules to generate the more damaging hydroxyl radical. In *H. pylori,* mechanisms for protecting against oxidative stress include the enzymes superoxide dismutase, alkyl hydroxide peroxidase and catalase (Naito and Yoshikawa, 2002; Harris et al, 2002). These enzyme systems interact to protect cells from reactive oxygen species. These enzymes are regulated in *H. pylori* by the protein Fur that responds to stimulants (e.g. Fe^{2+} levels) in the environment by regulating the oxidative-stress proteins and the iron storage protein Pfr (Bereswill et al, 1998a Bereswill et al, 1998b; Waidner et al, 2002).

In the human stomach iron is released from food by peptic degradation, while iron restriction can be encountered via the iron-chelating activity of lactoferrin (Nakao et al, 1997). Therefore, both iron starvation and iron overload may occur in relatively short time intervals. The gastric environment requires the development of regulatory systems that allow *H. pylori* to regulate its iron metabolism.

Pfr functions are essential for gastric adaptation and iron storage is necessary for the successful development of *H. pylori* infection (Waidner et al, 2002).

In *H. pylori* Pfr has been characterised as the protein involved in the storage of iron (Frazier et al, 1993; Evans et al, 1995). Studies on the function of Pfr indicate that it also plays a substantial role in the resistance of *H. pylori* to metal toxicity (Harrison and Arioso, 1996) and oxidative stress. Fe^{2+} interacts with superoxide and hydrogen peroxide species to generate the highly reactive and extremely damaging hydroxyl

radical (Andrews et al, 2003). The organism at the same time needs to protect itself from hydroxyl radicals as a result of the rapid alteration of iron levels in its environment. Ferritins catalyse a function that is the exact opposite to that of iron uptake systems, which increase cytoplasmic iron concentration. In an iron-rich environment, Pfr is up regulated and the iron transporter proteins are down regulated by Fur (Delany et al 2001a). It is believed that ferritins may protect the organism from other metals (nickel, copper, manganese and zinc) (Contreras et al 2003). Evidence of an interaction of ferritin with iron-dependent regulation mechanisms has also been obtained (Bereswill et al, 1998a; Bereswill et al, 1998b).

It has also been theorised that there may be other possible roles for ferritin other than iron storage, metal toxicity and oxidative stress. What the other possible functions of ferritin might be is still unknown. However, if there are other roles for ferritin they might play a factor in the disease outcome of *H. pylori*. See section 4.9.10 for further discussion on possible explanations on the down-regulation of Pfr in DU and NUD strains

JHP0650 (spot number 150)

This is the competence protein ComGF which is down regulated in DU and NUD isolates. It belongs to the cellular processes and signalling category and is involved in intracellular trafficking and secretion.

The ComG operon in *B. subtilis* encodes seven proteins (ComGC, GG, GD, GE, GF, GB and GA) that are all essential for the binding of transforming DNA to the competent cell surface (Chung et al, 1998). All of the proteins are membrane associated. ComGF behaves as an integral membrane protein while ComGA, a putative ATPase, is located on the inner face of the membrane as a peripheral membrane protein (Chung et al, 1998). The ComGF is a small protein with a predicted single membrane-spanning segment near its N-terminus (Chung and Dubnau, 1998). It does not appear to possess a processing site, has no known orthologue and its role is unknown, although it is required for DNA binding to the competent cell surface.

Other Proteins

Other differentially regulated proteins identified were HP0453 (spot number 153) and JHP0052 (spot number 156). JHP0052 has not been categorised, but the NCBI record shows that this protein has been inferred to have a biological process involved in the regulation of transcription that is DNA dependent.

4.10.8 NUD Differentially Regulated Proteins

Four proteins with no known function were identified that were differentially regulated (Figure 3.12, Table 3.15) for the NUD isolates these are:

- HP0879 (spot number 149). Uncharacterised and with no known function.
- HP0453 (spot number 153). Function unknown. It differed in spot intensity in the DU isolates.
- HP1520 (spot number 152). Function unknown.
- JHP0052 (spot number 156). It is related to proteins in transcription activity but has no known function. It also differed in spot intensity in the DU isolates.

4.10.9 Unidentified Proteins

Approximately 50 proteins remain unidentified in this study. The majority of these proteins were small M_r proteins, which would be expected to have only a few digestion cleavage sites. This makes it difficult to identify them as it would generate a large number of possible matches in the databases. A solution that may need to be considered in the future is to re-analyse these proteins using a technique such as MALDI-TOF Post Source Decay or ESI-MS. A partial amino acid sequence might then be utilised to acquire a more accurate match in the databases. Other protein identification technologies that can be utilised are liquid chromatography/mass spectrometry (LC-MS/MS) (Mann and Wilm, 1994; Yates et al., 1995) mass spectrometry sequencing (PMF+MS/MS) or Tandem mass spectrometry (MALDI TOF-TOF).

Also very little homology may have been found with proteins in the database because the unidentified proteins investigated in this study may not actually have been

entered into the protein database. Therefore these proteins would not be identified even if another technique was used. Therefore a peptide mass fingerprint that has a "no hit" result needs to be rechecked later when the database has been updated.

4.10.10 Stress response, cell division, cell membrane and other proteins

Spots selected for identification in this study included proteins

whose expression is altered in response to:

- Oxidative stress - KatA, Pfr, Icd, MetK, TrxA/TrxB
- DNA damage - RecA
- Iron restricted stress – KatA, Pfr, CeuE, Hsp12 variant C‡
- Nickel Excess stress – Pfr and CeuE
- Temperature – Hsp12 variant C

that affect:

- Cell Division – MetK and MinE
- Cell membrane - PlsX, Ptc1, ComGF, JHP0140, JHP0896 and SpaA

that have other roles:

- Digestion of plant fibre - JHP0295
- Hom – Aspartate pathway
- Ribosome binding to RpoN - HP0958
- Pathogenicity – Cag17/N

Many of the proteins identified in this study have more than one role or alternative roles. For example Pfr (see above), which has been extensively studied for *H. pylori,* is involved in stress, iron and nickel excess and iron starvation. Whereas another protein such as MetK, poorly understood in *H. pylori*, has been proposed to have different roles dependent on the organism studied, oxidative stress (humans) and cell division (*E. coli*). A comparison of the eleven *H. pylori* isolates studied identified eight proteins that have a role in protection of the organism from stress i.e. oxidative stress, iron-restriction, and temperature (KatA, Pfr, MetK, TrxA/TrxB, Hsp12 variant C, CeuE and RecA). Another eight proteins are associated with cell division (MetK and MinE) and the cell membrane (PlsX, Ptc1, ComGF, SpaA, JHP0140 and JHP0896) but have no apparent connection to stress. It is possible that stress may

indirectly affect these particular aspects of *H. pylori's* morphology. Figure 4.1 shows a representation of the possible connections between proteins involved in these particular aspects.

Oxidative stress

The induction of an inflammatory response by *H. pylori* infection leads to an increased potential for oxidative damage to the bacterium. KatA, Pfr, Icd, MetK, (involved in metabolism) and TrxA/TrxB (cellular processes and signalling) have been associated with protection from oxidative stress. Additionally other proteins unrelated to those involved in protection from oxidative stress may respond to damage caused by oxidative stress, e.g. DNA damage (Figure 4.1). KatA, essential for protection against oxidative stress, has been shown to spontaneously mutate and cease expression in response to a change in its environment (e.g. *in vivo* vs *in vitro*) (Manos et al, 1998). The protein Pfr is known to respond to various stimuli - oxidative stress, iron depletion and nickel excess (Waidner et al, 2002; Contreras et al, 2003). However, *icd* expression has only been found to respond to oxidative stress in *E. coli* (Lee et al, 1999) and humans with no similar event currently (up to 2005) known to occur in *H. pylori*. The TrxA/TrxB redox system is known to increase thioredoxin and decrease thioredoxin reductase production in response to oxidative stress (Berggren et al, 1996; Reddy et al, 1999). MetK is believed to be important to cell division in *E. coli* and oxidative stress in eukaryotic organisms (Wei and Newman, 2002; Caro and Cederbaum, 2004). Although no such association has been found in *H. pylori* it is possible that MetK could be linked with the other proteins that respond to oxidative stress. Oxidative stress will cause DNA damage, which the organism might respond to by increasing RecA production (Walker, 1984).

A comparison of the isolates revealed that the proteins that are essential in protecting the organism from oxidative stress, differed in spot intensities (three-fold or greater increase/decrease) for the different disease groups:

- DU isolates – KatA, Pfr, Icd and MetK;
- NUD isolates – Pfr; and
- GU isolates – TrxA/TrxB and RecA.

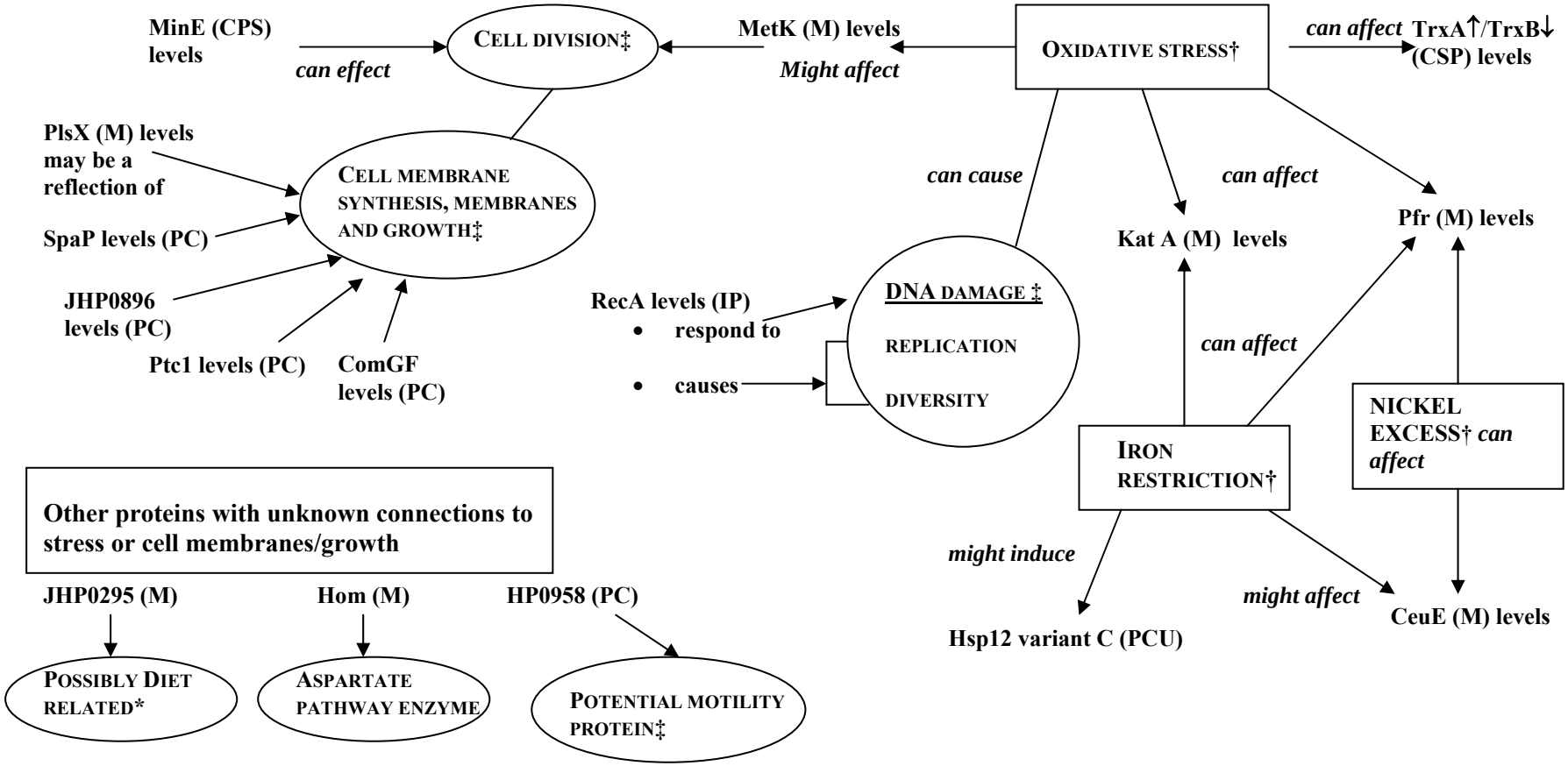


Figure 4.1 Proteins involved in stress response, morphology and other aspects of *H. pylori*

A diagrammatic representation showing the connection of proteins involved in stress response and morphology of *H. pylori*

IP=information processing M=metabolism CPS=cellular process and signalling PC= poorly characterised

† protein response to stress conditions ‡ proteins that affect physiology of *H. pylori* *protein response to non-stress related environment

Compared to the GU isolates the DU isolates produced the greatest number of proteins involved in the oxidative-stress response (KatA, Pfr, Icd and MetK). NUD, GI and DU isolates differed from GU isolates in spot intensities for TrxA/TrxB and RecA. While NUD isolates and DU isolates both differed in the spot intensity for Pfr in comparison to the GI and GU isolates.

Microorganisms adapt to environmental cues by turning on and off the expression of genes for survival within specific environmental niches. Such adaptations might include responses to changes in oxidative stress. The comparative proteome analysis revealed that isolates within a specific disease group displayed similar expression patterns for oxidative-stress proteins. These findings suggest that these proteins may play a significant factor in determining the disease outcome of a *H. pylori* strain.

Haeme and Non-haeme regulation

In *H. pylori* both iron acquisition and storage are governed by the ferric uptake regulator Fur (Bereswill et al, 1998a; Bereswill et al, 2000; Fassbinder et al 2000; van Vliet et al, 2002). In response to an increased iron concentration, Fur represses transcription of the iron uptake genes *fecA2* and *frpB1* (Delany et al, 2001a; Delany et al, 2001b; Fassbinder et al 2000) while transcription of the ferritin gene *pfr* increases (Bereswill et al, 2000, Delany et al, 2001a). Conversely, under iron-restricted conditions transcription of the iron uptake genes is stimulated and Pfr-mediated iron storage is repressed by Fur. Disruption of the *fur* gene was found to derepress *pfr* but was not found to be involved in the regulation *katA* or *ceuE*.

Expression of Pfr is also repressed by Fur upon increased nickel, copper, zinc, and manganese concentrations (Bereswill et al, 2000), leading to a model where modulation of Pfr transcription contributes greatly to the maintenance of iron and non-haeme metal homeostasis in *H. pylori*.

This study identified four proteins that might be affected by iron and non-haeme metal homeostasis, pH and temperature, these being Pfr, CeuE, KatA and Hsp12

variant C. These proteins have different expression responses to these environmental conditions:

- Iron restriction environment
  - Pfr – repressed
  - CeuE – upregulated
  - KatA – upregulated
  - Hsp12 variant C – upregulated
- Iron excess
  - Pfr – upregulated
  - CeuE – repressed
- Nickel excess
  - Pfr – repressed
  - CeuE – repressed
- pH stress
  - Hsp12 variant C – upregulated
- Temperature stress
  - Hsp12 variant C – upregulated

This study found that the expression of Pfr was decreased in both NUD and DU strains, however why this is the case may not necessarily be the same for both disease groups. One possible explanation is that iron environmental conditions affect DU strains and nickel conditions affect NUD strains.

Iron-restriction stress

Iron has an important role in maintaining basic metabolic functions, such as electron transport, and it has a high redox reactivity and toxicity. Pfr responds to oxidative stress, iron restriction and nickel excess. Under conditions of oxidative stress, Pfr expression increases. When iron-restriction and nickel excess are involved its production is reduced. KatA is expressed at high levels in response to oxidative stress *in vivo*. However, in an iron limited environment KatA expression is reduced, whereas normal iron levels promote KatA. Iron transport proteins, such as CeuE, are increased in response to iron restriction and are restricted in an iron rich environment. Comparative 2-DE analysis between the groups revealed that the DU group displayed differences in Pfr and KatA, while NUD displayed differences in Pfr, Hsp12 variant C and CeuE. While iron limitation may be a factor in determining the disease outcome of a particular *H. pylori* strain such as the DU isolates this may

not be the case for the NUD isolates. If iron limitation were a factor in the NUD isolates then the observed expression levels of CeuE would have been increased and KatA would have been decreased.

This study also identified Hsp12 variant C as a protein specific to the NUD isolates. Hsp12 variant C, an induced protein, that is known to respond to more than one type of stress, pH, temperature and iron restriction (de Vries et al, 2003). It is possible that Hsp12 was induced by iron restriction. However, no other corresponding variations in the expression of proteins involved in this type of stress, such as the CeuE and KatA response to iron restriction, were identified for the NUD isolates. Additionally because there are at least six variations of Hsp12 (de Vries et al, 2003) that may have different theoretical and observed pIs and M_rs it is uncertain if Hsp12 was not expressed by the other isolates in this study. It is also possible that Hsp12 may have other unidentified factors that alter its expression, for example nickel excess. Further examination would be required to determine other factors that might affect Hsp12.

Nickel excess stress

A constant supply of nickel ions is required for the synthesis and activity of a number of metalloenzymes that play a crucial role in prokaryotic and eukaryotic cells (Eitinger and Mandrand-Berthelot, 2000). However, when nickel ions accumulate, they inhibit growth and exhibit a toxic effect. Although Pfr and CeuE have been associated with nickel excess this stimulus also affects the expression of other proteins such as the ferric uptake regulator (Fur), iron-regulated outer membrane protein (FrpB), outer-membrane proteins omp11, omp31 and omp32, FlaA, FlaB etc (Contreras et al, 2003). Both DU and NUD isolates displayed differences in the spot intensity of Pfr but, NUD isolates were the only ones that differed in the expression of CeuE (specifically absent). The comparative analysis did not reveal any significant differences in the spot intensities of the other proteins known to be involved in nickel excess response. Additionally the NUD isolates did not have similar spot intensities for FlaA and FlaB. Also the spot intensity for FlaA and FlaB of the NUD3 isolate was at least six-fold greater than nine other isolates (GU1, GU2, GU3, GI1, GI2, NUD1, NUD2, DU1 and DU3). The DU2 was the only other

isolate with an abundance of FlaA and FlaB with at least a four-fold increase in spot intensity. This suggests that the expression of any other proteins that are involved in the nickel excess response would not be correlated with any disease groups. Further investigation into these proteins that respond to nickel excess is required to determine if it might be a factor in the NUD strains.

Cell division and membrane proteins

Stress conditions such as pH can also affect other essential proteins e.g. membrane proteins and cell growth proteins involved in cell division and proliferation.

Two proteins were identified for the DU isolates that have possible roles in cell division, MetK and MinE. As previously mentioned, MetK has possible roles in oxidative stress response and cell division (Wei and Newman, 2002; Caro and Cederbaum, 2004). Cell division can also be hampered by the overexpression of MinE (de Boer et al, 1992). A comparison of the groups revealed that the difference in the spot intensities of the cell division proteins was a characteristic of DU group. Ptc1, found in the GI isolates, plays a possible role in a variety of cellular mechanisms, including the control of metabolism, the cell cycle, cell proliferation and differentiation (Tatsuya et al, 1993). However, very little is known about the possible function of Ptc1 in *H. pylori*.

The proteins with possible roles in the cell membrane are PlsX (DU isolates), ComGF (DU and NUD), SpaP (DU) and JHP0896 a predicted dynamin GTPase (GI). Very little is known about the possible function of these four proteins. PlsX is possibly involved in cell membrane synthesis (Rock and Cronan, 1982), ComGF is a membrane protein for the binding of transforming DNA to the competent cell surface (Chung et al, 1998), SpaP may be involved in cell binding (Demuth and Irvine, 2002) and JHP0896, a predicted GTPase, has been implicated in vesicle formation and traffic (Rothman et al, 1990; Suzuki et al, 2001). The function of the cell growth and membrane proteins has only been predicted, further investigation would be required to ascertain if iron limitation and/or oxidative stress might affect these proteins.

Other proteins

The remaining protein differences that were identified had no apparent connection to protection from stress or cell division and cell membrane.

Hom, found in the NUD and DU isolates, is an enzyme in the aspartate pathway (Viola, 2000). Currently no study has examined the response of this protein to any influences, and no relation to the other conditions mentioned have been noted in either eukaryotes or prokaryotes.

Animals such as ruminants are dependent on the presence of microorganisms in the fore-stomach and rumen for the digestion of large amounts of plant material (Hespell and Whitehead, 1990). Small amounts of plant fibres are also degraded by the microorganisms in humans. JHP0295, which was identified as a specifically unique protein for the DU isolates, is a predicted xylanase that is produced by microorganisms to degrade xylan a component of plant material (Bastawde, 1992; Kulkarni et al, 1999). The regulation of xylanase secretion by microorganisms is still not completely understood. It is believed that the xylanase is induced in these microorganisms (Bastawde 1992; Kulkarni et al, 1999). Because xylan cannot enter the microorganism, the expression of xylanase is stimulated by low molecular mass xylan fragments. These low molecular mass fragments include xylose, xylobiose, xylooligosaccharides, heterodisaccharides of xylose and glucose and play a key role in the regulation of xylanase biosynthesis (Kulkarni et al, 1999). These fragments are released from xylan by a small amount of constitutively produced xylanase (Bastawde, 1992; Kulkarni et al, 1999). Cellulose has also been found to be an inducer (Hrmova et al 1984). It is not known if the production of xylanase is inducible or constitutive in *H. pylori*, however it most likely that it is inducible. When a protein is induced this would give the appearance of a specifically unique protein consequently any alterations in the levels of xylan might result in an increase or decrease in the protein spot intensity. Whereas if the enzyme is constitutive the protein spot would still be observable in the proteome map of all isolates. It is possible that xylanase is not a unique protein but might alternatively be a differentially expressed protein in DU strains. Further investigations need to be undertaken to understand the implications of this protein in DU strains.

HP0958 (also known as JHP0892), in the GU isolates, has recently been identified as a potential motility associated protein with strong interactions to RNA polymerase sigma-54 factor (RpoN) and FliH (Pereira and Hoover, 2005; Ryan et al, 2005). It has been found that when HP0958 is absent the production of FlaA and FlaB is decreased and the production of FlgM is increased. However, in this study any decrease in HP0958 was not found to correspond to alterations in FlaA, FlaB and FlgM as observed by Pereira and Hoover (2005). The expression of FlaA, FlaB and FlgM was variable irrespective of HP0958 production levels. This implies that HP0958 may have other possible functions other than motility, or that flagellin production is influenced by other factors.

YjgF - little is known about its function.

The proteins AhpC, NapA, GroEL, Rpo7/12, SodB, RpoA and Tsr [see section 4.7.1] were found to be positional shifts

Forty eight percent of the identified differences in the proteome were due to unknown or poorly characterised proteins.

In summary, the differences in the stress response, cell division and membrane proteins of *H. pylori* isolates could account in part for the different clinical outcomes of infection. These proteome variations between isolates that are associated with a specific disease outcome might indicate genes that would influence colonisation of a particular niche, such as the duodenum. Variations in protein expression may be in response to specific stimuli. These stimuli in turn can produce other effects that may alter the levels of other proteins. For example, oxidative stress alters the expression of KatA, Pfr, etc (Figure 4.1). Additionally, oxidative stress may cause DNA damage that alters the production of RecA.

4.11 Conclusions

Proteome analysis has proven to be a useful method for comparing *H. pylori* isolates and disease outcomes. Through pattern matching of 2-DE spots it was possible to

determine how many proteins were common to all isolates and to detect specific differences between isolates involved in different disease outcomes. One hundred and seventy three protein spots were identified in this study from the whole cell proteins of eleven *H. pylori* isolates. Twenty one were basic protein spots that have been characterized for the first time in *H. pylori*, with the remainder having been characterized in previous studies. The proteome analysis highlighted the extensive strain variation that *H. pylori* exhibits with only 471 spots (40.6%) out of ~1168 per isolate being located at the same 2-D position. Furthermore, through the dual application of two medium range IPGs and two gel systems this study has successfully improved the resolution of basic and low molecular mass proteins and provided further information towards the compilation of a master reference map of *H. pylori*.

In contrast to other microorganisms whose proteomes have been studied, basic proteins form a large proportion of all the 2-DE patterns of the *H. pylori* isolates investigated. Because the majority of the proteins in *H. pylori* have pIs within the range of pH 6-11 there is a need to resolve all these basic proteins. Previous comparative studies resolved proteins through the application of IPG with a pH range of 6-10 which did not provide efficient and reproducible separation of proteins. This study has shown that IPG 2-D gels in the range pH 6-11 improved the resolution of basic proteins within the pH range 8.5–10.5. However, despite the advances in alkaline range IPGs that have improved the reproducibility and quality of separations within pH range 6–11 many of the basic proteins still remain unrepresented. This lack at the more alkaline pH may be due to the instability of the gradient, or the instability of the proteins themselves, at an extremely alkaline pH. Additionally when *H. pylori* are removed from their acidic environment many of these basic proteins that may have been required for colonisation in a hostile *in vivo* environment may not be expressed *in vitro*.

This study utilised two gel systems - Tris-glycine and Tris-tricine. The application of tricine as the trailing ion improved the resolution of *H. pylori* proteins with masses between 7-20 kDa, however, no proteins with masses below 7 kDa could be visualised. In addition, the Tricine system also provided excellent resolution of

proteins between 20-62 kDa. However, to obtain efficient separation of proteins with high molecular masses between 62-150 kDa glycine was utilised. Approximately 38 protein spots migrating in the 7-20 kDa region were identified by matrix assisted laser desorption/ionization - mass spectrometry, of which 22 were identified for the first time. The digestion of the low mass proteins often produced only a few peptides, which were either insufficient for efficient identification by mass spectrometry or could not be matched with any proteins in the database. The gel system described here may be useful for the efficient separation of low molecular mass proteins in future studies to construct proteome maps of *H. pylori*.

Protein matching analysis of isolates grown in normal culture media indicated that the differences are mainly involved in stress responses, cell division and cell membrane protein production. The comparative proteome analysis revealed that isolates associated with a specific disease group displayed similar expression patterns of these proteins. Proteins that had known functions were essential to the organism's survival and growth in a hostile environment. Microorganisms adapt to environmental cues by turning on and off the expression of genes for survival within specific environmental niches. Such adaptations might include responses to changes in oxidative stress. These findings suggest that these disease-specific proteins identified in this study may play a significant factor in determining the preferential colonisation of particular *H. pylori* isolates in specific mucosal layer niches and/or the disease outcome of a *H. pylori* strain.

Most significantly this study has found, through the application of protein matching and cluster analysis, that protein differences in isolates could be correlated to four different clinical outcomes. While qualitative analysis divided isolates into four distinct clusters, quantitative analysis did not display this trend. However, both the qualitative and quantitative cluster analyses showed that isolates formed two major clusters, gastric and duodenal. Interestingly, the isolates in the gastric groups (GU, NUD, and GI) clustered closer together than the DU isolates. This indicates that isolates may adapt to particular microhabitats or that certain isolates are better suited to a particular microhabitat.

The protein differences observed between the disease groups included in this study may provide further insight into the physiology of this organism and its proteins. Of course further investigation is needed to determine if other isolates from these disease groups also have the same characteristic proteins. However, the results obtained have shown that protein matching and cluster analysis may provide a potentially powerful tool for diagnostic, taxonomic and epidemiological investigations of *H. pylori*.

4.12 Future Directions

This study should be expanded in at least four directions. The first includes increasing the number of isolates included in the comparison, particularly in the Gastritis group. Secondly, improving the technique to detect and resolve low molecular mass proteins below 7 kDa. Thirdly, using alternative techniques for the analysis of high molecular mass proteins greater than 150 kDa and finally repeating the analysis of basic proteins between the pI 10-12.

The gastritis group was represented by two isolates NCTC11638 and 26695, which have been cultured in the laboratory a number of years. Therefore, proteins identified in the gastritis disease group may not be indicative of disease markers. Isolates from more recent biopsies might provide a more accurate representation and therefore more gastritis isolates should be studied. Because so many proteins of NCTC11638 (Lock et al, 2001) and 26695 (Jungblut et al, 2000) have been identified and referenced these isolates provide valuable information on conserved proteins.

When basic proteins are of interest and need to be displayed in the gel an alternative extraction procedure (alkaline) might need to be utilised. Extraction could be performed by two extraction procedures. The first extraction procedure (for the acidic proteins pH 4-7) could use a Tris base up to 40 mmol (Westermeier and Naven, 2002) and in the second extraction the basic proteins greater than pH 7 might then be precipitated in TCA/acetone (Ohlmeier et al, 2000).

Currently techniques to resolve and observe extremely basic proteins are not commercially available, while the methods for detection of low molecular mass proteins still require further improvements. However, new methods such as HPLC that are currently being utilised for the detection of high and low molecular mass proteins show promise.

This study compared the proteome of these isolates under normal *in vitro* culture conditions. Additionally many of the proteins observed were high abundance proteins. Therefore, in order to obtain a complete proteome map of this organism we may require the application of two different approaches. The first would require prefractionation techniques to enrich for low abundance proteins. The second approach would include examination of the organism under different conditions of culture. This would include variables such as starvation of the organism, culturing the organism in the presence of antibiotics, and cell culture techniques to observe how the organism interacts in the presence of human gastrointestinal cells.

Although the proteome of 26695 has been examined this has not been the case for *H. pylori* strain J99. The genomes of 26695 and J99 have been compared (Alm and Trust, 1999) and found to be largely conserved so it would be of interest to compare the proteome of these two isolates to determine how their 2-D maps compare.

Finally, this study should be expanded to involve the large-scale screening of isolates associated with specific disease outcomes, in particular GU, NUD, DU and GI. The traditional method used to be 1D Western Blotting of isolates. However, the new method in proteomics using protein microarrays has great potential. Therefore the use of current screening techniques might include the use of protein microarrays for the detection of possible disease markers within other isolates. At least 60 proteins present themselves as possible marker proteins of interest for future large-scale screening of *H. pylori*. In order to ascertain if other isolates from cases of GU, NUD, DU and GI contain these marker proteins at least one hundred isolates should be screened from each of the four groups.

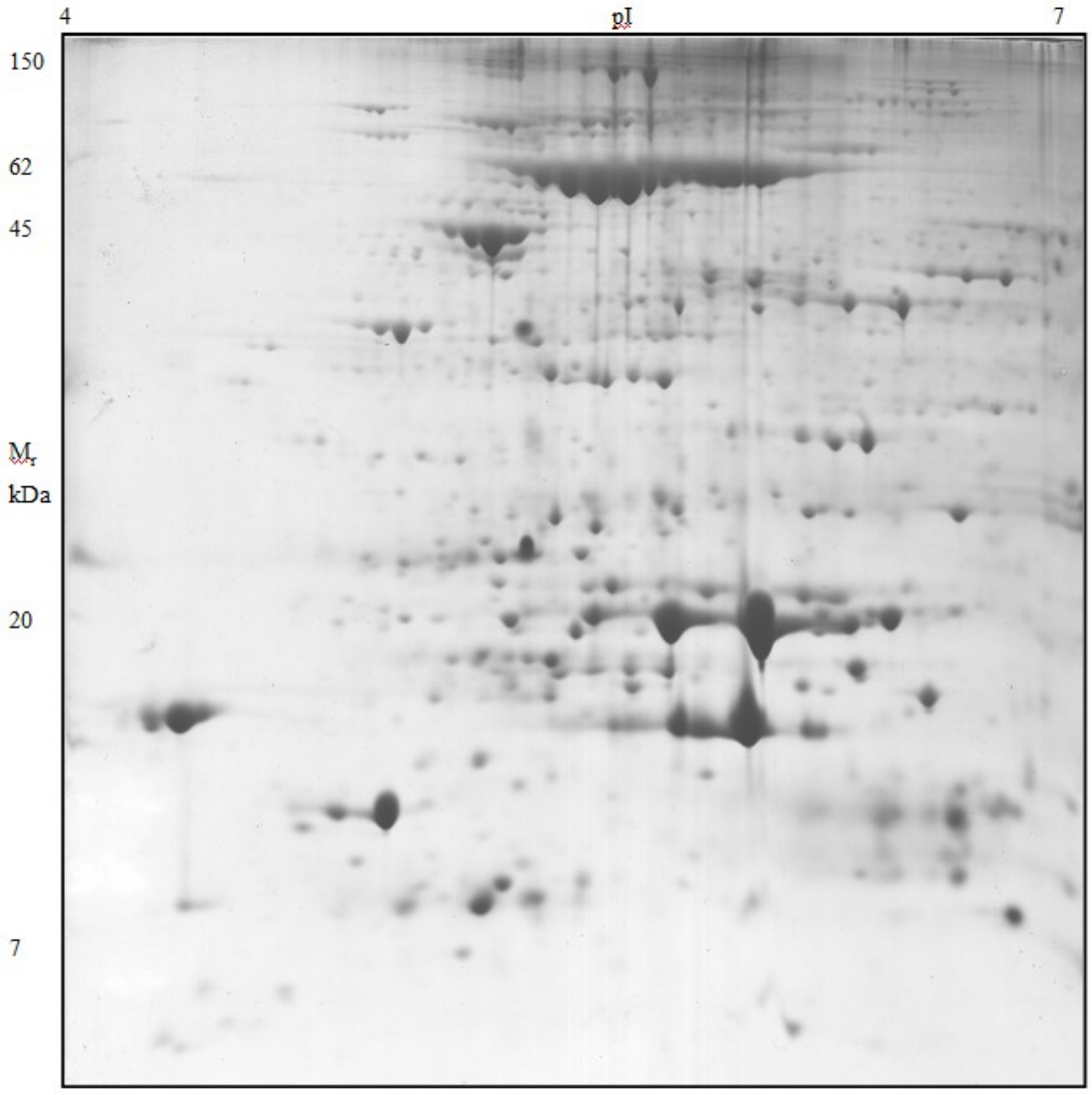


Figure A.1 *H. pylori* GU1 pH 4-7

GU1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

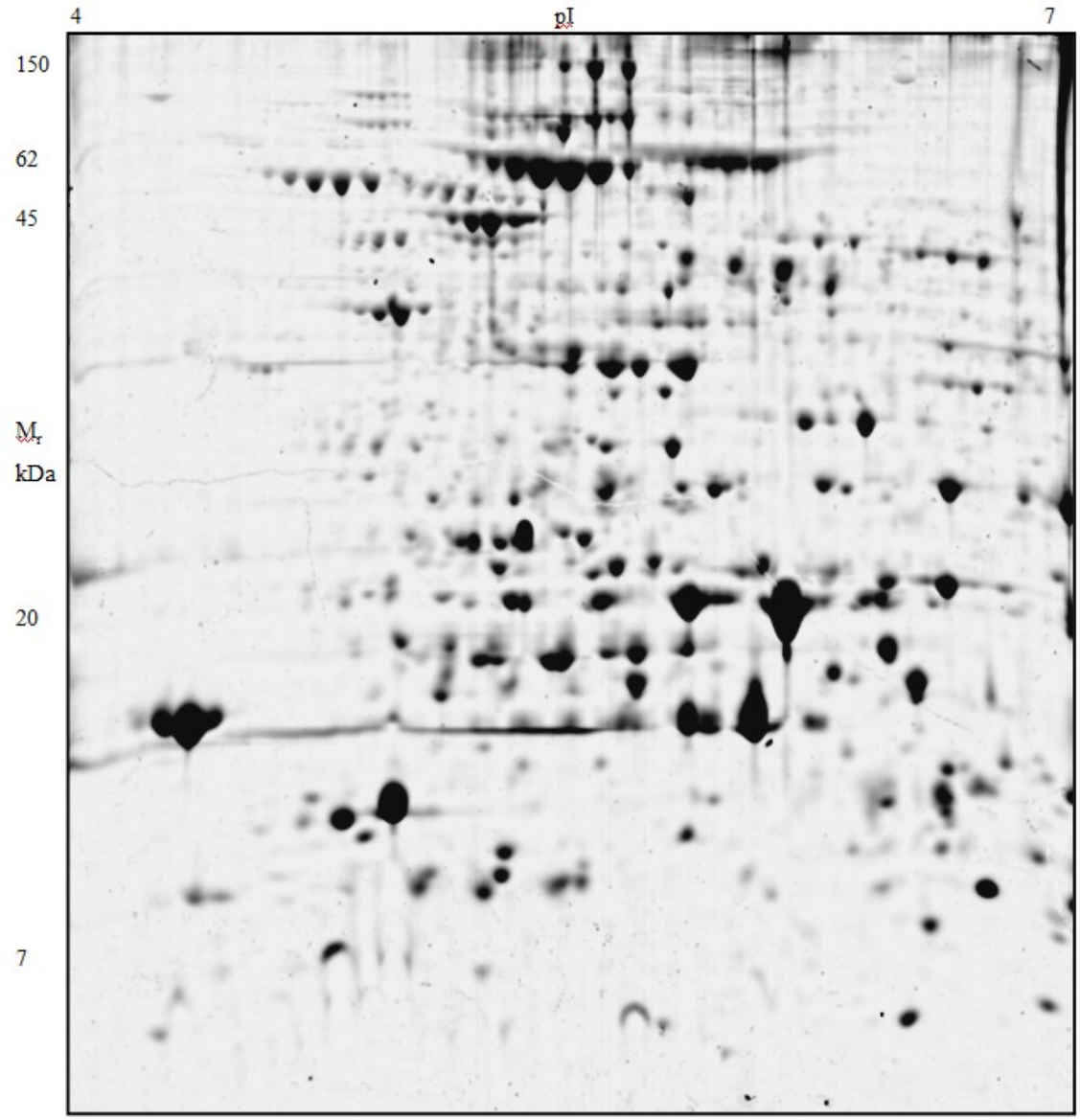


Figure A.2 ***H. pylori*** **GU2 pH 4-7**

GU2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

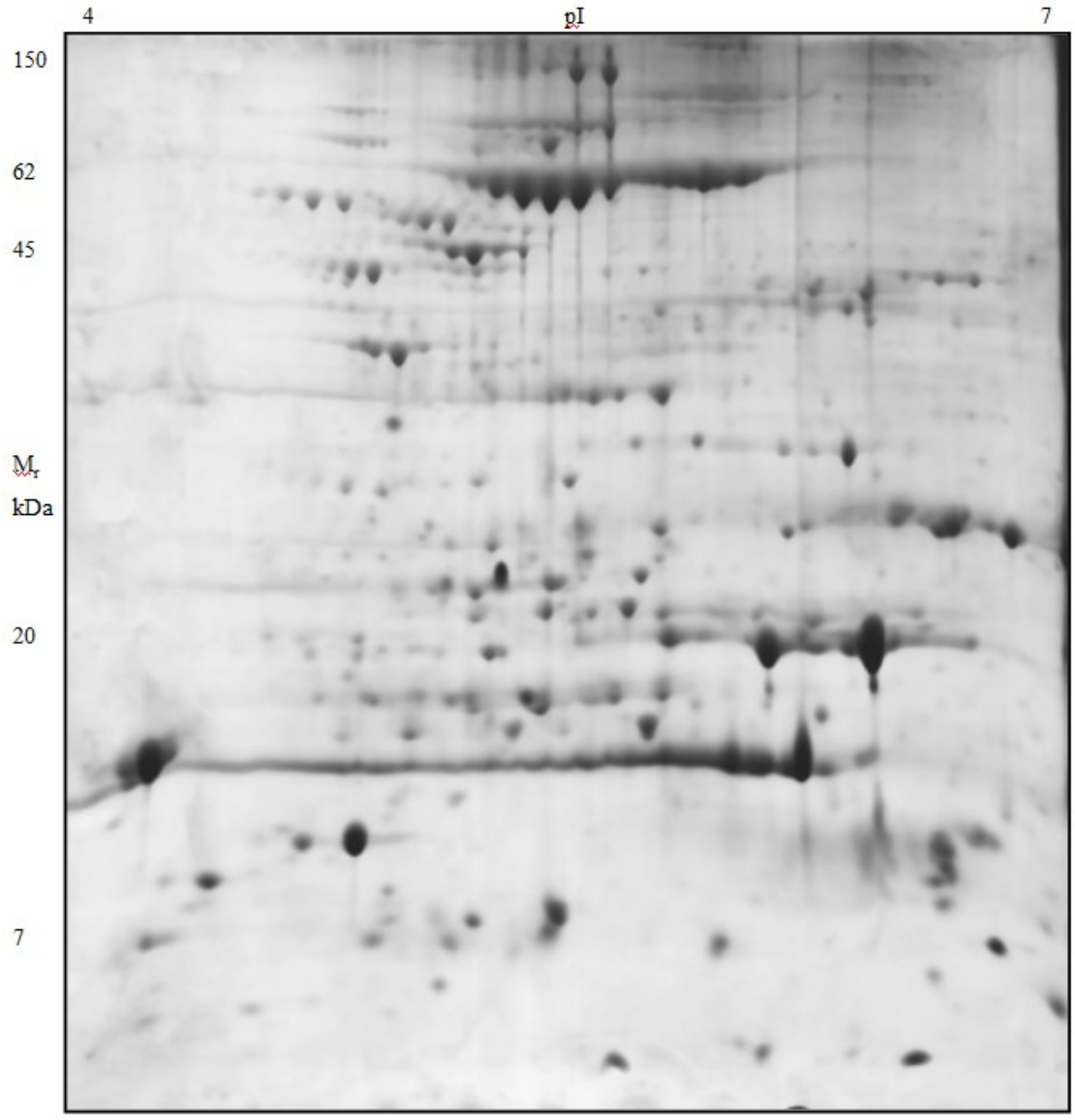


Figure A.3 *H. pylori* GU3 pH 4-7

GU3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

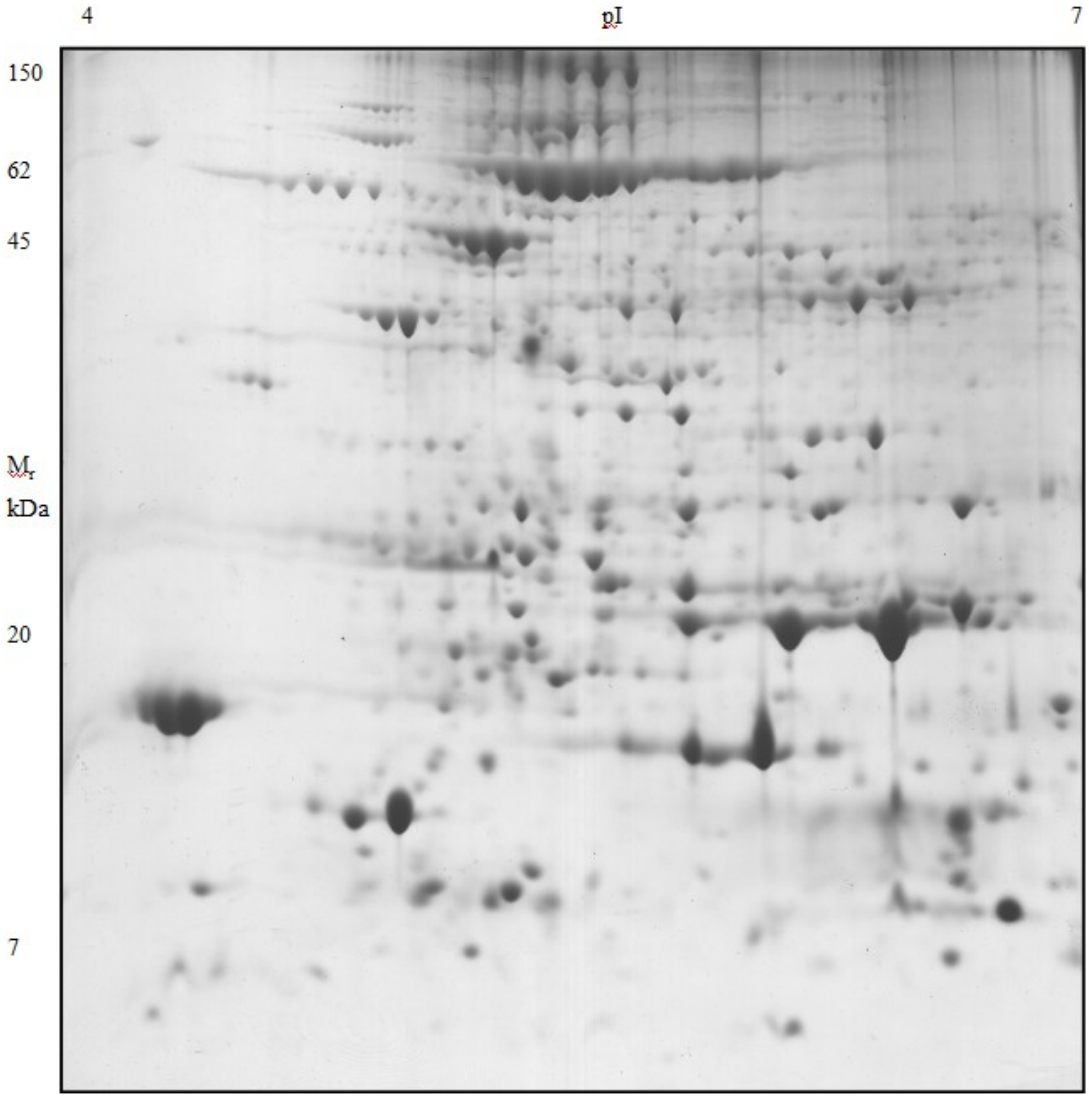


Figure A.4 *H. pylori* NUD1 pH 4-7

NUD1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

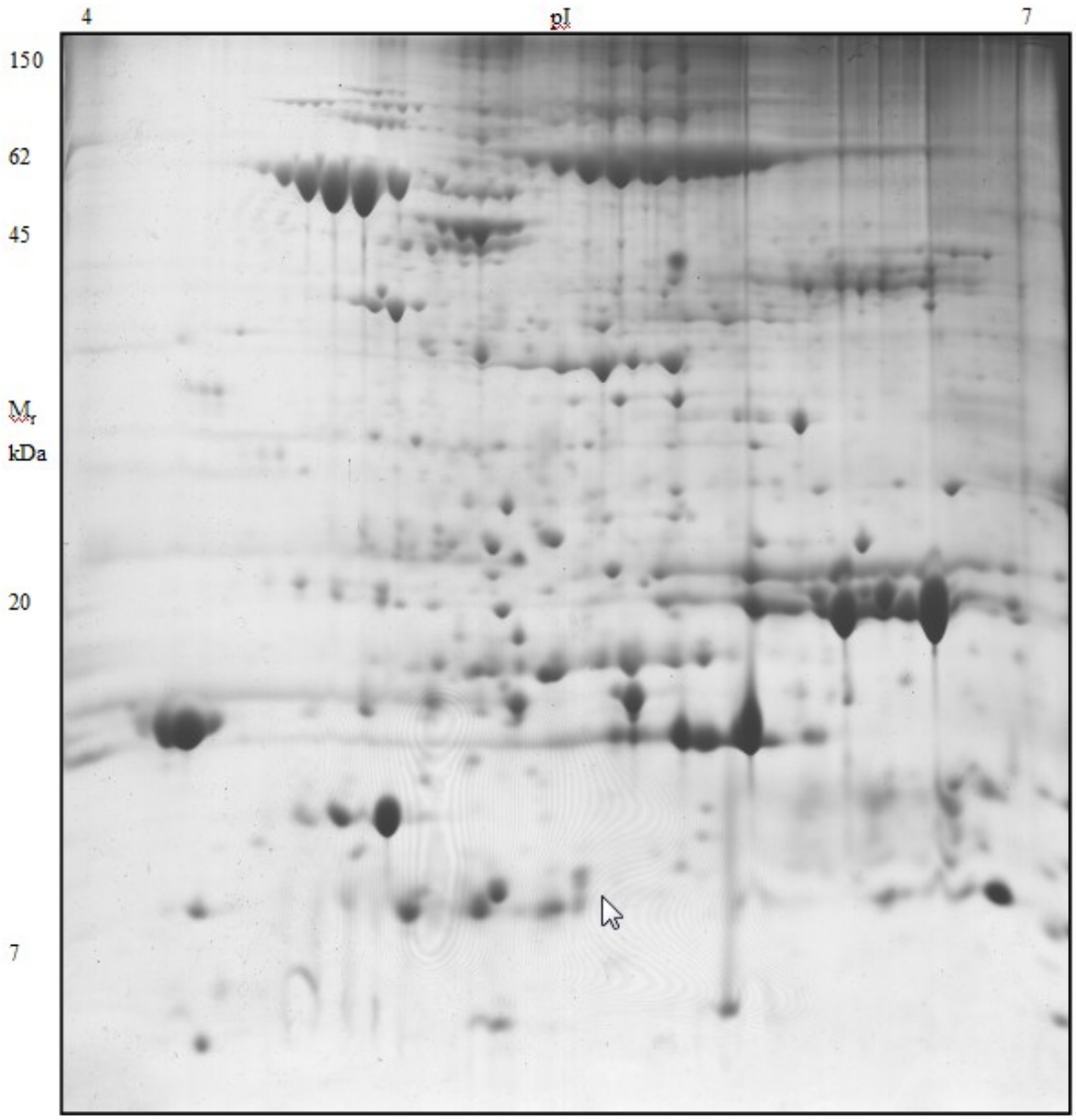


Figure A.5 ***H. pylori*** **NUD2 pH 4-7**

NUD2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

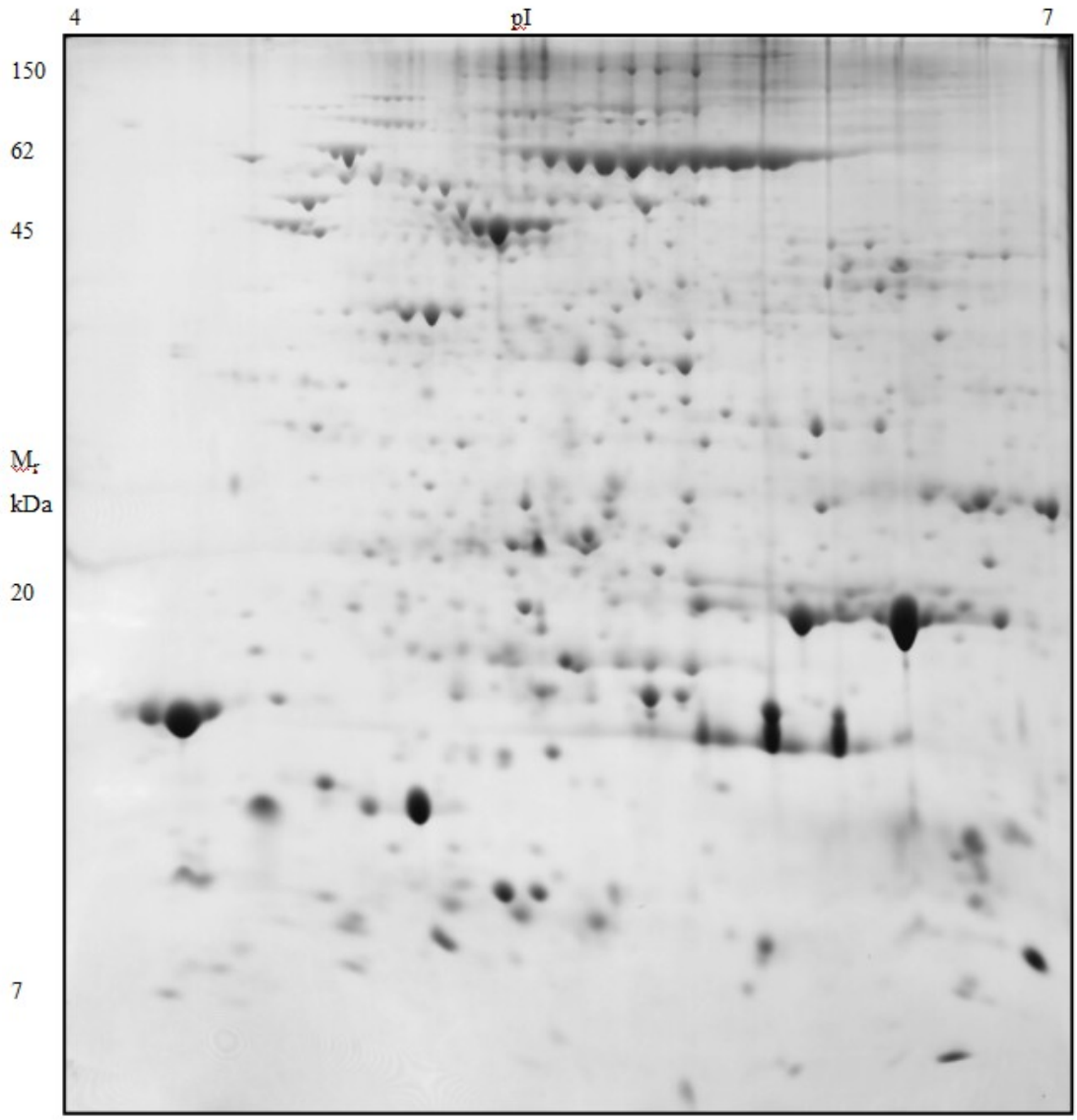


Figure A.6 *H. pylori* NUD3 pH 4-7

NUD3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

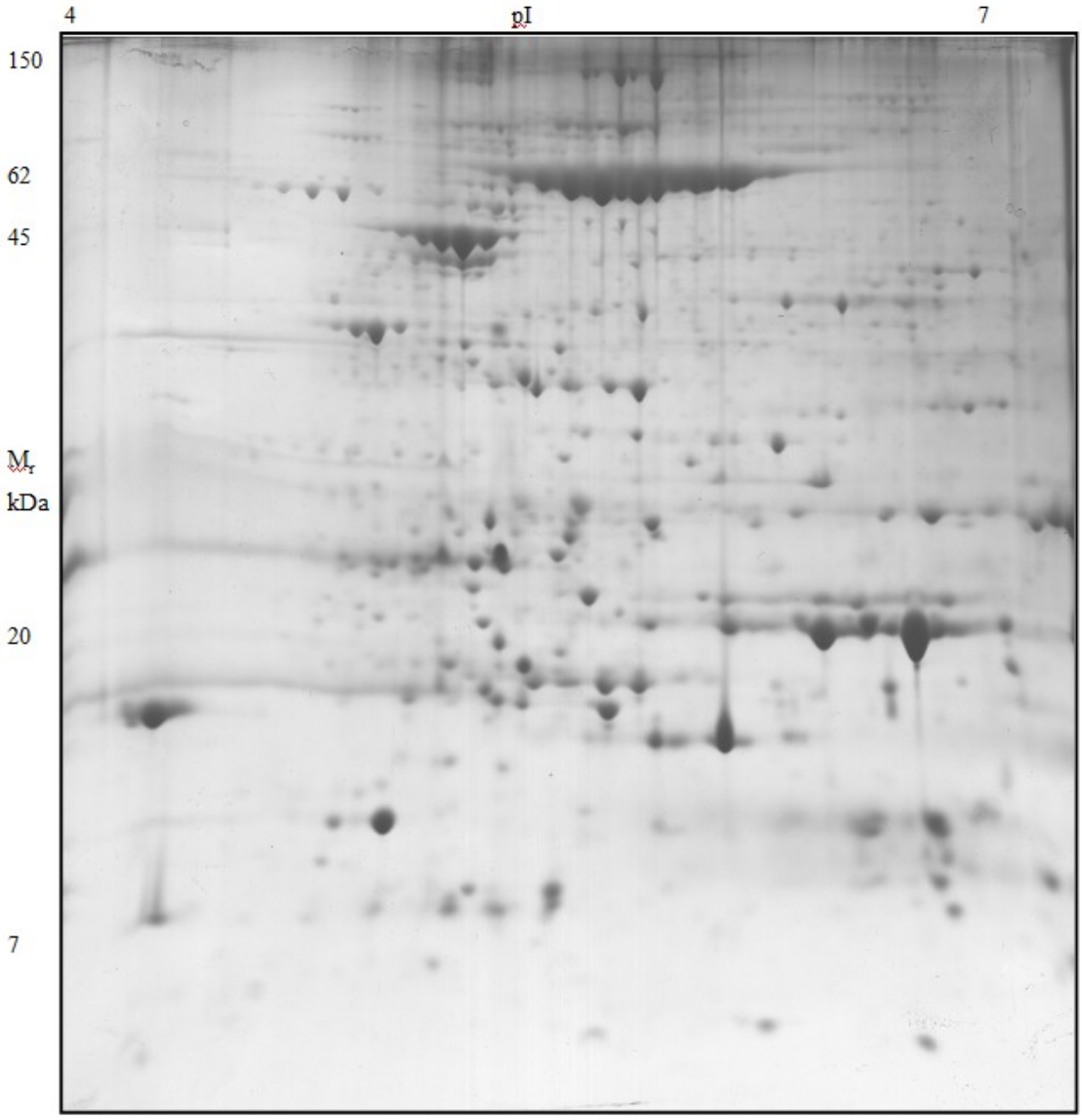


Figure A.7 *H. pylori* DU1 pH 4-7

DU1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

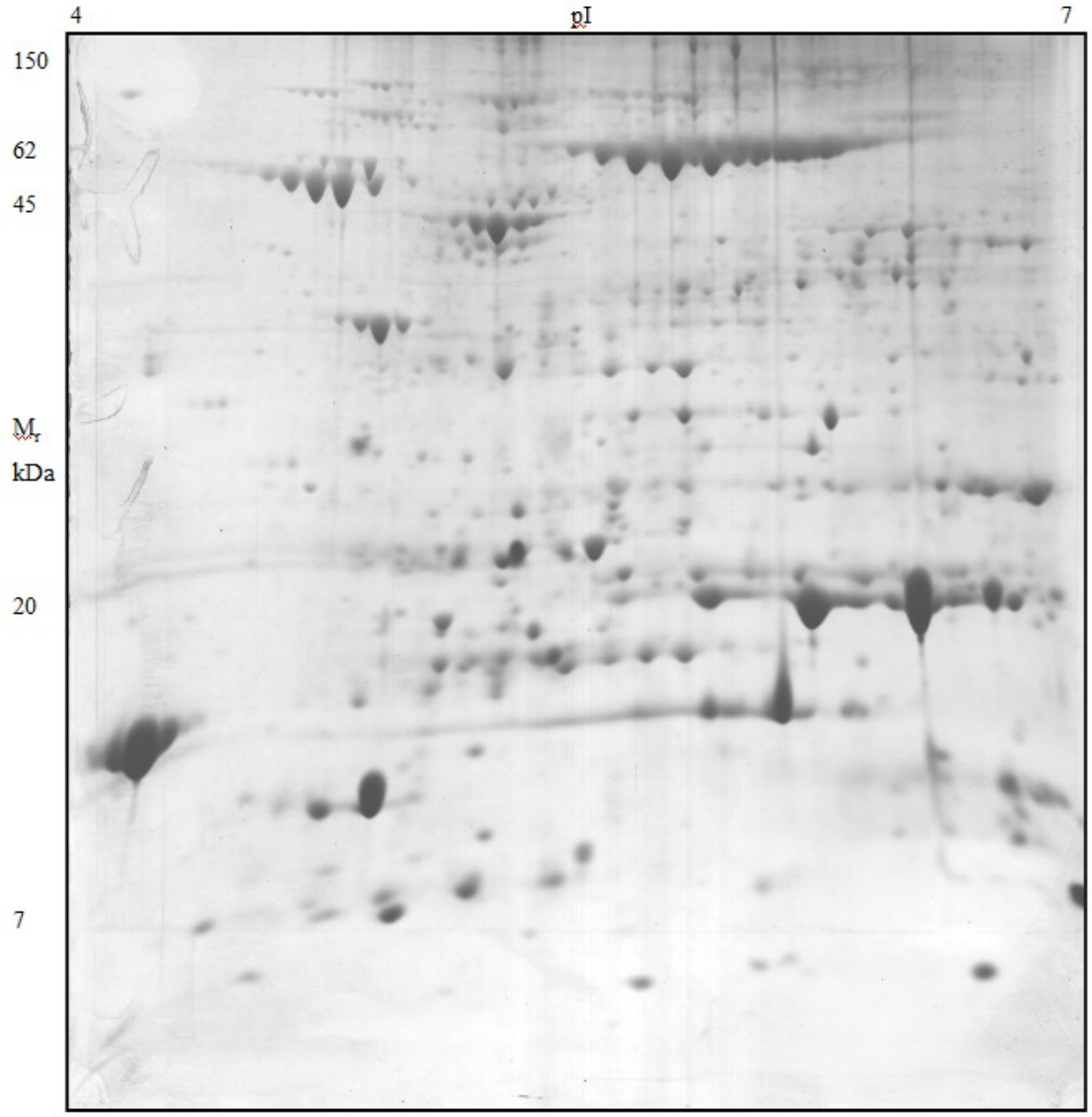


Figure A.8 *H. pylori* DU2 pH 4-7

DU2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

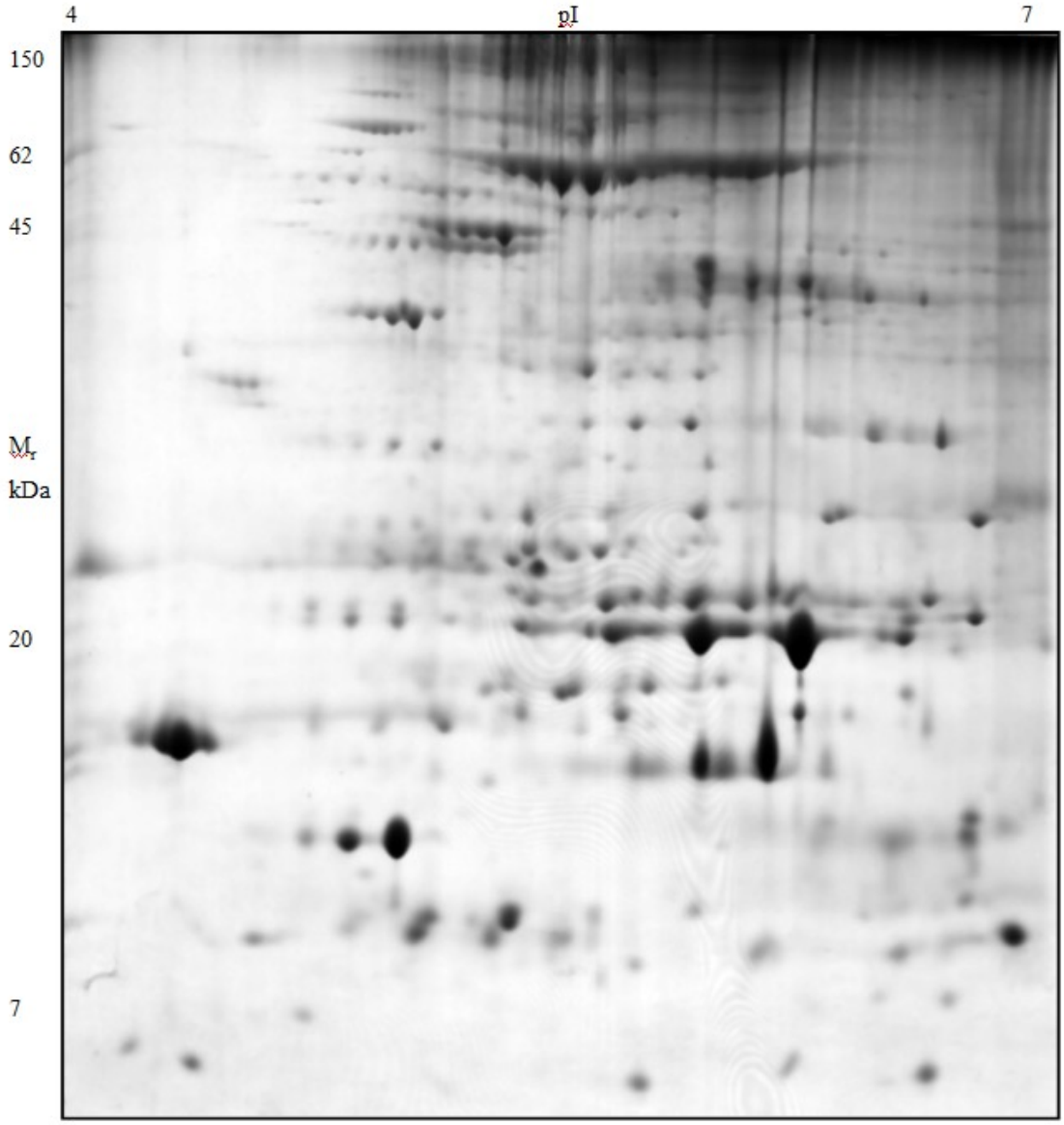


Figure A.9 *H. pylori* DU3 pH 4-7

DU3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

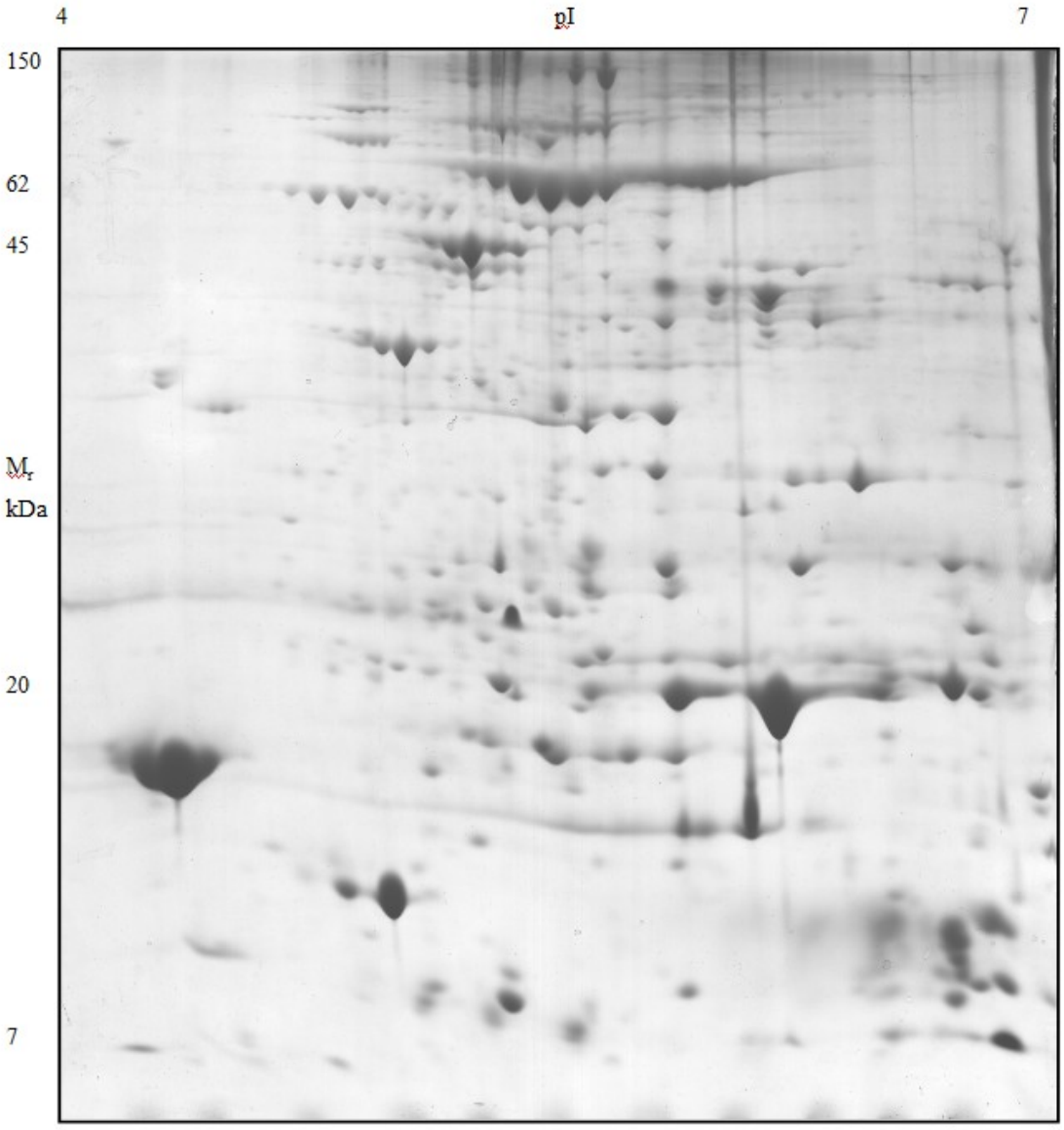


Figure A.10 *H. pylori* GI1 pH 4-7

GI1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

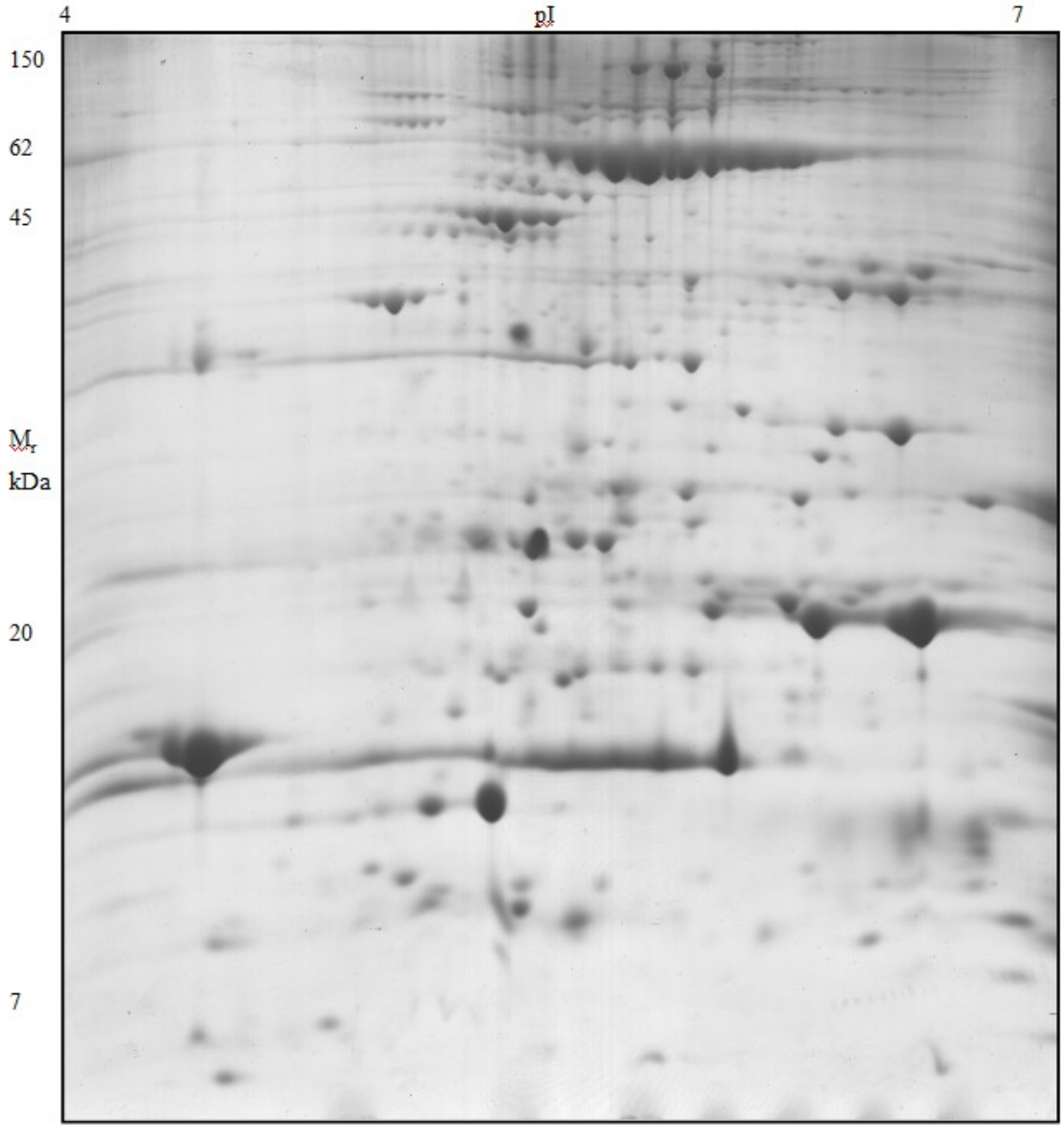


Figure A.11 *H. pylori* GI2 pH 4-7

GI2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards

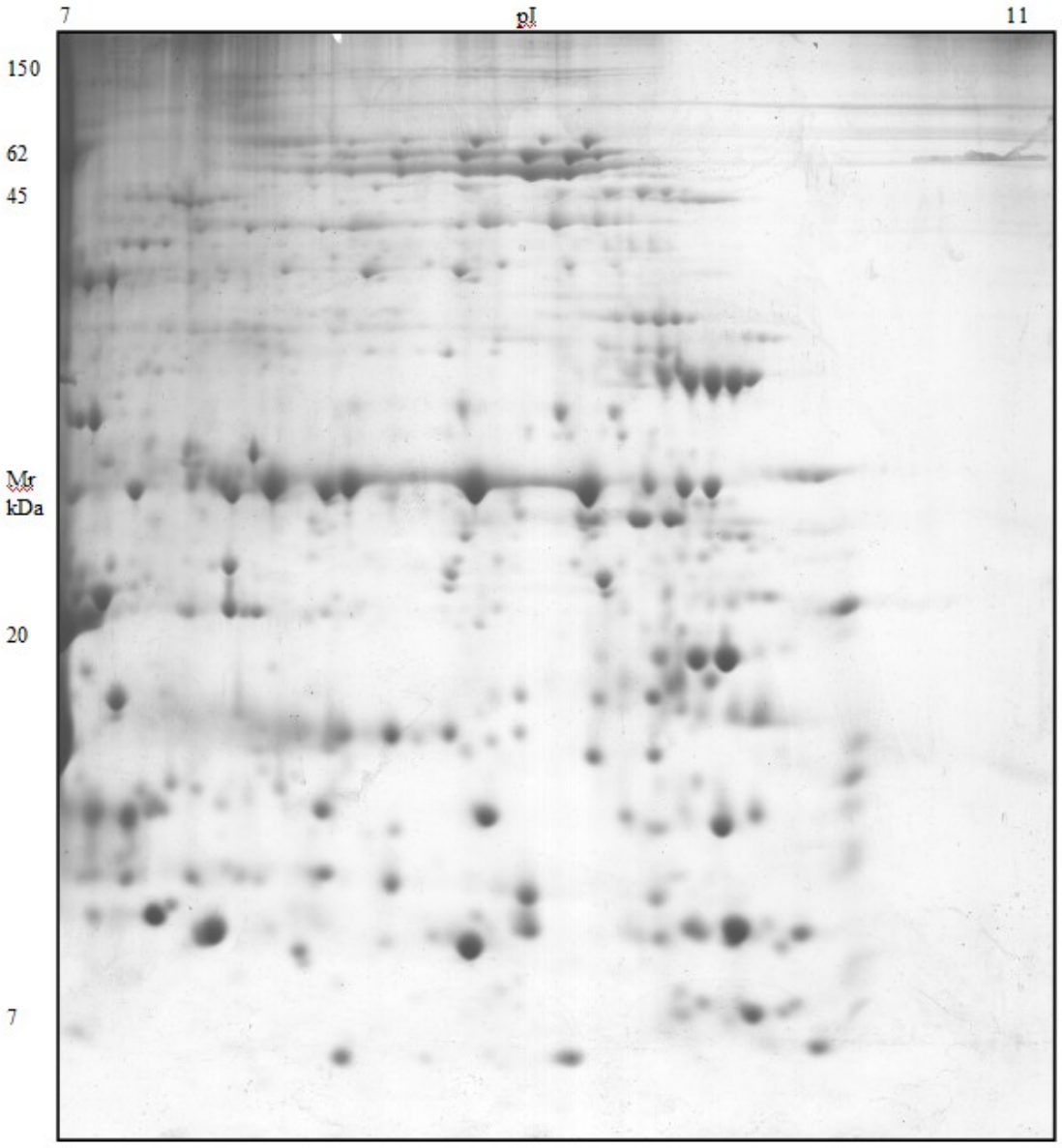


Figure A.12 *H. pylori* GU1 pH 6-11

GU1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

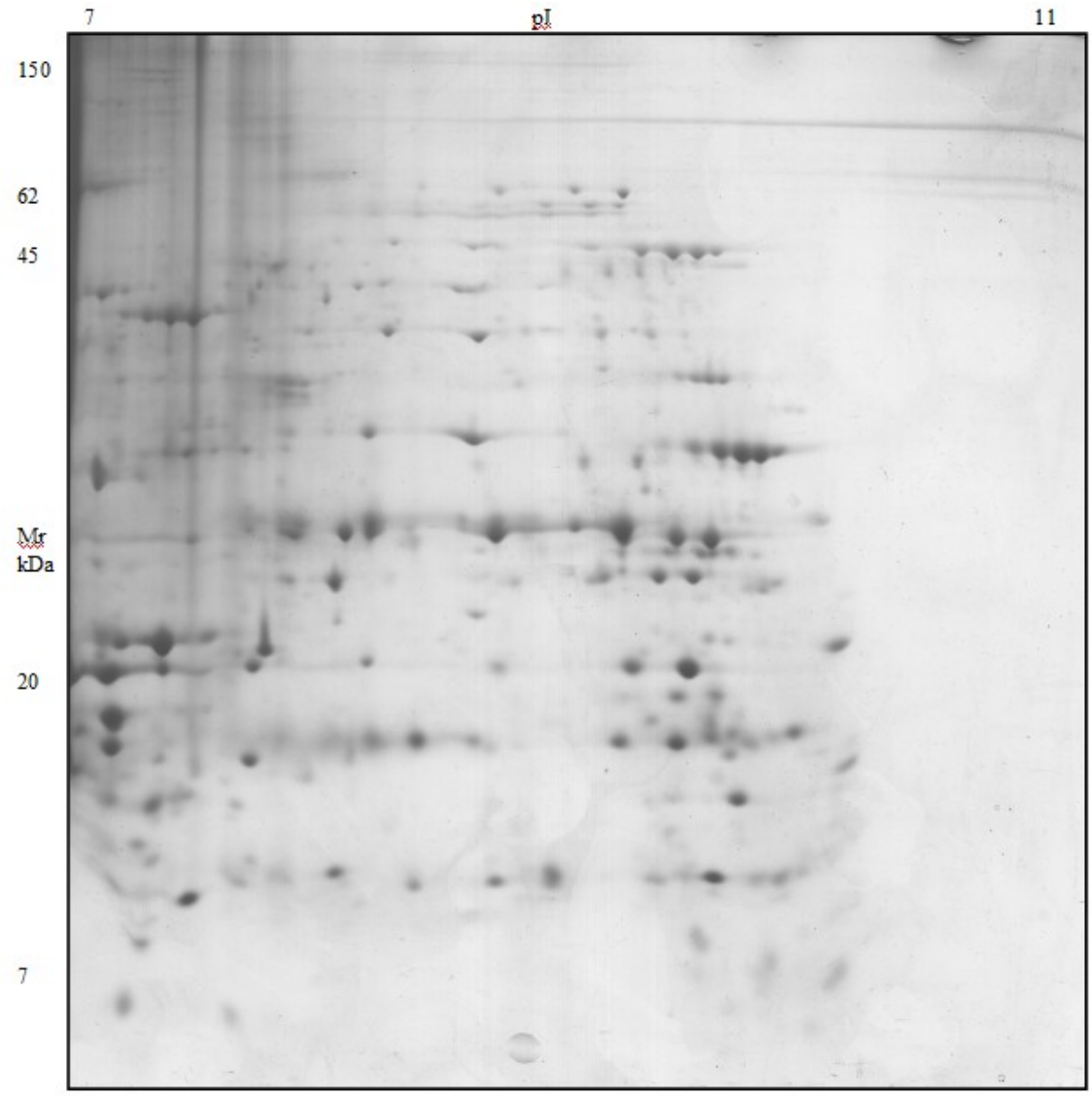


Figure A.13 *H. pylori* GU2 pH 6-11

GU2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

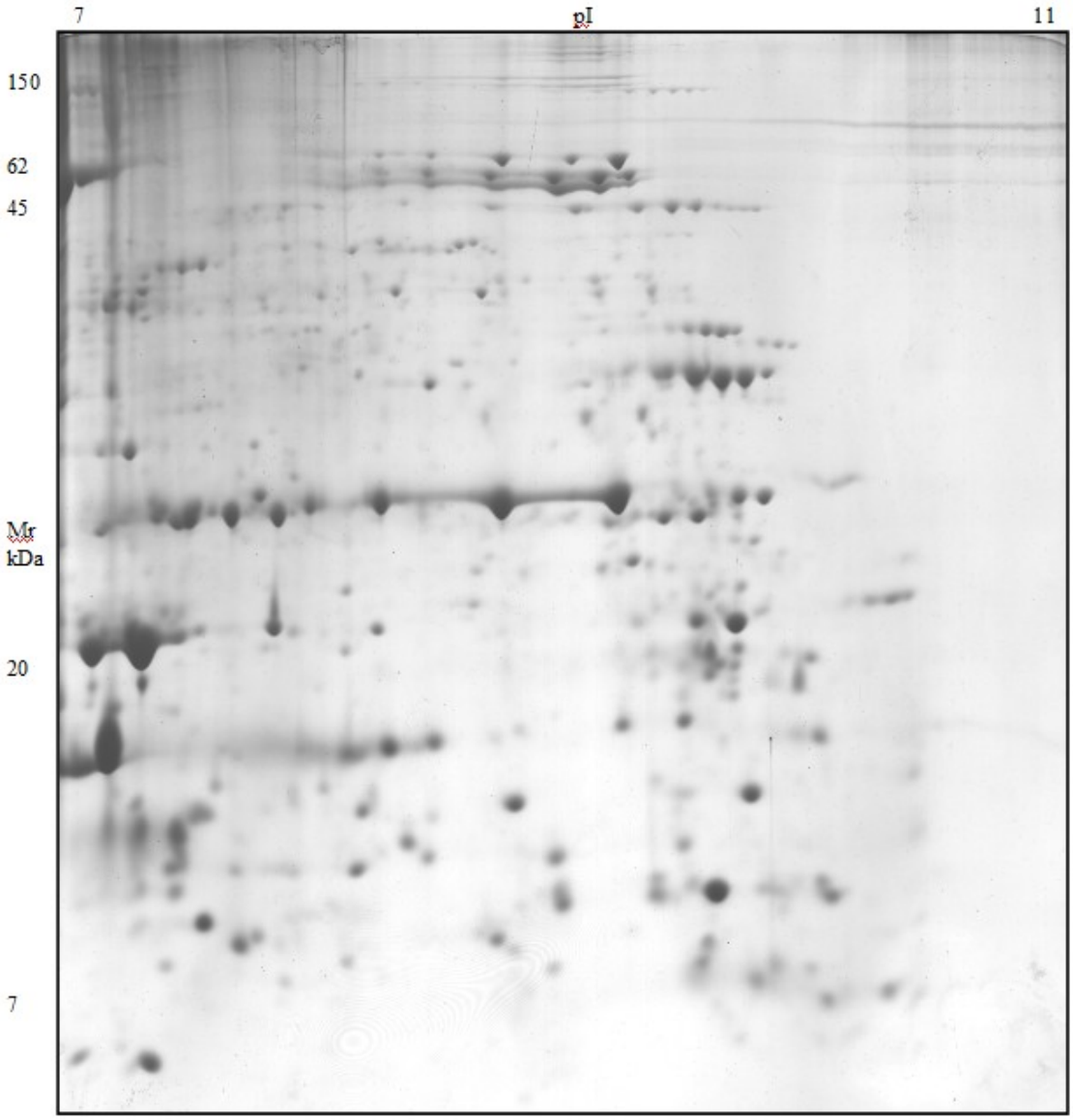


Figure A.14 *H. pylori* GU3 pH 6-11

GU3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

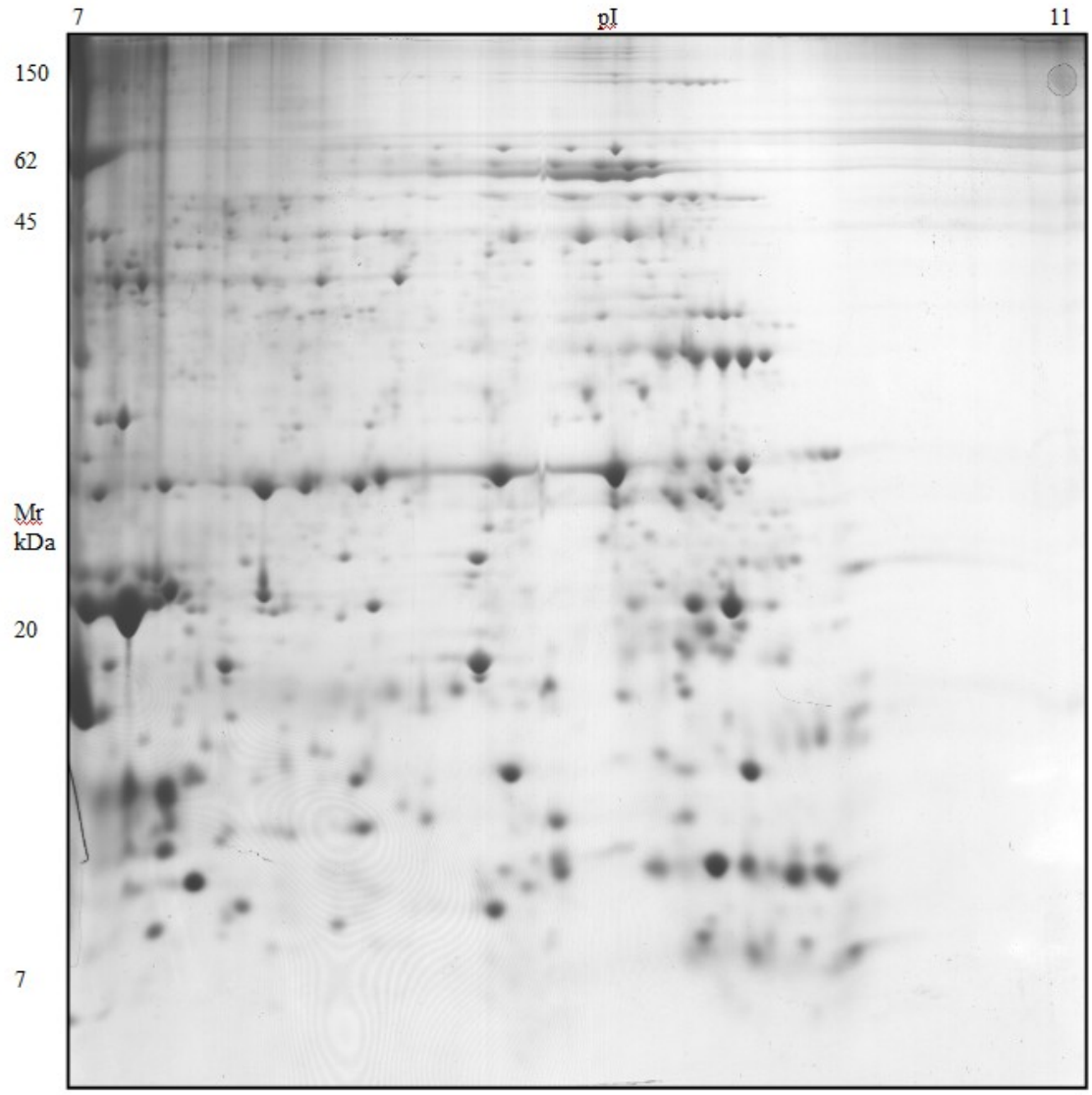


Figure A.15 *H. pylori* NUD1 pH 6-11

NUD1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

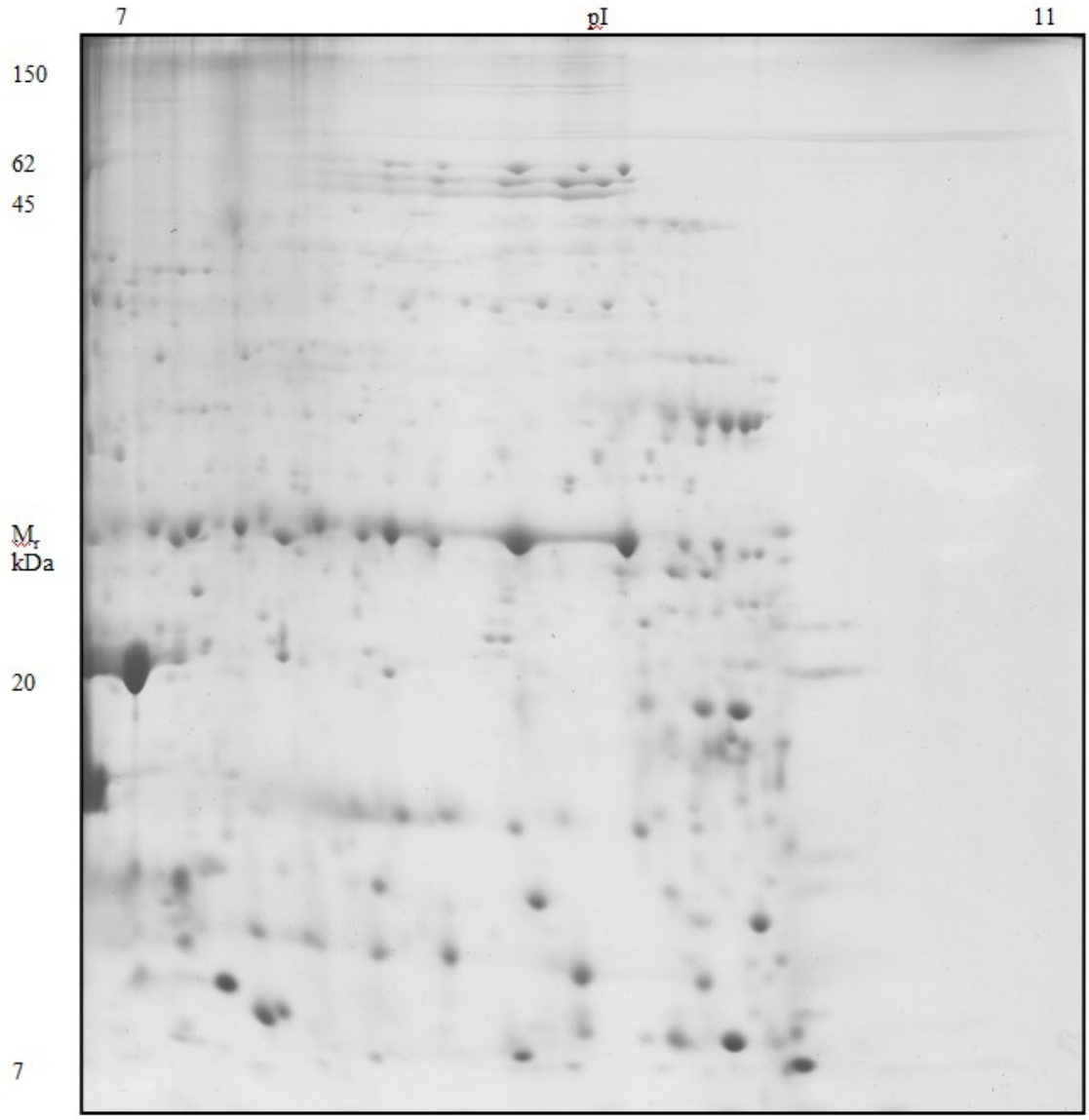


Figure A.16 *H. pylori* NUD2 pH 6-11

NUD2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

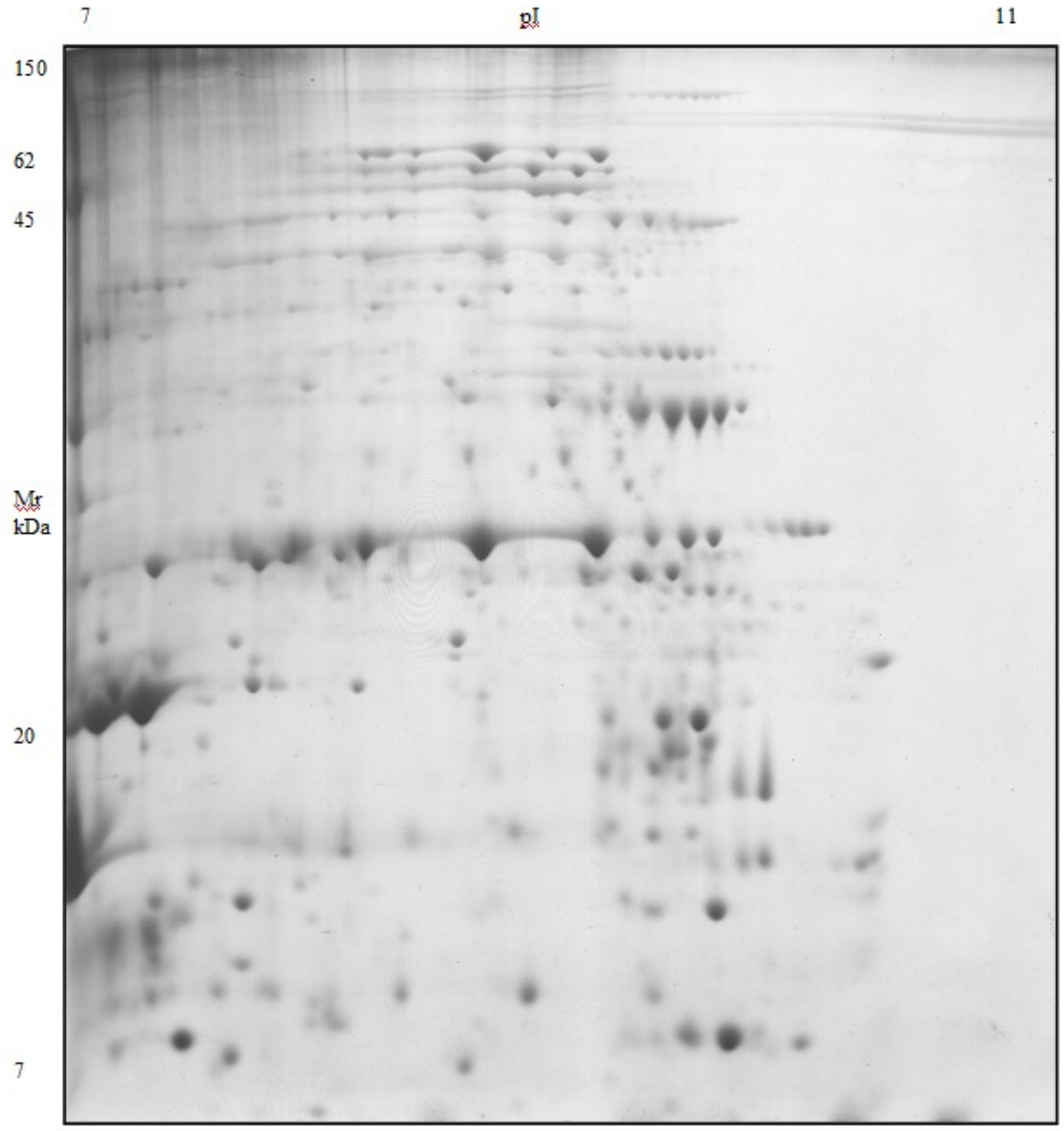


Figure A.17 *H. pylori* NUD3 pH 6-11

NUD3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

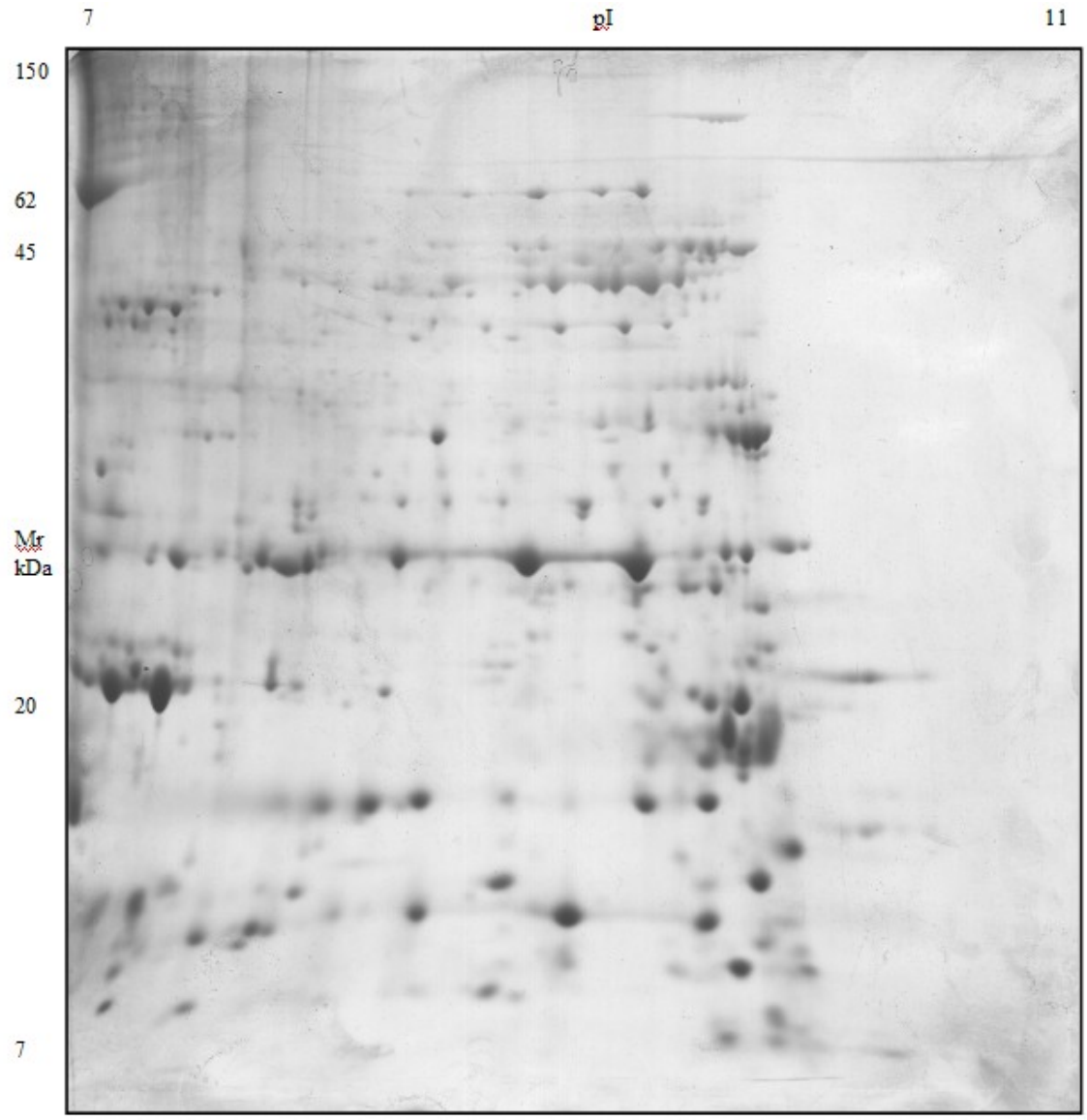


Figure A.18 *H. pylori* DU1 pH 6-11

DU1 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

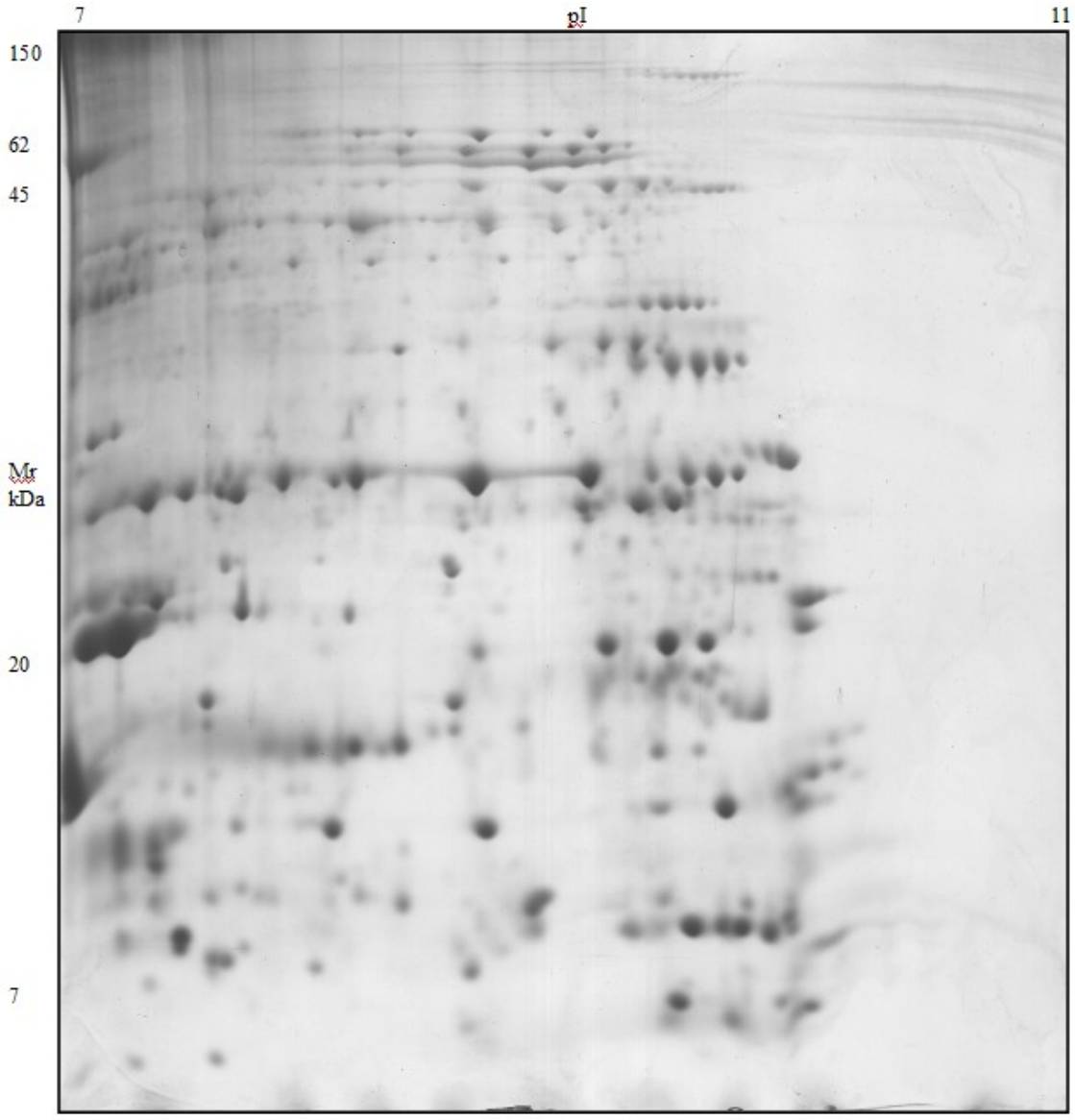


Figure A.19 *H. pylori* DU2 pH 6-11

DU2 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

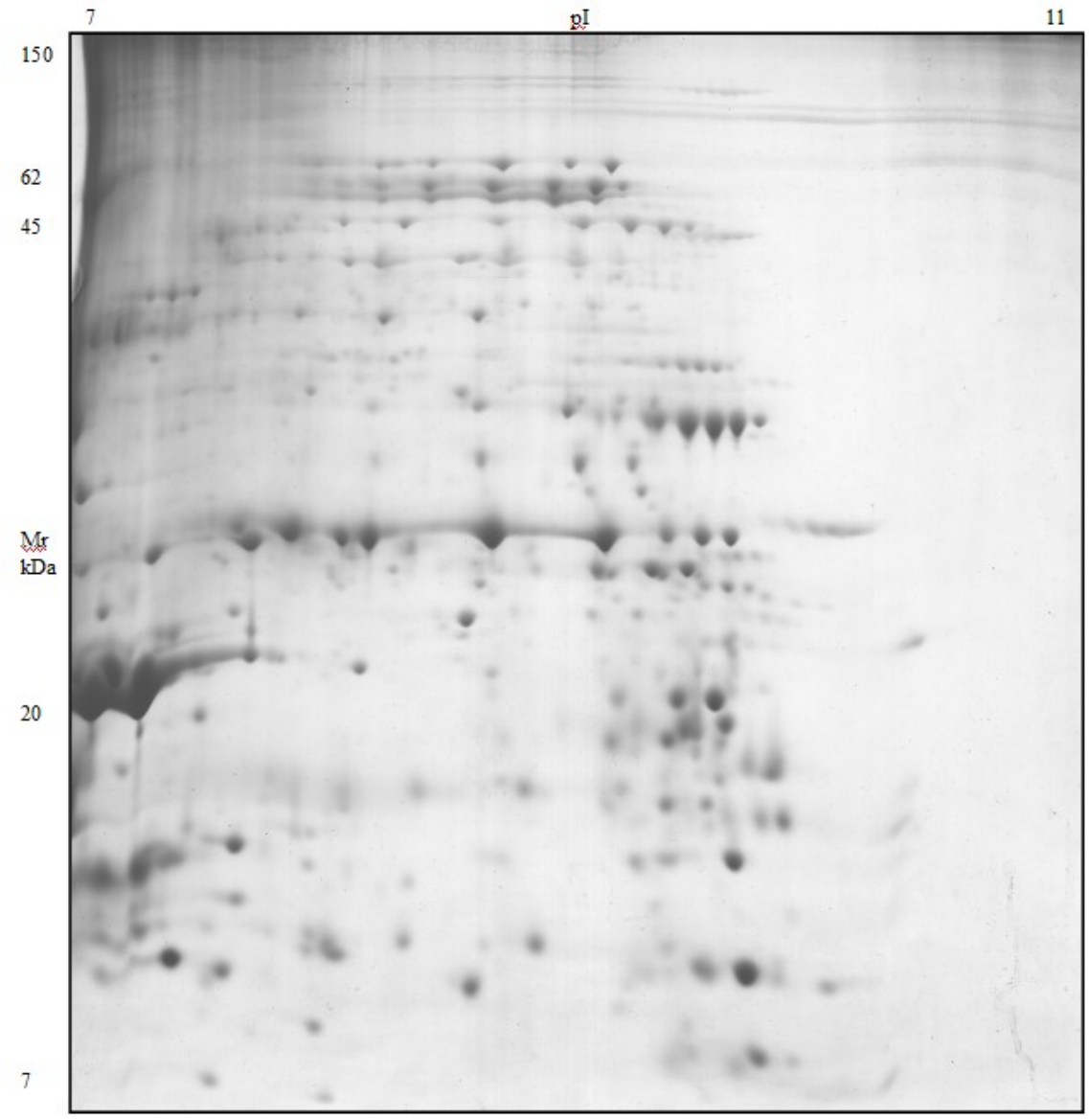


Figure A.20 *H. pylori* DU3 pH 6-11

DU3 proteins separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

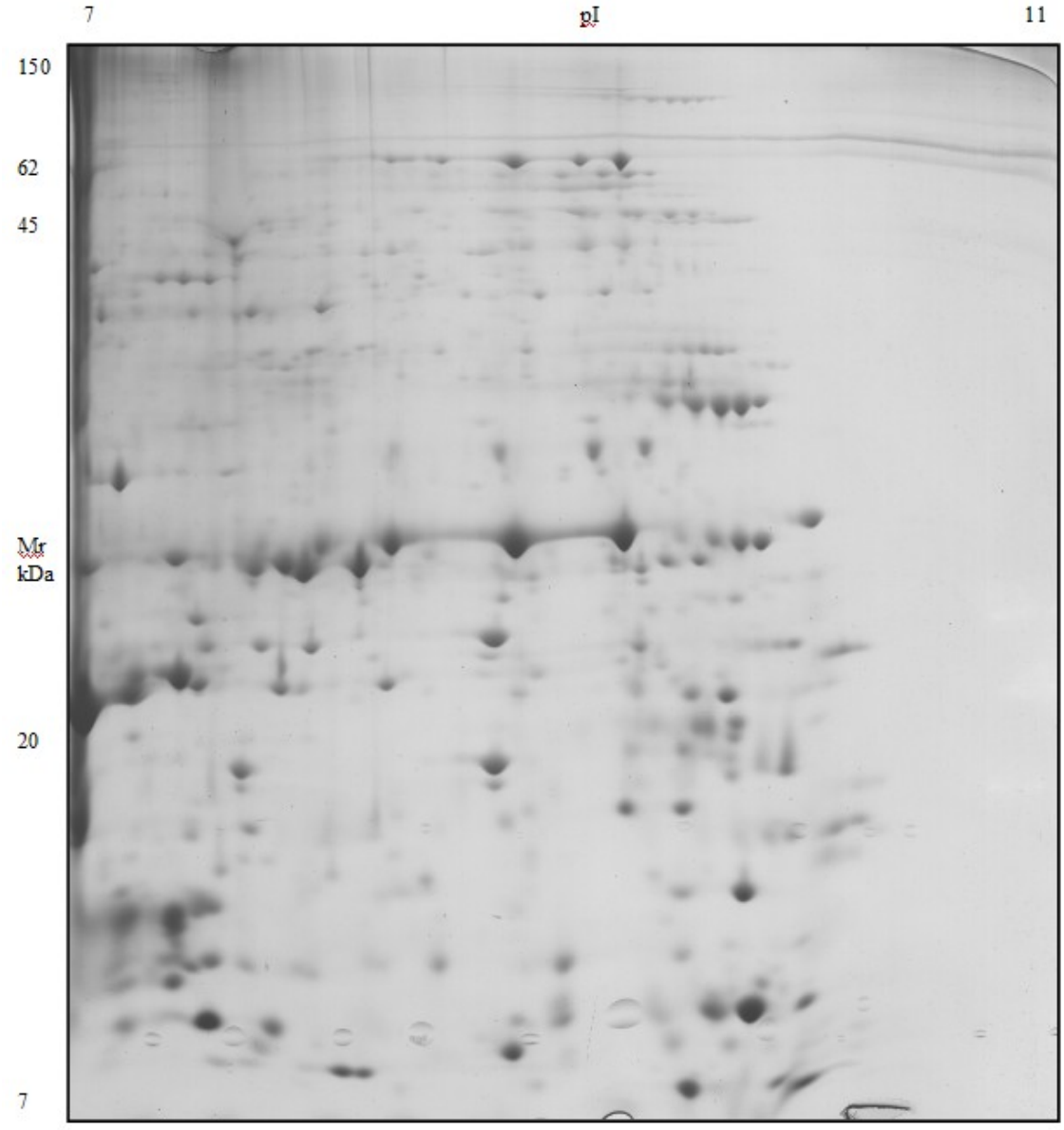


Figure A.21 *H. pylori* GI1 pH 6-11

GI1 proteins were separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

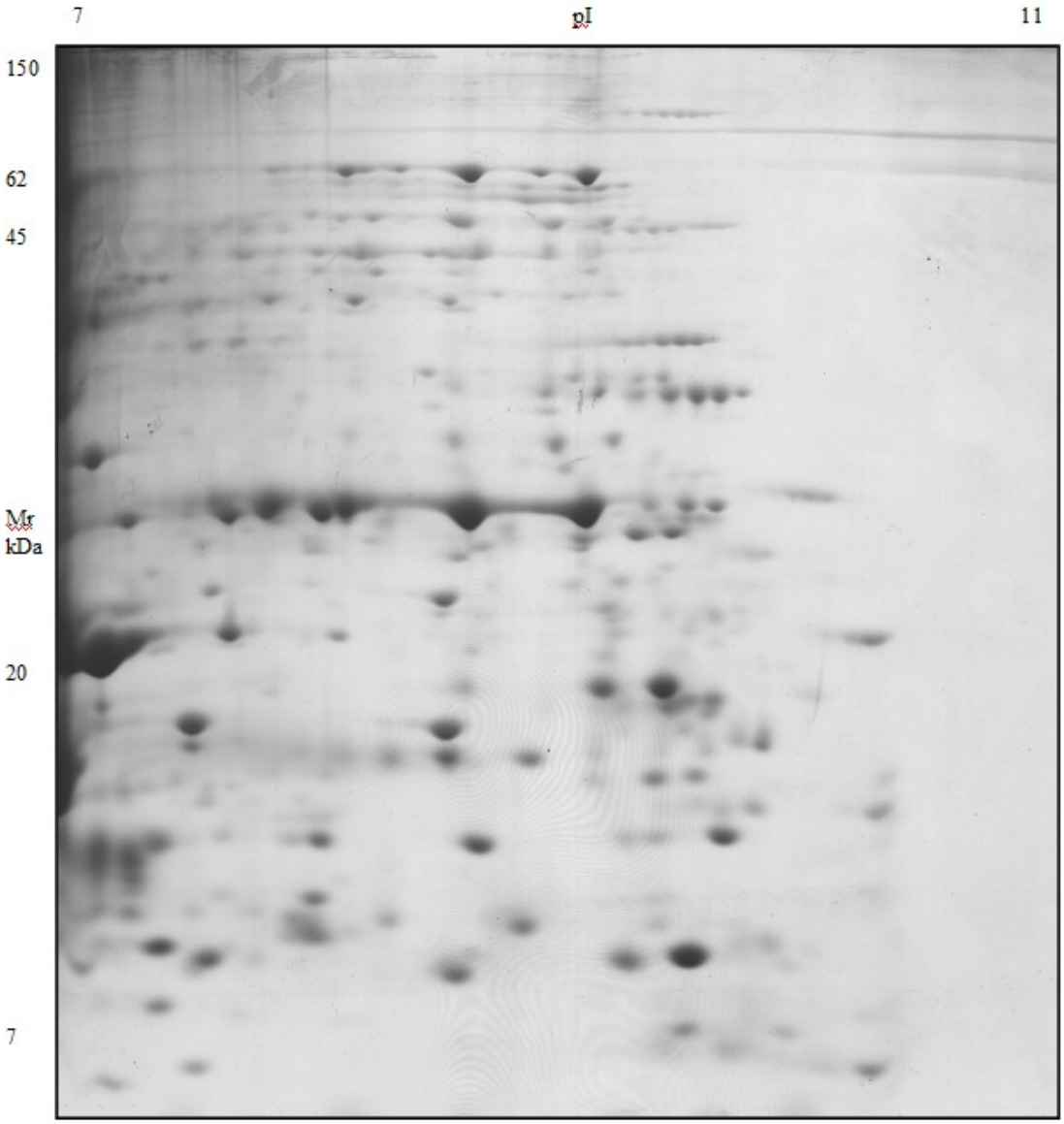


Figure A.22 *H. pylori* GI2 pH 6-11

GI2 proteins were separated on 8.5% polyacrylamide slab gels by a Tris-tricine electrode buffer. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

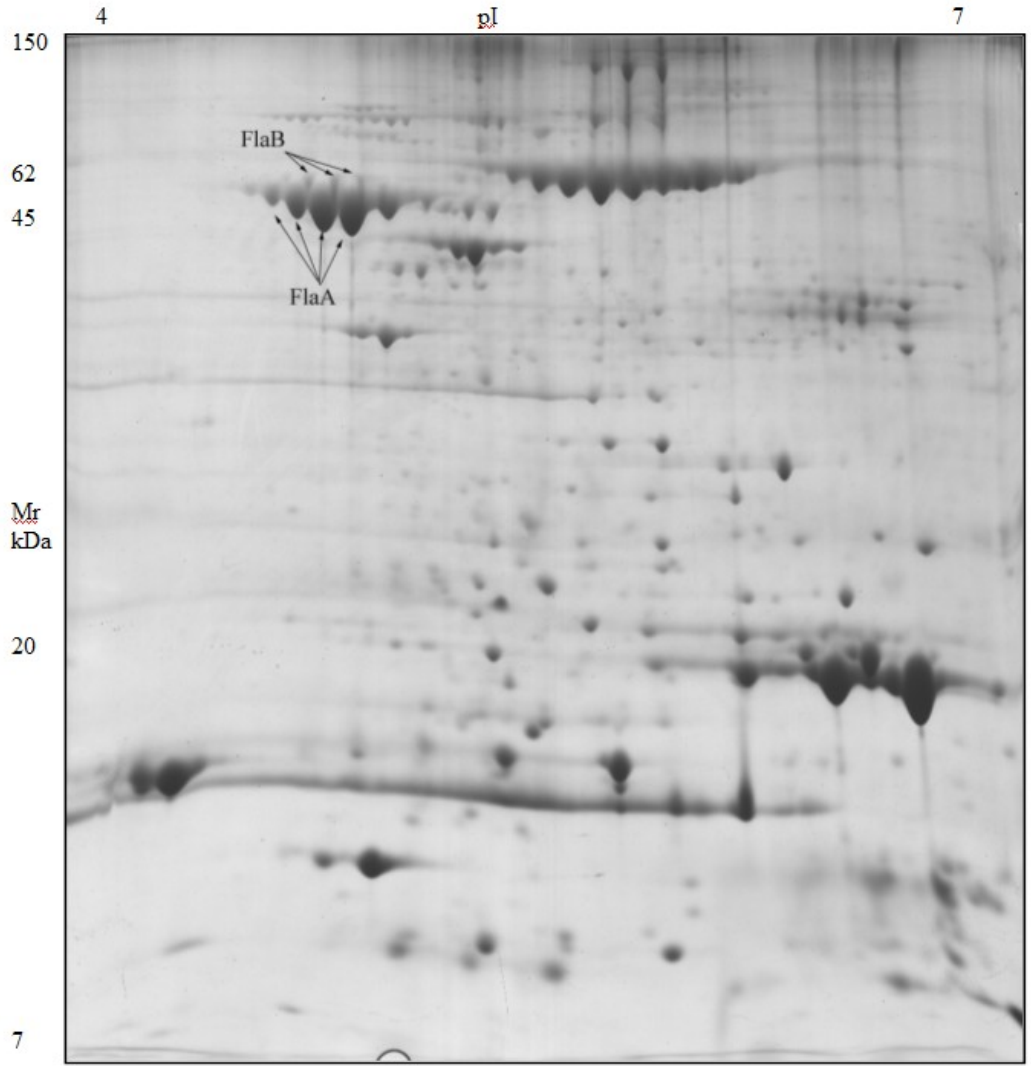


Figure A.23 NUD3 Identified Flagellin Spots

The NUD3 strain is an example of a *H. pylori* strain which had the highest concentration of FlaA, FlaB protein spots. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

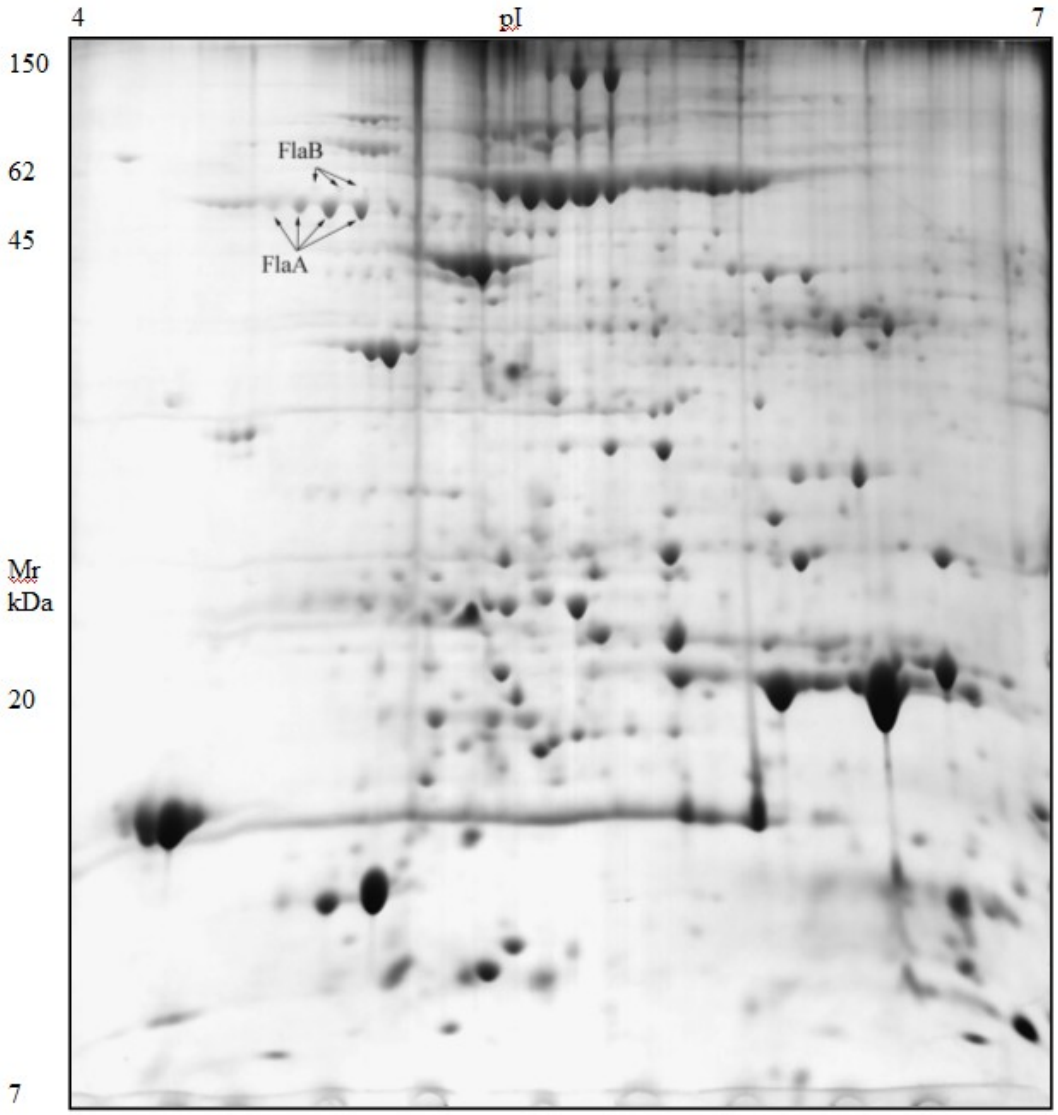


Figure A.24 NUD1 Identified Flagellin Spots

The NUD1 strain is an example of a *H. pylori* strain which had a medium level concentration of FlaA, FlaB protein spots. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

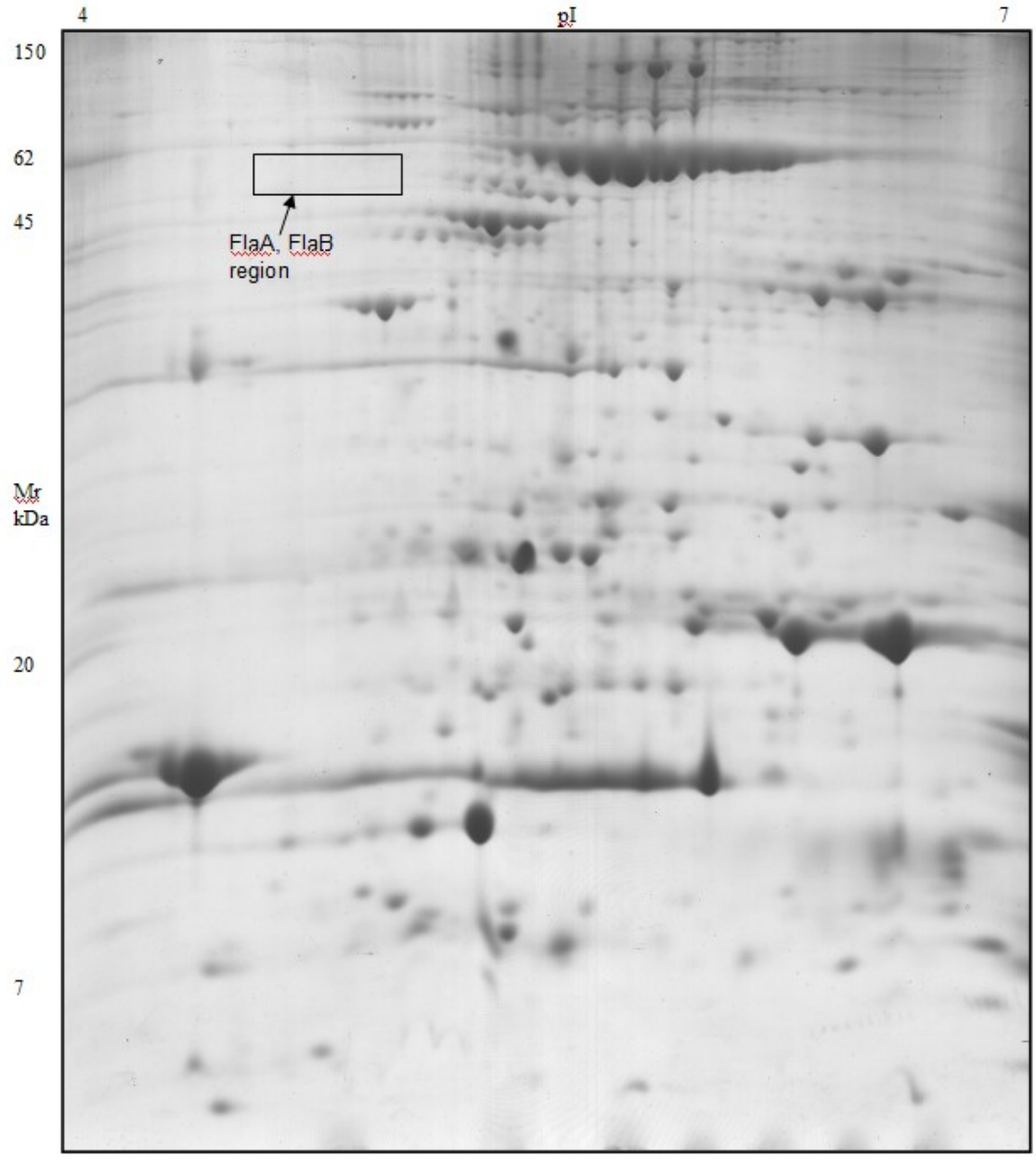


Figure A.25 GI2 Identified Flagellin Spots

The GI2 strain is an example of a *H. pylori* strain which displayed no FlaA, FlaB protein spots. Determination of the molecular mass and pI of protein spots was obtained from the use of 2-D protein standards.

Alm R A, Ling L S, Moir D T, King B L, Brown E D, Doig P C, Smith D R, Noonan B, Guild B C, deJonge B L, Carmel G, Tummino P J, Caruso A, Uria-Nickelsen M, Mills D M, Ives C, Gibson R, Merberg D, Mills S D, Jiang Q, Taylor D E, Vovis G F and Trust T J. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 1999 **397**:176-80.

Alm R A and Trust T J. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *Journal of Molecular Medicine* 1999 **77**:834-46.

Ames G F and Nikaido K. Two-dimensional gel electrophoresis of membrane proteins. *Biochemistry* 1976 **15**:616-23.

Anderson L and Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997 **18**:533-7.

Ando T, Israel D A, Kusugami K and Blaser M J. HP0333, a member of the *dprA* family, is involved in natural transformation in *Helicobacter pylori*. *Journal of Bacteriology* 1999 **181:**5572–5580.

Andrews S C, Robinson A K and Rodríguez-Quiñones F. Bacterial iron homeostasis. *FEMS Microbiology Reviews*. 2003 **27:**215-37.

Anonymous. *Campylobacter pylori* becomes *Helicobacter pylori*. *Lancet* 1989 **2**:1019-20.

Apostolopoulos P, Vafiadis-Zouboulis I, Tzivras M, Kourtessas D, Katsilambros N and Archimandritis A. *Helicobacter pylori* (*H pylori*) infection in Greece: the changing prevalence during a ten-year period and its antigenic profile. *BMC Gastroenterology* 2002 **2**:11.

Aras RA, Lee Y, Kim S K, Israel D, Peek R M Jr and Blaser M J. Natural variation in populations of persistently colonizing bacteria affect human host cell phenotype. *The Journal of Infectious Diseases* 2003 **188**:486-96.

Arber W. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiology Reviews* 2000 **24**:1-7.

Asaka M, Kimura T, Kudo M, Takeda H, Mitani S, Miyazaki T, Miki K and Graham D Y. Relationship of *Helicobacter pylori* to serum pepsinogens in an asymptomatic Japanese population. *Gastroenterology*. 1992 **102**:760-6.

Azuma T, Ito Y, Suto H, Ohtani M, Dojo M, Muramatsu A, Kuriyama M and Kato T. The effect of *Helicobacter pylori* eradication therapy on dyspepsia symptoms in industrial workers in Japan. *Alimentary Pharmacology and Therapy* 2001 **15**:805-11.

Baddour L M. Virulence factors among gram-positive bacteria in experimental endocarditis. *Infection and Immunity* 1994 **62:**2143–2148.

Bae S H, Harris A G, Hains P G, Chen H, Garfin D E, Hazell S L, Paik Y K and Walsh B J, Cordwell S J. Strategies for the enrichment and identification of basic proteins in proteome projects. *Proteomics* 2003 **3**:569-79.

Balakrishnan H, Srinivasan M C and Rele M V. Extracellular protease activities in relation to xylanase secretion in an alkalophilic *Bacillus* sp. *Biotechnology Letters* 1997 **19**:599–601.

Bandyopadhyay A K and Sonawat H M. Salt dependent stability and unfolding of [Fe2-S2] ferredoxin of *Halobacterium salinarum*: spectroscopic investigations. *Biophysical Journal* 2000 **79**:501-10.

Barabino A. *Helicobacter pylori*-related iron deficiency anaemia: a review. *Helicobacter* 2002 **7**:71-5.

Bardhan P K. Epidemiological features of *Helicobacter pylori* infection in developing countries. *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America* 1997 **25**:973-8.

Bastawde K B. Xylan structure, microbial xylanases, and their mode of action. *World Journal of Microbiology and Biotechnology* 1992 **8**:353–368.

Berkelman T, Stenstedt T Bjellqvist B, Laird N, McDowell M, Olsson I and Westermeier R. 2D electrophoresis using immobilised pH gradients: principles and methods. Amersham Biosciences 1998.

Bereswill S, Lichte F, Vey T, Fassbinder F and Kist M. Cloning and characterization of the *fur* gene from *Helicobacter pylori*. *FEMS Microbiology Letters* 1998a **159:**193–200.

Bereswill S, Waidner U, Odenbreit S, Lichte F, Fassbinder F, Bode G and Kist M. Structural, functional and mutational analysis of the pfr gene encoding a ferritin from *Helicobacter pylori*. *Microbiology* 1998b **144**:2505-16.

Berggren M, Gallegos A, Gasdaska J R, Gasdaska P Y, Warneke J and Powis G. Thioredoxin and thioredoxin reductase gene expression in human tumours and cell lines, and the effects of serum stimulation and hypoxia. *Anticancer Research* 1996 **16**:3459-66.

Bernhardt J, Buttner K, Scharf C and Hecker M. Dual channel imaging of two-dimensional electropherograms in *Bacillus subtilis*. *Electrophoresis*. 1999 **20**:2225-40.

Bjellqvist B, Ek K, Righetti P G, Gianazza E, Gőrg A, Westermeier R and Postel W. Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods* 1982 **6**:317-39.

Blaser L. Measured breath. *Environmental Health Perspectives* 1996 **104**: 1292-4.

Blaser M J. Helicobacter pylori and the pathogenesis of gastroduodenal inflammation. *Journal of Infectious Diseases* 1990 **161**:626-33.

Blaser M J. *Helicobacter pylori* eradication and its implications for the future. *Alimentary Pharmacology and Therapy* 1997 **11** Suppl 1:103-7.

Blattner F R, Plunkett G, Bloch C A, Perna NT, Burland V, Riley, M, Collado-Vides J, Glasner J D, Rode C K, Mayhew G F, Gregor J, Davis N W, Kirkpatrick H A, Goeden M A, Rose D J, Mau B, Shao Y. The complete genome sequence of *Escherichia coli* K-12. *Science* 1997 **277:** 1453–1474.

Bode G, Mauch F and Malfertheiner P. The coccoid forms of *Helicobacter pylori*. Criteria for their viability. *Epidemiology and Infection*. 1993 **111**:483-90.

De Boer P A. J, Crossley R E, and Rothfield L I. Roles of MinC and MinD in the Site-Specific Septation Block Mediated by the MinCDE System of *Escherichia coli.* *Journal of Bacteriology* 1992 **174**:63-70.

Bourne, H R, Sanders, D A and McCormick F. The GTPase superfamily: a conserved switch of diverse cell functions. *Nature* 1990 **348**:125–132.

Broder S and Venter J C. Whole genomes: the foundation of new biology and medicine. *Current Opinion in Biotechnology* 2000 **11**:581-5.

Brooke L, Urquhart I, Cordwell S J and Humphery-Smith I. Comparison of Predicted and Observed Properties of Proteins Encoded in the Genome of *Mycobacterium Tuberculosis* H37Rv. *Biochemical and Biophysical Research Communications* 1998 **253:**70–79.

Buck M, McCloskey J A, Basile B and Ames B N**.** *cis* 2-Methylthioribosylzeatin (ms^2io^6A) is present in the transfer RNA of *Salmonella typhimurium*, but not *Escherichia coli*. *Nucleic Acids Research* 1982 **10:**5649–5662.

Calam J. Clinicians Guide to *Helicobacter pylori.* Chapman and Hall Medical London 1996 pages 1-92.

Cadenas E. Biochemistry of oxygen toxicity. *Annual Review of Biochemistry* 1989 **58**:79-110.

Cantoni. Activation of methionine for transmethylation. *Journal of Biological Chemistry* 1951 **189**:745-754.

Caro A A and Cederbaum A I. Antioxidant properties of S-adenosyl-L-methionine in Fe(2+)-initiated oxidations. *Free Radical Biology & Medicine* 2004 **36**:1303-16.

Censini S, Lange C, Xiang Z Crabtree J E, Ghiara P, Borodovsky M, Rappuoli R, Covacci A. *cagA*, a pathogenicity island of *Helicobacter pylori*, encodes type-I specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences of the United States of America* 1996 **93**:14648–53.

Chanto G, Occhialini A, Gras N, Alm R A, Megraud F and Marais A. Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of *Helicobacter pylori. Microbiology* 2002 **148**:3671-3680.

Charbonneau H and Tonks N K. 1002 protein phosphatases? *Annual Review of Cell Biology* 1992 **8**:463-493.

Choi I Y, Sup K I, Kim H J and Park J W. Thermosensitive phenotype of *Escherichia coli* mutant lacking NADP+-dependent isocitrate dehydrogenase. *Redox Report: Communications in Free Radical Research* 2003 **8**:51-6.

Chung Y S, Breidt F and Dubnau D. Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of Bacillus subtilis. *Molecular Microbiology*. 1998 **29**:905-13.

Chung Y S and Dubnau D. All seven *comG* open reading frames are required for DNA binding during transformation of competent Bacillus subtilis. *Journal of Bacteriology* 1998 **180**:41-5.

Clements M O and Foster S J. Starvation recovery of *Staphylococcus aureus* 8325-4. *Microbiology*. 1998 **144**:1755-63.

Cohen G N. The aspartokinases and homoserine dehydrogenase of *Escherichia coli*. *Current Topics in Cellular Regulation* 1969 **1**:183-231.

Colland F, Rain J C, Gounon P, Labigne A, Legrain P and De Reuse H. Identification of the *Helicobacter pylori* anti-sigma28 factor. *Molecular Microbiology* 2001 **41**:477-87.

Colledge M and Scott J D. AKAPs: from structure to function. *Trends in Cell Biology* 1999 **9**:216-21.

Cohen J C. Evolving therapies for peptic ulcer disease: *Helicobacter pylori* treatment. *Gastroenterologist*. 1995 **3**:289-300.

Connolly D M and Winkler M E. Genetic and physiological relationships among the *miaA* gene, 2-methylthio-N6-(delta 2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in *Escherichia coli* K-12. *Journal of Bacteriology* 1989 **171**:3233–3246.

Contreras M and Thiberge J M, Mandrand-Berthelot M A and Labigne A. Characterization of the roles of NikR, a nickel-responsive pleiotropic autoregulator of *Helicobacter pylori.* *Molecular Microbiology* 2003 **49**:947-63.

Cordwell S J, Larsen M R, Cole RT and Walsh B J. Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. *Microbiology* 2002 **148**:2765-81.

Correa P. *Helicobacter pylori* infection and gastric cancer. *Cancer Epidemiology Biomarkers and Prevention* 2003 **12**:238s-241.

Cremer F and van de Walle C. Method for extraction of proteins from green plant tissues for two-dimensional polyacrylamide gel electrophoresis. *Analytical Biochemistry* 1985 **147**:22-6.

De Ungria M C, Kolesnikow T, Cox P T and Lee A. Molecular characterization and interstrain variability of pHPS1, a plasmid isolated from the Sydney strain (SS1) of *Helicobacter pylori*. *Plasmid* 1999 **41**:97–109.

Delany I, Spohn G, Rappuoli R and Scarlato V. The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*. *Molecular Microbiology* 2001a **42**:1297–1309.

Delany I, Pacheco AB, Spohn G, Rappuoli R and Scarlato V. Iron-dependent transcription of the *frpB* gene of *Helicobacter pylori* is controlled by the Fur repressor protein. *Journal of Bacteriology* 2001b **183**:4932-7.

Demuth D R and Irvine D C. Structural and functional variation within the alanine-rich repetitive domain of streptococcal antigen I/II. *Infection and Immunity* 2002 **70**: 6389-98.

de Vries N, van Ark E M, Stoof J, Kuipers E J, van Vliet A H and Kusters J G. The stress-induced *hsp12* gene shows genetic variation among *Helicobacter pylori* strains. *FEMS Immunology and Medical Microbiology* 2003 **38**:45-51.

Dhaenens L, Szczebara F and Husson M O. Identification, characterization, and immunogenicity of the lactoferrin-binding protein from *Helicobacter pylori*. *Infection and Immunity*. 1997 **65**:514-8.

Dixon M F. Patterns of inflammation linked to ulcer disease. *Bailliere's Best Practice & Research. Clinical Gastroenterology* 2000 **14**:27-40.

Doig P, de Jonge B L, Alm R A, Brown E D, Uria-Nickelsen M, Noonan B, Mills S D, Tummino P, Carmel G, Guild B C, Moir D T, Vovis G F and Trust T J. *Helicobacter pylori* physiology predicted from genomic comparison of two strains. *Microbiology and Molecular Biology Reviews* 1999 **63**:675-707.

Driessen A J M, Rosen B P and Konings W N. Diversity of transport mechanisms: common structural principles. *Trends and Current Opinions* 2000 **25**:397-401.

Dunn B E, Cohen H and Blaser M J. *Helicobacter pylori. Clinical Microbiology Review* 1997 **10**:720-41.

Dwek M V, Ross H A and Leathem A J. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. *Proteomics* 2001 **1**:756-62.

Eaton K A, Morgan D R, and Krakowka S. *Campylobacter pylori* virulence factors in gnotobiotic piglets. *Infection and Immunity* 1989 **57**: 1119–1125.

Eaton K A, Suerbaum S, Josenhans C and Krakowka S. Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. *Infection and Immunity*. 1996 **64**:2445-8.

Enroth H, Akerlund T, Sillen A and Engstrand L. Clustering of clinical strains of *Helicobacter pylori* analysed by two-dimensional gel electrophoresis. *Clinical Diagnosis of Laboratory and Immunology*. 2000 **7**:301-6.

Ernst F D, Bereswill S, Waidnet B, Stoof J, Mader U, Kusters J G, Kuipers E J, Kist M, van Vliet A H M and Homuth G. Transcriptional profiling of *Helicobacter pylori* Fur- and iron-regulated gene expression. *Microbiology* 2005 **151**:533-546.

Esberg B, Leung H C, Tsui H C, Bjork G R and Winkler M E. Identification of the miaB gene, involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. *Journal of Bacteriology* 1999 **181**:7256-65.

Evans DJ Jr, Evans D G, Lampert HC and Nakano H. Identification of four new prokaryotic bacterioferritins, from *Helicobacter pylori*, *Anabaena variabilis*, *Bacillus subtilis* and *Treponema pallidum*, by analysis of gene sequences. *Gene* 1995 **153**:123-7.

Everitt B. Cluster analysis is a generic term for a wide range of numerical methods for examining data. *Statistical Methods in Medical Research* 2004 **13**:343-5.

Everitt B S. Cluster analysis: a brief discussion of some of the problems. *The British Journal of Psychiatry : the Journal of Mental Science* 1972 **120**:143-5.

Farr S, Touati D and Kogoma T. Effects of oxygen stress on membrane functions in *Escherichia coli* : Role of HPI catalase. *Journal of Bacteriology* 1988 **170**:1837-1842.

Fiedorek S C, Malaty H M, Evans D L, Pumphrey C L, Casteel H B, Evans D J Jr and Graham D Y. Factors influencing the epidemiology of *Helicobacter pylori* infection in children. *Pediatrics*. 1991 **88**:578-82.

Figura N, Trabalzini L, Mini R, Bernardini G, Scaloni A, Talamo F, Lusini P, Ferro E, Martelli P and Santucci A. Inactivation of *Helicobacter pylori cagA* gene affects motility. *Helicobacter*. 2004 **9**:185-93.

Fischer W and Haas R. The RecA protein of *Helicobacter pylori* requires a posttranslational modification for full activity. *Journal of Bacteriology* 2004 **186**:777-84.

Flengsrud R and Kobro G. A method for two-dimensional electrophoresis of proteins from green plant tissues. *Analytical Biochemistry* 1989 **177**:33-6.

Fountoulakis M, Juranville J F, Roder D, Evers S, Berndt P and Langen H. Reference map of the low molecular mass proteins of *Haemophilus influenzae*. *Electrophoresis*. 1998 **19**:1819-27.

Frazier B A, Pfeifer J D, Russell D G, Falk P, Olsen A N, Hammar M, Westblom T U and Normark S J. Paracrystalline inclusions of a novel ferritin containing nonhaeme iron, produced by the human gastric pathogen *Helicobacter pylori*: evidence for a third class of ferritins. *Journal of Bacteriology* 1993 **175**:966-72.

Frick D N and Richardson C C. DNA primases. *Annual Review of Biochemistry* 2001 **70**:39-80.

Fung W P. *Helicobacter pylori*: the first description. *The American Journal of Gastroenterology* 1995 **90**:2070.

Galvani M, Rovatti L, Hamdan M, Herbert B and Righetti P G. Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? *Electrophoresis*. 2001 **22**:2066-74.

Geis G, Suerbaum S, Forsthoff B, Leying H and Opferkuch W. Ultrastructure and biochemical studies of the flagellar sheath of *Helicobacter pylori*. *Journal of Medical Microbiology* 1993 **38**:371-7.

Glupczynski Y, Bourdeaux L, Verhas M, DePrez C, DeVos D and Devreker T. Use of a urea breath test versus invasive methods to determine the prevalence of *Helicobacter pylori* in Zaire. *European Journal of Clinical Microbiology and Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 1992 **1**: 322-7.

Goodwin C S. How *Helicobacter pylori* acquired its name, and how it overcomes gastric defence mechanisms. *Journal of Gastroenterology and Hepatology* 1994 **9**:S1-3.

Goodwin C S and Armstrong J A. Microbiological aspects of *Helicobacter pylori* (*Campylobacter pylori*). *European Journal of Clinical Microbiology and Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*. 1990 **9**:1-13.

Goodwin C S and Worsley B W. *Helicobacter pylori* Biology and Clinical Practice CRS Press Boca Raton, Florida (1993a) pages 1-82.

Goodwin C S and Worsley B W. Microbiology of *Helicobacter pylori*. *Gastroenterology Clinics of North America* 1993b **22**:5-19.

Gorelik E, Galili U and Raz A. On the role of cell surface carbohydrates and their binding proteins (lectins) in tumour metastasis. *Cancer Metastasis Reviews* 2001 **20**:245-77.

Gőrg A, Obermaier C, Boguth G, Csordas A, Diaz J J and Madjar J J. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis* 1997 **18**:328-37.

Görg A, Weiss W and Dunn M J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 2004 **4**:3665–3685.

Govorun V M, Moshkovskii S A, Tikhonova O V, Goufman E I, Serebryakova M V, Momynaliev K T, Lokhov P G, Khryapova E V, Kudryavtseva L V, Smirnova O V, Toropyguine I Y, Maksimov B I and Archakov A I. Comparative Analysis of Proteome Maps of *Helicobacter pylori* Clinical Isolates. *Biochemistry (Mosc)* 2003 **68**:42-9.

Graham D Y, Adam E, Reddy G T, Agarwal J P, Agarwal R, Evans D J Jr, Malaty H M and Evans D G. Seroepidemiology of *Helicobacter pylori* infection in India. Comparison of developing and developed countries. *Digestive Diseases and Sciences* 1991a **36**:1084-8.

Graham D Y, Malaty H M, Evans D G, Evans D J Jr, Klein P D and Adam E. Epidemiology of *Helicobacter pylori* in an asymptomatic population in the United States. Effect of age, race, and socioeconomic status. *Gastroenterology* 1991b **100**:1495-501.

Graham D Y and Yamaoka Y. Disease-specific *Helicobacter pylori* virulence factors: the unfulfilled promise. *Helicobacter* 2000 **5**:S3-9.

Gupta S, Bhushan B and Hoondal G S. Enhanced production of xylanase from *Staphylococcus* sp. SG-13 using amino acids. *World Journal of Microbiology and Biotechnology* 1999 **15**:511–512.

Gupta S, Bhushan B and Hoondal G S. Isolation, purification and characterization of xylanase from *Staphylococcus* sp. SG-13 and its application in biobleaching of kraft pulp. *Journal of Applied Microbiology* 2000 **88**:325–334.

Guy G R, Philip R and Tan Y H. Analysis of cellular phosphoproteins by two-dimensional gel electrophoresis: applications for cell signalling in normal and cancer cells. *Electrophoresis*. 1994 **15**:417-40.

Harris A G, Hinds F E, Beckhouse A G, Kolesnikow T and Hazell S L. Resistance to hydrogen peroxide in *Helicobacter pylori*: role of catalase (KatA) and Fur, and functional analysis of a novel gene product designated 'KatA-associated protein', KapA (HP0874). *Microbiology* 2002 **148**:3813-25.

Harrison P M and Arosio P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochimica et Biophysica Acta* 1996 **1275**:161-203.

Hashimoto W, Suzuki H, Nohara S, Tachi H, Yamamoto K and Kumagai H Subunit association of gamma-glutamyltranspeptidase of *Escherichia coli* K-12. *Journal of Biochemistry* (Tokyo). 1995 **118**:1216-23.

He X Y, Zhuang Y H, Zhang X G and Li G L. Comparative proteome analysis of culture supernatant proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra. *Microbes Infections* 2003 **5**:851-6.

Herbert B. Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis*. 1999 **20**:660-3.

Herbert B R, Molloy M P, Gooley A A, Walsh B J, Bryson W G and Williams K L. Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. *Electrophoresis*. 1998 **19**:845-51.

Hespell R B and Whitehead T R. Physiology and genetics of xylan degradation by gastrointestinal tract bacteria. *Journal of Dairy Science* 1990 **73**:3013-22.

Higashikubo A, Tanaka N, Noda N, Maeda I, Yagi K, Mizoguchi T and Nanri H. Increase in thioredoxin activity of intestinal epithelial cells mediated by oxidative stress. *Biological and Pharmaceutical Bulletin* 1999 **22**:900-3.

Higgins C F, Haag P D, Nikaido K, Ardeshir F, Garcia G, and Ames G L. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. *Nature* 1982 **298**:723–727.

Hoffman PS, Goodwin A, Johnsen J, Magee K and Veldhuyzen van Zanten S J. Metabolic activities of metronidazole-sensitive and -resistant strains of Helicobacter pylori: repression of pyruvate oxidoreductase and expression of isocitrate lyase activity correlate with resistance. *Journal of Bacteriology* 1996 **178**:4822-9.

Hofreuter D, Odenbreit S, Henke G, and Haas R. Natural competence for DNA transformation in *Helicobacter pylori*: identification and genetic characterization of the *comB* locus. *Molecular Microbiology*. 1998 **28**:1027–1038.

Hofreuter D and Haas R. Characterisation of two cryptic *Helicobacter pylori* a putative source for horizontal gene transfer and gene shuffling. *Journal of Bacteriology* 2002 **184**:2755-2766.

Holcombe C, Kaluba J and Lucas S B. Non-ulcer dyspepsia in Nigeria: a case-control study. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 1991 **85**:553-5.

Holcombe C, Omotara B A, Eldridge J and Jones D M. *H. pylori*, the most common bacterial infection in Africa: a random serological study. *American Journal of Gastroenterology* 1992 **87**:28-30.

Holger K, Burns B P, Angermüller K, Odenbreit S, Fischer W, Melchers K and Haas R. Identification and Characterization of *Helicobacter pylori* Genes Essential for Gastric Colonization. *The Journal of Experimental Medicine*, 2003 **197**:813-822.

Holmgren A. Thioredoxin. *Annual Review of Biochemistry* 1985 **54**:237-71.

Holms W H. Control of flux through the citric acid cycle and the glyoxylate bypass in *Escherichia coli. Biochemical Society symposium* 1987 **54**:17-31.

Hrmova M, Biely P, Vrsanka M and Petrakova E. Induction of cellulose-and xylan-degrading enzyme complex in yeast *Trichosporon cutaneum*. *Archive of Microbiology* 1984 **161:**371-376.

Hu L T and Mobley H L. Expression of catalytically active recombinant *Helicobacter pylori* urease at wild-type levels in *Escherichia coli*. *Infection and Immunology* 1993 **61**:2563-9.

Humphrey-Smith I, Cordwell S J and Blackstock W P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. *Electrophoresis* 1997 **18:**1217-1242.

Hunter T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling. *Cell* 1995 **80**:225-36.

Hussell T, Isaacson P G, Crabtree J E and Spencer J. The response of cells from low-grade B-cell gastric lymphomas of mucosa-associated lymphoid tissue to *Helicobacter pylori*. *Lancet* 1993 **342**:571-4.

Husson M O, Legrand D, Spik G and Leclerc H. Iron acquisition by *Helicobacter pylori*: importance of human lactoferrin. *Infection and Immunity* 1993 **61**:2694-7.

Ikenoue T, Maeda S, Ogura K, Akanuma M, Mitsuno Y, Imai Y, Yoshida H, Shiratori Y and Omata M. Determination of *Helicobacter pylori* virulence by simple gene analysis of the cag pathogenicity island. *Clinical and Diagnostic Laboratory Immunology* 2001 **8**:181-6.

Isobe M, Uyakul D, Liu K L, Goto T. FAB-MS/MS spectrometry in determining the primary structure of gamma-glutamyl-containing peptides. *Agricultural and Biological Chemistry* 1990 **54:**1651-60.

Ito M, Haruma K, Kaya S, Kamada T, Kim S, Sasaki A, Sumii M, Tanaka S, Yoshihara M, Wagner S and Chayama K. Serological comparison of serum pepsinogen and anti-parietal cell antibody levels between Japanese and German patients. *European Journal of Gastroenterology & Hepatology* 2002 **14**:123-7.

James P, Quadroni M, Carafoli E, Gonnet G. Protein identification by mass profile fingerprinting**.** *Biochemical and biophysical research communications* 1993 **195:**58-64.

Jeon C, Yoon H and Agarwal K. The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 1994 **91**:9106-10.

Jiang Q, Hiratsuka K and Taylor D E. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Molecular Microbiology* 1996 **20**:833-42.

Jin J B, Bae H, Kim S J, Jin Y H, Goh C H, Kim D H, Lee Y J, Tse Y C, Jiang L and Hwang I. The Arabidopsis Dynamin-Like Proteins ADL1C and ADL1E Play a Critical Role in Mitochondrial Morphogenesis. *The Plant Cell* 2003 **15**:2357–2369.

Jones D M, Lessells A M and Eldrige J. Campylobacter like organisms on the gastric mucosa: culture, histological, and serological studies. *Journal of Clinical Pathology* 1984 **37**:1002-6.

Josenhans C, Niehus E, Amersbach S, Horster A, Betz C, Drescher B, Hughes K T and Suerbaum S. Functional characterization of the antagonistic flagellar late regulators FliA and FlgM of *Helicobacter pylori* and their effects on the *H. pylori* transcriptome. *Molecular Microbiology* 2002 **43**:307–322.

Jungblut P R, Schaible U E, Mollenkopf H J, Zimny-Arndt U, Raupach B, Mattow J, Halada P, Lamer S, Hagens K and Kaufmann S H. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Molecular Microbiology* 1999 **33**:1103-17.

Jungblut P R, Bumann D, Haas G, Zimney-Ardnt U, Holland P, Lamer S, Siejak F, Aebischer A and Meyer T F. Comparative proteome analysis of *H. pylori. Molecular Microbiology* 2000 **36**:710-725.

Kather B, Stingl K, van der Rest M E, Altendorf K and Molenaar D. Another unusual type of citric acid cycle enzyme in *Helicobacter pylori*: the malate:quinone oxidoreductase. *Journal of Bacteriology* 2000 **182**:3204-9.

Kersten B, Burkle L, Kuhn E J, Giavalisco P, Konthur Z, Lueking A, Walter G, Eickhoff H and Schneider U. Large-scale plant proteomics. *Plant Molecular Biology* 2002 **48**:133-41.

Kettman J R, Frey J R and Lefkovits I. Proteome, transcriptome and genome: top down or bottom up analysis? *Biomolecular Engineering* 2001 **18**:207-12.

Khasin A, Alchanati I and Shoham Y. Purification and characterization of a thermostable xylanase from *Bacillus stearothermophilus* T-6. *Applied and Environmental Microbiology* 1993 **59**:1725-30.

Kirschner M. Intracellular proteolysis. *Trends in Cell Biology* 1999 **9**:M42-5.

Kirchhausen T. Vesicle formation: dynamic dynamin lives up to its name. *Current Biology* 1998 **8**:R792–R794.

Kleanthous H, Clayton C L and Tabaqchali S. Characterization of a plasmid from *Helicobacter pylori* encoding a replication protein common to plasmids in gram-positive bacteria. *Molecular Microbiology* 1991 **5**:2377-89.

Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* 1975 **26**:231-43.

Koonin E V. A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends in Biochemical Sciences* 1996 **21**:242–243.

Krah A, Schmidt F, Becher D, Schmid M, Albrecht D, Rack A, Büttner K and Jungblut P R. Analysis of Automatically Generated Peptide Mass Fingerprints of Cellular Proteins and Antigens from *Helicobacter pylori* 26695 Separated by Two-dimensional Electrophoresis. *Molecular and Cellular Proteomics* 2003 **2**:1271-1283.

Kreitzer G, Marmorstein A, Okamoto P, Vallee R, and Rodriguez-Boulan E. Kinesin and dynamin are required for post-Golgi transport of a plasma-membrane protein. *Nature Cell Biology* 2000 **2**:125–127.

Krishna R G and Wold F. Post-translational modification of proteins. *Advances in Enzymology and Related Areas of Molecular Biology* 1993 **67**:265-98.

Krishnamurthy P, Parlow M H, Schneider J, Burroughs S, Wickland C, Vakil N B, Dunn, B E, and Phadnis H. Identification of a novel penicillin-binding protein from *Helicobacter pylori. Journal of Bacteriology* 1999 **181**:5107-5110.

Kruft V, Eubel H, Jänsch L, Werhahn W, and Braun H P. Proteomic Approach to Identify Novel Mitochondrial Proteins in Arabidopsis. *Plant Physiology* **127**:1694-1710.

Kulkarni N Shendye A and Rao M. Molecular and biotechnological aspects of xylanases. *FEMS Microbiology Reviews* 1999 **23**:411–456.

Kusnezow W and Hoheisel J D. Solid supports for microarray immunoassays. *Journal of Molecular Recognition* 2003 **16**:165-76.

Labenz J and Borsch G. Evidence for the essential role of *Helicobacter pylori* in gastric ulcer disease. *Gut* 1994 **35**:19-22.

Langen H, Takacs B, Evers S, Berndt P, Lahm H W, Wipf B, Gray C and Fountoulakis M. Two-dimensional map of the proteome of *Haemophilus influenzae. Electrophoresis* 2000 **21**:411-29.

Lee A., O'Rourke J, De Ungria M C, Robertson B, Daskalopoulos G, and Dixon M F. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. *Gastroenterology* 1997 **112:**1386-1397.

Lee S M, Koh H. J, Huh T L and Park, J W. Radiation sensitivity of an *Escherichia coli* mutant lacking NADP-dependent isocitrate dehydrogenase. *Biochemical and Biophysical Research Communications* 1999 **254:** 647–650.

Leiros I, McSweeney S and Hough E. The reaction mechanism of phospholipase D from *Streptomyces* sp. strain PMF. Snapshots along the reaction pathway reveal a pentacoordinate reaction intermediate and an unexpected final product. *Journal of Molecular Biology* 2004 **339**:805-20.

Levin L and Forschiassin F. Influence of growth conditions on the production of xylanolytic enzymes by *Trametes trogii*. *World Journal of Microbiology and Biotechnology* 1998 **14**:443–446.

Lindner C, Stulke J and Hecker M. Regulation of xylanolytic enzymes in *Bacillus subtilis*. *Microbiology* 1994 **140**:753-7.

Link A J, Hays L G, Carmack E B and Yates J R 3rd. Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC8143. *Electrophoresis* 1997 **18**:1314-34.

Link A. J. edited 2-DE Proteome Analysis Protocols. *Methods in Molecular Biology* vol 112 Humana Press New Jersey 1999.

Liu W, Zhu W, Lu Y, Kong Y and Ma G. Production, partial purification and characterization of xylanase from *Trichosporon cutaneum* SL409. *Process Biochemistry* 1998 **33**:331–326.

Lock R A, Cordwell S J, Coombs G W, Walsh B J and Forbes J M. Proteome analysis of *Helicobacter pylori*: major proteins of type strain NCTC 11637. *Pathology* 2001 **33**: 365-74.

Locke V L, Gibson T S, Thomas T M, Corthals G L and Rylatt D B. Gradiflow as a prefractionation tool for two-dimensional electrophoresis. *Proteomics* 2002, **2**:1254–1260.

Lopez M F. Better approaches to finding the needle in a haystack: optimizing proteome analysis through automation. *Electrophoresis* 2000 **21**:1082-93.

Lopez M F and Pluskal M G. Protein micro- and macroarrays: digitizing the proteome. *Journal of chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 2003 **787**:19-27.

Madi A, Mikkat S, Ringel B, Ulbrich M, Thiesen H J and Glocker M O. Mass spectrometric proteome analysis for profiling temperature-dependent changes of protein expression in wild-type *Caenorhabditis elegans*. *Proteomics* 2003 **3**:1526-34.

Maeng O, Kim Y C, Shin H J, Lee J O, Huh T L, Kang K I, Kim Y S, Paik S G and Lee H. Cytosolic NADP(+)-dependent isocitrate dehydrogenase protects macrophages from LPS-induced nitric oxide and reactive oxygen species. *Biochemical and Biophysical Research Communications* 2004 **317**:558-64.

Manos J, Kolesnikow T and Hazell S L. An investigation of the molecular basis of the spontaneous occurrence of catalase-negative phenotype of *Helicobacter pylori*. *Helicobacter* 1998 **3**:28-38.

Marengo E, Robotti E, Antonucci F, Cecconi D, Campostrini N and Righetti P G. Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems. *Proteomics* 2005 **5**:654-66.

Marieb E. (1995) Human Anatomy and Physiology Third Edition.

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

Marshall B J and Warren J R. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1984 **1**:1311-5.

Martin-de-Argila C, Boixeda D, Canton R, Mir N, de Rafael L, Gisbert J, Arocena C, and Garcia P A. *Helicobacter pylori* infection in a healthy population in Spain. *European Journal of Gastroenterology & Hepatology* 1996 **8**:1165-8.

Matic I, Rayssiguier C and Radman M. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* 1995 **80:**507–515.

Mato J M, Alvarez L and Pajares M A. SAM synthesis: molecular mechanisms and clinical implications. *Pharmacology & Therapeutics* 1997 **73**:265– 280.

McColl K E. *Helicobacter pylori* and dyspepsia. Pro and against. *Digestive and Liver Disease : Official Journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 2000 **32**:S199-201.

McNulty C A and Watson D M. Spiral bacteria of the gastric antrum. *Lancet* 1984 **1**:1068-9.

Megraud F, Brassens-Rabbe M P, Denis F, Belbouri A and Hoa D Q. Seroepidemiology of *Campylobacter pylori* infection in various populations. *Journal of Clinical Microbiology* 1989 **27**:1870-3.

Minnis J A, Taylor T E, Knesek J E, Peterson W L and McIntire S A. Characterization of a 3.5-kbp plasmid from *Helicobacter pylori*. *Plasmid* 1995 **34**:22-36.

Mitra P, Pal A K, Basu D and Hati R N. A staining procedure using Coomassie brilliant blue G-250 in phosphoric acid for detection of protein bands with high resolution in polyacrylamide gel and nitrocellulose membrane. *Analytical Biochemistry* 1994 **223**: 327-9.

Molloy M. P, Herbert B R, Walsh B J, Tyler M I, Traini M and Sanchez J C, Hochstrasser D F, Williams K L and Gooley A. A Extraction of membrane proteins by differential solubilisation for separation using two-dimensional gel electrophoresis. *Electrophoresis* 1998 **19**:837-844.

Morris A J, Engebrecht J and Frohman M A. Structure and regulation of phospholipase D. *Trends in Pharmacological Sciences* 1996 **17**:182–185.

MS-Fit information website http://prospector.ucsf.edu/html/instruct/allman.htm #prot_MW.

Muller D R, Schindler P, Towbin H, Wirth U, Voshol H, Hoving S and Steinmetz M O. Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Analytical Chemistry* 2001 **73**:1927-34.

Naito Y and Yoshikawa T. Molecular and cellular mechanisms involved in *Helicobacter pylori*-induced inflammation and oxidative stress. *Free Radical Biology and Medicine* 2002 **33**:323-336.

Nakao K, Imoto I, Ikemura N, Shibata T, Takaji S, Taguchi Y, Misaki M, Yamauchi K and Yamazaki N. Relation of lactoferrin levels in gastric mucosa with *Helicobacter pylori* infection and with the degree of gastric inflammation. *American Journal of Gastroenterology* 1997 **92**:1005-11.

Neuhoff V, Arold N, Taube D and Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 1988 **9**:255-62.

Nishihara J C and Champion K M. Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* 2002 **23**:2203-15.

Nomura A, Stemmermann G N, Chyou P H, Kato I, Perez-Perez G I and Blaser M J. *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. *The New England Journal of Medicine* 1991 **325**:1132-6.

Nouwens A S, Willcox M D, Walsh B J and Cordwell S J. Proteomic comparison of membrane and extracellular proteins from invasive (PAO1) and cytotoxic (6206) strains of *Pseudomonas aeruginosa*. *Proteomics* 2002 **2**:1325-46.

Oakley B, Kirsch D and Morris N. A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Analytical Biochemistry* 1980 **105**:361-3.

Obar R A, Collins C A, Hammarback J A, Shpetner H S and Vallee R B**.** Molecular cloning of the microtubule-associated mechanochemical enzyme dynamin reveals homology with a new family of GTP-binding proteins. *Nature* 1990 **347:** 256–261.

Obis D. Genetic and biochemical characterization of a high-affinity betaine uptake system (BusA) in Lactococcus lactis reveals a new functional organization within bacterial ABC transporters. *Journal of Bacteriology* 1999 **181**: 6238–6246.

O'Farrell P H. High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry* 1975 **250**:4007-21.

O'Toole P W, Kostrzynska M and Trust T J. Nonmotile mutants of *Helicobacter pylori* and *Helicobacter mustelae* defective in flagellar hook production. *Molecular Microbiology* 1994 **14**:691–703.

Ofman J J, Etchason J, Fullerton S, Kahn K L and Soll A H. Management strategies for *Helicobacter pylori*-seropositive patients with dyspepsia: clinical and economic consequences. *Annals of Internal Medicine* 1997 **126**:280-91.

Oh-Ishi M and Maeda T. Separation techniques for high-molecular-mass proteins. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 2002 **771**:49-66.

Ohlmeier S, Scharf C and Hecker M. Alkaline proteins of *Bacillus subtilis*: first steps towards a two-dimensional alkaline master gel. *Electrophoresis* 2000 **21**:3701-9.

Olivieri E, Herbert B and Righetti P G. The effect of protease inhibitors on the two-dimensional electrophoresis pattern of red blood cell membranes. *Electrophoresis* 2001 **22**:560-5.

Ottemann K M and Lowenthal A C. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infection and Immunity* 2002 **70**:1984–90.

Pandey A and Lewitter F. Nucleotide sequence databases: a gold mine for biologists. *Trends in Biochemical Ssciences* 1999 **24**:276-80.

Pantoflickova D and Blum A L. Antagonist: should we eradicate *Helicobacter pylori* in non-ulcer dyspepsia? *Gut* 2001 **48**:758-9.

Pardanani A, Wieban E D, Spelsberg T C and Tefferi A. Primer on medical genomics part IV: expression proteomics. *Mayo Clinic Proceedings* 2002 **77**:1185 – 1196.

Park S M, Yoo B C, Lee H R, Yoon J H and Cha Y J. Antral *Helicobacter pylori* infection, hypergastrinemia and peptic ulcers: effect of eradicating the organism. *The Korean Journal of Internal Medicine* 1993 **8**:19-24.

Parsonnet J, Friedman G D, Vandersteen D P, Chang Y, Vogelman J H, Orentreich N and Sibley R K. *Helicobacter pylori* infection and the risk of gastric carcinoma. *The New England Journal of Medicine* 1991 **325**:1127-31.

Patton W F, Lim M J and Shepro D. Protein detection using reversible metal chelate stains. *Methods in Molecular Biology (Clifton, N.J.)* 1999 **112**:331-9.

Patton W F. Detection technologies in proteome analysis. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 2002 **771**:3-31.

Peisker K. Application of Neuhoff's optimized Coomassie brilliant blue G-250/ammonium sulphate/phosphoric acid protein staining to ultrathin polyacrylamide gels on polyester films. *Electrophoresis* 1988 **9**:236-8.

Peng J and Gygi S P. Proteomics: the move to mixtures. *Journal of Mass Spectrometry* 2001 **36**:1083-91.

Penfold S S, Lastovica A J, and Elisha B G. Demonstration of plasmids in *Campylobacter pylori*. *Journal of Infectious Diseases* 1988 **157:**850–851.

Pereira L and Hoover T R. Stable accumulation of sigma54 in *Helicobacter pylori* requires the novel protein HP0958. *Journal of Bacteriology* 2005 **187**:4463-9.

Perez-Perez G I, Taylor D N, Bodhidatta L, Wongsrichanalai J, Baze W B, Dunn B E, Echeverria P D and Blaser M J. Seroprevalence of *Helicobacter pylori* infections in Thailand. *The Journal of Infectious Diseases* 1990 **161**:1237-41.

Pichoff S, Vollrath B and Bouche J P. MinCD-independent inhibition of cell division by a protein that fuses MalE to the topological specificity factor MinE. *Journal of Bacteriology* 1997 **179**: 4616-4619.

Pichoff S, Vollrath B, Touriol C, and Bouche J P. Deletion analysis of the gene coding for the topological specificity of cell division in *Escherichia coli*. *Molecular Microbiology* 1995 **18:**321–329.

Pierrel F, Bjork G R, Fontecave M and Atta M. Enzymatic Modification of tRNAs. *Journal of Biological Chemistry* 2002 **277**:13367-70.

Pitson S M, Mendz G L, Srinivasan S and Hazell S L. The tricarboxylic acid cycle of *Helicobacter pylori*. *European Journal of Biochemistry* 1999 **260**:258-67.

Podkovyrov S and Larson T J. Lipid biosynthetic genes and a ribosomal protein gene are cotranscribed. *FEBS Letters* 1995 **368**:429-31.

Quadroni M and James P. Proteomics and automation. *Electrophoresis* 1999 **20**:664-77.

Rabilloud T. Solubilisation of proteins for electrophoretic analyses. *Electrophoresis* 1996 **17**:813-29.

Rabilloud T. Detecting proteins separated by 2-D gel electrophoresis. *Analytical Chemistry* 2000 **72**:48A-55A.

Rappu P, Shin B S, Zalkin H and Mantsala P. A Role for a Highly Conserved Protein of Unknown Function in Regulation of *Bacillus subtilis purA* by the Purine Repressor. *Journal of Bacteriology* 1999 **181**:3810-3815.

Reddy P G, Bhuyan D K and Bhuyan K C. Lens-specific regulation of the thioredoxin-1 gene, but not thioredoxin-2, upon *in vivo* photochemical oxidative

stress in the Emory mouse. *Biochemical and Biophysical Research Communications* 1999 **265**:345-9.

Replogle M L, Glaser S L, Hiatt R A and Parsonnet J. Biologic sex as a risk factor for *Helicobacter pylori* infection in healthy young adults. *American Journal of Epidemiology* 1995 **142**:856-63.

Righetti P G. Isoelectric focusing: theory, methodology and applications. Elsevier Amsterdam 1983 1-386.

Rock C O and Cronan J E Jr. *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochimica Biophysica Acta* 1996 **1302:**1–16.

Rothman J H, Raymond C K, Gilbert T, O'Hara P J and Stevens T H. A putative GTP binding protein homologous to interferon-inducible Mx proteins performs an essential function in yeast protein sorting. *Cell* 1990 **61:**1063–1074.

Ryan K A, Karim N, Worku M, Moore S A, Pen C W and O'Toole P W. HP0958 is an essential motility gene in *Helicobacter pylori*. *FEMS Microbiology Letters* 2005 **248**:47-55.

Scheffler N K, Falick A M, Hall S C, Ray W C, Post D M, Munson R S Jr and Gibson B W. Proteome of *Haemophilus ducreyi* by 2-D SDS-PAGE and mass spectrometry: strain variation, virulence, and carbohydrate expression. *Journal of Proteome Research* 2003 **2**:523-33.

Scheler C, Lamer S, Pan Z, Li XP, Salnikow J and Jungblut P. Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* 1998 **19**:918-27.

Schmiedeknecht G, Kerkhoff C, Orsó E, Stöhr J, Aslanidis C, Nagy G M, Knuechel R and Schmitz G. Isolation and characterization of a 14.5-kDa trichloroacetic-acid-soluble translational inhibitor protein from human monocytes that is upregulated upon cellular differentiation. *European Journal of Biochemistry.* 1996 **242**:339-51.

Schmitz G and Downs D M. Reduced transaminase B (IlvE) activity caused by the lack of *yjgF* is dependent on the status of threonine deaminase (IlvA) in *Salmonella enterica* serovar Typhimurium. *Journal of Bacteriology* 2004 **186**:803-10.

Schulenberg B, Beechem J M and Patton W F. Mapping glycosylation changes related to cancer using the Multiplexed Proteomics technology: a protein differential display approach. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 2003 **793**:127-39.

Shevchenko A, Wilm M, Vorm O and Mann M. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Analytical Chemistry* 1996 **68**: 850–858.

Shevchenko A, Loboda A, Shevchenko A, Ens W and Standing K G. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Analytical Chemistry* 2000 **72**:2132-41.

Singer M and Berg P. George Beadle: from genes to proteins. *Nature Reviews Genetics* 2004 **5**:949-54.

Sitas F, Smallwood R, Jewell D, Millard P R, Newell D G, Meuwissen S G, Moses S, Zwiers A and Forman D. Serum anti-*Helicobacter pylori* IgG antibodies and pepsinogens A and C as serological markers of chronic atrophic gastritis. *Cancer Epidemiology, Biomarkers & Prevention : a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 1993 **2**:119-23.

Sjostrom J E and Larsson H. Factors affecting growth and antibiotic susceptibility of *Helicobacter pylori*: effect of pH and urea on the survival of a wild-type strain and a urease-deficient mutant. *Journal of Medical Microbiology* 1996 **44**:425-33.

Stanley J, Linton D, Burnens A P, Dewhirst F E, On S L, Porter A, Owen R J and Costas M. *Helicobacter pullorum* sp. nov.-genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. *Microbiology* 1994 **140**:3441-9.

Steinberg T H, Haugland R P and Singer V L. Applications of SYPRO orange and SYPRO red protein gel stains. *Analytical Biochemistry* 1996a **239**:238-45.

Steinberg T H, Jones L J, Haugland R P and Singer V L. SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Analytical Biochemistry* 1996b **239**:223-37.

Su A I, Cooke M P, Ching K A, Hakak Y, Walker J R, Wiltshire T, Orth A P, Vega R G, Sapinoso L M, Moqrich A, Patapoutian A, Hampton G M, Schultz P G, Hogenesch J B. Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* 2002 **99**: 4465–70.

Suerbaum S, Smith J M, Bapumia K, Morelli G, Smith N H, Kunstmann E, Dyrek I, and Achtman M. Free recombination within *Helicobacter pylori. Proceedings of the National Academy of Science USA* 1998.**95:**12619–12624.

Suzuki J, Ohnsihi H, Shibata H, Wada A, Hirayama T, Iiri T, Ueda N, Kanamaru C, Tsuchida T, Mashima H, Yasuda H and Fujita T. Dynamin is involved in human epithelial cell vacuolation caused by the *Helicobacter pylori*–produced cytotoxin VacA. *Journal of Clinical Investigations* 2001 **107**:363-70.

Switzer R C 3rd, Merril C R, and Shifrin S. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Analytical Biochemistry* 1979 **98** :231-7.

Talley N J, Vakil N, Ballard E D 2nd and Fennerty M B. Absence of benefit of eradicating *Helicobacter pylori* in patients with non-ulcer dyspepsia. *The New England Journal of Medicine* 1999 **341**: 1106-11.

Tatsuya M, Tsai A Y M, and Saito H. Mutations in a Protein Tyrosine Phosphatase Gene (PTP2) and a Protein Serine/ Threonine Phosphatase Gene (PTCJ) Cause a Synthetic Growth Defect in *Saccharomyces cerevisiae. Molecular and Cellular Biology* 1993 **13**: 5408-5417.

Teh B H, Lin J T, Pan W H, Lin S H, Wang L Y, Lee T K and Chen C J. Seroprevalence and associated risk factors of *Helicobacter pylori* infection in Taiwan. *Anticancer Research* 1994 **14**:1389-92.

Thiede B, Hohenwarter W, Krah A, Mattow J, Schmid M, Schmidt F and Jungblut P R. Peptide mass fingerprinting. *Methods* 2005 **35**:237-47.

Thompson W G and Heaton K W. Functional bowel disorders in apparently healthy people. *Gastroenterology* 1980 **79**:283-8.

Tomb J F, White O, Kerlavage A R, Clayton R A, Sutton G G, Fleischmann R D, Ketchum K A, Klenk H P, Gill S, Dougherty B A, Nelson K, Quackenbush J, Zhou L, Kirkness E F, Peterson S, Loftus B, Richardson D, Dodson R, Khalak H G, Glodek A, McKenney K, Fitzegerald L M, Lee N, Adams M D, Hickey E K, Berg D E, Gocayne J D, Utterback T R, Peterson J D, Kelley J M, Cotton M D, Weidman J M, Fujii C, Bowman C, Watthey L, Wallin E, Hayes W S, Borodovsky M, Karp P D, Smith H O, Fraser C M anVenter J C. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997 **388**:539-47.

Tominaga K, Hamasaki N, Watanabe T, Uchida T, Fujiwara Y, Takaishi O, Higuchi K, Arakawa T, Ishii E, Kobayashi K, Yano I and Kuroki T. Effect of culture conditions on morphological changes of *Helicobacter pylori. Journal of Gastroenterology* 1999 **34** Suppl 11:28-31.

Truffa-Bachi P, Veron M and Cohen G N. Structure, Function, and Possible Origin of a Bifunctional Allosteric Enzyme, *Escherichia coli* Aspartokinase I-Homoserine Dehydrogenase I. *Critical Reviews in Biochemistry and Molecular Biology* 1974 **2**: 379-415.

Trust T J, Doig P, Emody L, Kienle Z, Wadstrom T and O'Toole P. High-affinity binding of the basement membrane proteins collagen type IV and laminin to the gastric pathogen *Helicobacter pylori*. *Infection and Immunity* 1991 **59**:4398-4404.

Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N and Schlemper R J. *Helicobacter pylori* infection and the development of gastric cancer. *The New England Journal of Medicine* 2001 **345**:784-9.

Vallee B L and Falchuk K H. The biochemical basis of zinc physiology. *Physiological Reviews* 1993 **73**:79-118.

Van Bogelen R A, Schiller E E, Thomas J D and Neidhardt F C. Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis* 1999 **20**: 2149–2159.

Van der Ende A, Rauws E A, Feller M, Mulder C J, Tytgat G N and Dankert J. Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. *Gastroenterology* 1996 **111**:638-47.

van Vliet A H, Stoof J, Vlasblom R, Wainwright S A, Hughes N J, Kelly D J, Bereswill S, Bijlsma J J, Hoogenboezem T, Vandenbroucke-Grauls C M, Kist M,

Kuipers E J and Kusters J G. The Role of the ferric uptake regulator (Fur) in regulation of *Helicobacter pylori* iron uptake. *Helicobacter* 2002 **7**:237-44.

Velayudhan J, Hughes N J, McColm A A, Bagshaw J, Clayton C L, Andrews S C and Kelly D J. Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. *Molecular Microbiology*. 2000 **37**:274-86.

Velazquez M and Feirtag J M. *Helicobacter pylori*: characteristics, pathogenicity, detection methods and mode of transmission implicating foods and water. *International Journal of Food Microbiology* 1999 **53**:95-104.

Verdon G, Albers S V, van Oosterwijk N, Dijkstra B W, Driessen A J M and Thunnissen A W H. Formation of the Productive ATP-Mg21-bound Dimer of GlcV, an ABC-ATPase from *Sulfolobus solfataricus*. *Journal of Molecular Biolology* 2003 **334**: 255–267

Vicente M, Chater K F and De Lorenzo V. Bacterial transcription factors involved in global regulation. *Molecular Microbiology* 1999 **33**:8-17.

Viola R E. The central enzymes of the aspartate family of amino acids biosynthesis. *Accounts of Chemical Research* 2000 **34**:339-349.

Waidner B, Greiner S, Odenbreit S, Kavermann H, Velayudhan J, Stahler F, Guhl J, Bisse E, van Vliet A H M and Andrews S C, Kusters J G, Kelly D J, Haas R, Kist M and Bereswill S. Essential Role of Ferritin Pfr in *Helicobacter pylori* Iron Metabolism and Gastric Colonization. *Infection and Immunity* 2002 **70**: 3923-3929.

Waite M. The PLD superfamily: insights into catalysis. *Biochimica Biophysica Acta* 1999 **1439**:187–197.

Walker G C. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiology Reviews* 1984 **48**:60-93.

Wang B, Jones D N, Kaine B P and Weiss M A. High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure* 1998 **6**:555-69.

Warren J R. Spiral bacteria of the gastric antrum. *Medical Journal of Australia* 1984 **141**:477-8.

Washburn M P, Wolters D and Yates J R 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* 2001 **19**:242-7.

Wasinger V C, Cordwell S J, Cerpa-Poljak A, Yan J X, Gooley A A, Wilkins M R, Duncan M W, Harris R, Williams K L and Humphery-Smith I. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 1995 **16**:1090-4.

Wei Y and Newman E B. Studies on the role of the *metK* gene product of *Escherichia coli* K-12. *Molecular Microbiology* 2002 **43**:1651-6.

Westblom T U, Phadnis S, Langenberg W, Yoneda K, Madan E and Midkiff B R. Catalase negative mutants of *Helicobacter pylori*. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology* 1992 **11**:522-6.

Westbrook J A, Yan J X, Wait R, Welson S Y and Dunn M J. Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms. *Electrophoresis* 2001 **22**:2865-71.

Westermeier R and Naven T. Proteomics in Practice: A Laboratory Manual of Proteome Analysis Wiley-VCH 2002 pages 1-160.

White I R, Pickford R, Wood J, Skehel J M, Gangadharan B and Cutler P. A statistical comparison of silver and SYPRO Ruby staining for proteomic analysis. *Electrophoresis* 2004, **25**:3048–54.

Wilkins M R, Williams K L Appel R D and Hochstrasser D F. Proteome Research: New Frontiers in Functional Genomics. Springer 1997.

Windle H J, Fox A, Ni Eidhin D and Kelleher D. The thioredoxin system of *Helicobacter pylori.* *Journal of Biological Chemistry* 2000 **275**:5081-9.

Wood Z A, Poole L B, Hantgan R R and Karplus P A. Dimers to doughnuts: redox-sensitive oligomerization of 2-cysteine peroxiredoxins. *Biochemistry* 2002 **41:** 5493–5504

Worst D J, Otto B R and de Graaff J. Iron-repressible outer membrane proteins of *Helicobacter pylori* involved in heme uptake. *Infection and Immunity* 1995 **63**:4161-5.

Worku M L, Sidebotham R L, Walker M M, Keshavarz T and Karim Q N. The relationship between *Helicobacter pylori* motility, morphology and phase of growth: implications for gastric colonization and pathology. *Microbiology* 1999 **145**:2803-11.

Wotherspoon A C, Doglioni C, Diss T C, Pan L, Moschini A, de Boni M and Isaacson P G. Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori*. *Lancet* 1993 **342**:575-7.

Wu F S and Wang M Y. Extraction of proteins for sodium dodecyl sulphate-polyacrylamide gel electrophoresis from protease-rich plant tissues. *Analytical Biochemistry* 1984 **139**:100-3.

Xie Z, Ho W T and Exton J H. Association of N- and C-terminal domains of phospholipase D is required for catalytic activity. *The Journal of Biological Chemistry* 1998 **273**:34679-82.

Yan J X, Kett W C, Herbert B R, Gooley A A, Packer N H and Williams K L. Identification and quantitation of cysteine in proteins separated by gel electrophoresis. *Journal of Chromatography A* 1998 **813**:187-200.

Yeatman T J. Future of cancer management: translating the genome, transcriptome and proteome. *Annals of Surgical Oncology* 2003 **10**: 7-14.

Zhang Y and Cronan J E Jr. Transcriptional analysis of essential genes of the Escherichia coli fatty acid biosynthesis gene cluster by functional replacement with the analogous *Salmonella typhimurium* gene cluster. *Journal of Bacteriology* 1998 **180**:3295-303.