



UNIVERSITY OF  
**LEICESTER**

**Phosphoglycerate kinase and phosphoenolpyruvate  
synthase of the enteric pathogen *Helicobacter pylori***

Thesis submitted for the degree of

**Doctor of Philosophy**

At the University of Leicester

By

**Aziz Yasir Hasan Al-Ethari**

Department of Molecular and Cell Biology

University of Leicester

June 2018

# Phosphoglycerate kinase and phosphoenolpyruvate synthase of the enteric pathogen *Helicobacter pylori*

Aziz Yasir Hasan Al-Ethari

## Abstract

*Helicobacter pylori* is a globally distributed enteric pathogen implicated in several serious diseases. Understanding the genetics and metabolism of the pathogen is of significant importance to developing new therapies for eradication. However, its metabolism is poorly characterised. The genome lacks coding sequences of some key glycolytic enzymes, however the gluconeogenic enzymes fructose-1,6-bisphosphatase and phosphoenolpyruvate synthase (hpPPSA) are present. This suggests *H. pylori* uses the glycolytic/gluconeogenic pathway for anabolic biosynthesis rather than for catabolic energy production. This study examines the structure and function of hpPPSA and phosphoglycerate kinase (hpPGK) and investigates the conditional essentiality of these genes, which were identified by *in silico* double deletion mutational studies of *H. pylori*.

The *ppsA* and *pgk* mutants (with controls) were constructed using experimental knock out strategies, and their role in synthetic lethality was investigated. The *ppsA*-mutated allele alone showed evidence of essentiality. The Krebs cycle in *H. pylori* deviates from the text book examples such as in humans and *E. coli*, thus *pgk* may be essential alone or in combination with other enzymes. The X-ray crystal structure of apo hpPGK was determined and compared to human PGK. Structural superposition showed that both the substrate and the nucleotide binding residues are well conserved. Four sulphate ions were identified bound in the hpPGK dimer. The positions of these molecules allowed the path of phosphoryl transfer during catalysis to be modelled. The enzyme was further characterised using kinetic techniques and compared to homologous enzymes. Prediction of the hpPPSA structure by homology modelling analysis located the essential His and Cys catalytic conserved residues in their respective domains. Superimposition of the N-terminal ATP binding domain superposition showed that these residues forming this binding site are well conserved.

Understanding *H. pylori* metabolism may provide directions for the development of therapeutics.

## Acknowledgements

It would not have been possible to write this thesis without the help and support of kind people, only some of whom it is possible to give particular mention here.

I am grateful to my supervisor Prof. Peter Moody whose expertise, understanding, guidance and support made it possible for me to work on a topic that was of great interest to me. It was a pleasure working under his supervision and learning from him as a scientist and human. I would like to express the deepest appreciation to my committee members Prof. Julian Ketley (for his suggestions during and after meetings and for giving me the opportunity to do the genetics part of my research in his lab and under his supervision) and Dr. Mark Leyland (for his suggestions during and after meetings). I would like to acknowledge all the members of Prof. Peter Moody's research group for assistance and advice and for providing me with a supportive and friendly environment to work in. In particular, I would like to thank Dr. Adnan Ayna, Dr. Samah Almehmadi and Dr Hanna Kwon. My thanks also go to Prof. Julian Ketley's lab, in particular Dr. Mohamed Elamen Fadel and Dr Richard Haigh for their unrestrained help during my genetic experimental work of this research. I would also like to thank Prof. John Schwabe's lab especially Dr. Louise Fairall (for her advices in protein crystallisation trials and helping in data collection from Diamond), Prof. Emma Raven's lab, in particular Dr Sofia M. Kapetanaki (for her support and providing all the experimental material when needed), Dr Mohammed El-Mezgueldi and Dr Jaswir Basran (for helping on enzyme kinetics), Dr Ralf Schmid (for useful discussions on bioinformatics work), Sura M. Abbood (for the scientific discussions and support) and lastly Dr. Xiowen Yang of blessed memory (for advises during and after cloning in expression vectors).

My acknowledgment would not be complete without giving special thanks to my beloved wife Hawraa Mahdi Hussein Al-Ethari and our beloved children Tabarek, Mohammed Sadeq, Baha Alddin and Amina. I could not make this work without your support. Thanks for always believing in me and for being there for me every single day of my life. For all that, thank you so very much.

Finally, my gratitude is to the Ministry of Higher Education and Scientific Research in Iraq for providing the funding for this PhD.

*To my beloved  
Father and mother,  
Wife and children,  
Brothers and sisters,  
I dedicate this thesis.*

## Table of Contents

Abstract .....	I
Acknowledgements .....	II
Table of Contents .....	IV
Abbreviations.....	XI
Chapter 1 Introduction .....	1
1.1 Introduction .....	1
1.2 Helicobacter pylori.....	1
1.2.1 Taxonomy.....	2
1.2.2 Morphology and growth requirements .....	4
1.2.3 Stomach colonisation and acid acclimation.....	5
1.3 H. pylori diseases and treatment.....	8
1.3.1 Epidemiology.....	8
1.3.2 Transmission .....	10
1.3.3 Disease spectrum .....	11
1.3.3.1 Gastritis.....	11
1.3.3.2 Peptic ulcer disease .....	12
1.3.3.3 Atrophic gastritis, intestinal metaplasia and gastric cancer .....	13
1.3.3.4 MALT lymphoma .....	13
1.3.4 Therapy regimes.....	14
1.3.5 Emergence of antibiotic resistance.....	16

1.3.5.1 Clarithromycin resistance.....	17
1.3.5.2 Metronidazole .....	17
1.4 Genome and strain diversity of <i>H. pylori</i> .....	19
1.5 Metabolism within <i>H. pylori</i> .....	20
1.5.1 Carbohydrate metabolism .....	21
1.5.2 Pyruvate metabolism .....	25
1.5.3 The Krebs cycle.....	27
1.6 Phosphoglycerate kinase and Phosphoenolpyruvate synthase .....	30
1.6.1 Phosphoglycerate kinase .....	30
1.6.2 PGK as drug target.....	32
1.6.3 Phosphoenolpyruvate synthase.....	32
1.6.4 Gene essentiality of <i>ppsA</i> .....	35
1.7 Aims of this study.....	37
Chapter 2. Materials and methods.....	39
2.1 Materials.....	39
2.1.1 Chemicals and reagents .....	39
2.1.2 Chromatographic media and membranes .....	39
2.1.3 Crystallisation accessories.....	39
2.2 Bacterial culture conditions and storage.....	40
2.3 Chromosomal DNA extraction .....	41
2.4 Plasmid DNA Extraction .....	42
2.5 Molecular Biology and cloning methods .....	42

2.5.1 Primer design .....	42
2.5.2 Gene amplification by polymerase chain reaction (PCR).....	43
2.5.2.1 PCR for cloning in pUC19.....	43
2.5.2.2 Colony PCR.....	45
2.5.2.3 PCR for cloning in overexpression vectors .....	45
2.5.2.3.1 pgk.....	45
2.5.2.3.2 ppsa.....	45
2.5.3 Agarose gel electrophoresis.....	46
2.5.4 Enzymatic modification of DNA .....	46
2.5.5 DNA dephosphorylation and ligation.....	46
2.5.6 Ethanol precipitation.....	47
2.5.7 Transformation by electroporation .....	47
2.5.7.1 Preparation of electro-competent cells .....	47
2.5.7.2 Electroporation.....	47
2.5.8 Verification of gene recombination in pUC19 .....	48
2.5.9 Plasmid minipreparation (miniprep) and PCR .....	48
2.5.10 Sequencing .....	48
2.5.11 Cloning and sequencing in overexpression vectors.....	49
2.5.12 Natural transformation of <i>H. pylori</i> strains.....	49
2.6 Protein expression and purification methods .....	50
2.6.1 Transformation.....	50
2.6.2 Overexpression optimization .....	51

2.6.3 Small scale purification of hpPGK .....	51
2.6.4 hpPGK expression and purification .....	52
2.6.5 SDS-PAGE of proteins .....	52
2.6.6 Peptide mass finger printing of hpPGK .....	53
2.6.7 Size exclusion chromatography (SEC) .....	53
2.7 Enzymatic characterization .....	54
2.7.1 Principle of hpPGK assay .....	54
2.7.2 Circular Dichroism (CD) spectroscopy .....	55
2.8 Crystallographic methods .....	55
2.8.1 Protein crystallisation .....	56
2.8.1.1 Crystallisation trials of hpPGK .....	57
2.8.1.2 Cryoprotectant selection for cryo-cooling of hpPGK crystals .....	58
2.8.2 Data collection and data processing .....	58
2.8.3 The phase problem .....	60
2.8.4 Refinement .....	61
2.8.5 Structure validation .....	62
2.9 Bioinformatics .....	62
2.9.1 Primary sequence alignment .....	62
2.9.2 Homology modelling of hpPPSA .....	63
Chapter 3. Investigation of ppsA and pgk synthetic lethality in <i>H. pylori</i> 26695 .....	64
3.1 Introduction .....	64
3.2 pgk, ppsA and rocF cloning into pUC19 .....	66

3.2.1 Cloning pgk into pUC19 .....	66
3.2.2 Cloning ppsA in pUC19 .....	68
3.2.3 Cloning rocF in pUC19 .....	70
3.3 Mutagenesis of pgk, ppsA and rocF in E. coli DH5 $\alpha$ .....	73
3.3.1 Insertion- inactivation of pgk .....	73
3.3.2 Deletion-inactivation of ppsA .....	75
3.3.3 Deletion-inactivation of rocF .....	78
3.3.3.1 cat insertion within pUC19-rocF inverse PCR product .....	79
3.3.3.2 Kanamycin insertion within pUC19-rocF inverse PCR product .....	82
3.4 Natural transformation of mutated pgk, ppsA and rocF in H. pylori 26695 .....	84
3.4.1 Transformation of pUC19-pgk::cat and pUC19 $\Delta$ rocF::cat plasmids .....	85
3.4.1.1 $\Delta$ rocF::cat verification in H. pylori .....	85
3.4.1.2 pgk::cat verification in H. pylori .....	87
3.4.2 Transformation of pUC19 $\Delta$ ppsA::kan and pUC19 $\Delta$ rocF::kan plasmids .....	91
3.5 Synthetic lethality of ppsA and pgk .....	94
3.6 Wild type and inactivated pgk alleles variation in pgk::cat H. pylori .....	95
3.7 Summary .....	98
Chapter 4. Structure and function of phosphoglycerate kinase and Bioinformatics of phosphoenolpyruvate synthase in H. pylori 26695 .....	100
4.1 Introduction .....	100
4.2 Cloning of pgk and ppsA in expression vectors .....	101
4.3 Expression and purification of hpPGK and hpPPSA .....	103

4.3.1 Expression trials of hpPGK and hpPPSA .....	103
4.3.2 Purification of hpPGK .....	106
4.3.3 Peptide mass fingerprinting of hpPGK and hpPPSA .....	108
4.3.4 Size exclusion chromatography (SEC) of hpPGK .....	109
4.4 Characterization of hpPGK.....	110
4.4.1 Enzymatic analysis of hpPGK.....	111
4.4.2 Far ultraviolet (UV) Circular Dichroism (CD) spectrum of hpPGK .....	113
4.5 X-Ray crystallography of apo- N-His <sub>6</sub> - hpPGK.....	113
4.5.1 Crystallization of hpPGK.....	113
4.5.2 Data collection and processing of apo-hpPGK.....	115
4.5.3 Structure determination of apo-hpPGK.....	117
4.5.4 Overall structure of apo hpPGK .....	120
4.5.5 ATP and 3,PGA binding sites of hpPGK using human PGK.....	123
4.5.6 Sulphate ligands bound to hpPGK.....	125
4.6 Homology modelling of hpPPSA .....	131
4.6.1 Primary sequence alignment of hpPPSA.....	131
4.6.2 Homology modelling of hpPPSA by Phyre2 .....	132
4.6.3 Ligand binding sites of hpPPSA ATP binding domain.....	136
4.7 Summary .....	140
Chapter 5. Discussion .....	141
5.1 Introduction .....	141
5.2 Synthetic lethality investigation in <i>H. pylori</i> .....	141

5.2.1 Mutagenesis of pgk.....	143
5.2.2 Mutagenesis of ppsA.....	146
5.3 Structure and function of hpPGK.....	148
5.3.1 Kinetics and solution characterisation of hpPGK.....	148
5.3.2 Crystal structure of hpPGK.....	149
5.3.3 Implications for drug design .....	151
5.4 Homology model of hpPPSA .....	152
5.5 Implications for H. pylori central metabolism .....	153
5.6 Conclusions .....	154
Appendix.....	156
References.....	159

## Abbreviations

1,3BPG	1,3bisphosphoglycerate
3-PGA	3-phosphoglycerate
6-PFK	6-phosphofructokinase
Å	Angstrom (1 Å = 0.1 nm)
a, b, c	Crystallographic cell vectors
ADP	Adenosine di phosphate
AMP	Adenosine mono phosphate
ATCC	American Types Culture Collection
ATP	Adenosine tri phosphate
bp	Base pair
cat	Chloramphenicol acetyl transferase cassette
CD	Circular Dichroism
Cj	<i>Campylobacter jejuni</i>
Da	Dalton
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
F	Forward primer
FBPase	Fructose-1,6-bisphosphatase
Fc	Calculated structure factors
Fo	Observed structure factors
GAP	Glyceraldehyde-3-phosphate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GST	Glutathione-S-transferase
Hp	<i>Helicobacter pylori</i>
I	Intensity of each reflection
IPTG	Isopropyl β-D-1-thiogalactopyranoside

$\sigma$	Uncertainties of intensity
kan	Kanamycin resistance cassette
kcat	Turn over number (overall rate constant)
Km	Michaelis constant
LB	Luria-Bertani
M	Molar
MHA	Mueller-Hinton Agar
MHB	Mueller-Hinton Broth
mM	Millimolar
MR	Molecular Replacement
NAD	$\beta$ -nicotinamide adenine dinucleotide
NADH	$\beta$ -nicotinamide adenine dinucleotide reduced form
NCS	Non-crystallographic symmetry
NEB	New England Biolabs
Ni-NTA	Nickel-charged nitriloacetic acid
NMR	Nuclear Magnetic Resonance
$^{\circ}$	Degree
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
ppsA	Phosphoenolpyruvate synthase
PGK	Phosphoglycerate kinase
R	Reverse primer
Rfree	R-factor calculated with 5% of observations used

Rmerge	Merging R-factor
Rmsd	Root mean square deviation
Rwork	Crystallographic R-factor
SAXS	Small angle X-ray scattering
SDS	Sodium dodecylsulphate
TAE	Tris Acetate EDTA
TEV	Tobacco Etch Virus
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
WHO	World Health Organisation
$\alpha, \beta, \gamma$	Crystallographic cell angles
$\mu\text{l}$	Microliter
$\mu\text{m}$	Micrometre

## Chapter 1 Introduction

### 1.1 Introduction

*Helicobacter pylori* (*H. pylori*) is an enteric pathogen that is found worldwide. In order to understand the pathogen and develop new therapies to combat it, it is necessary to gain the fullest possible knowledge of its metabolism biochemistry and genetics. The publication of the complete genome sequences of *H. pylori* strains started a new era in the understanding of the workings of this enteric pathogen. This study examines the metabolic enzymes phosphoglycerate kinase (PGK) and phosphoenolpyruvate synthase (PPSA) and the relationship between their annotated genes. Although these enzymes form part of the glycolytic/gluconeogenic pathways these are poorly characterised in this significantly important human pathogenic bacteria.

### 1.2 *Helicobacter pylori*

In 1889 the spiral organisms named *Vibrio rugula* found in the sediments of gastric washings from humans was first described in Poland by Walery Jaworsky (Jaworski, 1889). Subsequently, spirochetes were identified in the gastric mucosa of autopsied dogs by the Italian anatomist Giulio Bizzozero (Bizzozero, 1893). Early in the 20<sup>th</sup> century the German physician Walter Krienitz detected this spiral microbe in a human patient with gastric carcinoma (Krienitz, 1906).

For over a century this bacteria was thought to be a contaminant of digested food rather than being gastric colonisers. The organism emerged from relative obscurity when Barry Marshall and Robin Warren successfully isolated and cultured a spiral bacteria from the human stomach, and this later became known as *Helicobacter pylori* (Warren and Marshall, 1983). Self-ingestion experiments done by Marshall (Marshall et al., 1985), Morris (Morris and Nicholson, 1987) and long-term voluntary experiments (Morris et al., 1991) revealed that this bacteria colonise the epithelial layer of the human gut. This can lead to the symptoms of gastric mucosal inflammation and can develop into chronic and dangerous human diseases.

For these findings Warren and Marshall were awarded Nobel Prize in physiology and medicine in 2005. Since the discovery, *H. pylori* has been intensely investigated, producing a large body of literature. This pathogenic bacteria is now considered to one of the most important in terms of global impact on human health (Fox, 2002, Demiray-Gürbüz et al., 2017).

The following sections describe *H. pylori* classification, morphology, growth requirements and an overview of the reported mechanisms underlying *H. pylori*'s pathogenicity and host interactions, diseases, the current therapy regimens as well as an overview of the emerging antibiotic resistance.

### 1.2.1 Taxonomy

Cautiously, Marshall and Warren concluded that their newly discovered spiral bacilli appeared to be new species closely related to *Campylobacter*. Accordingly they preferred to name it as “pyloric *Campylobacter*” (Marshall and Warren, 1984). Nevertheless, the name *Campylobacter pyloridis* was proposed and the type strain designated as NCTC 13487 (Marshall et al., 1984). A few years later the species name was revised and grammatically corrected to *pylori* (the genitive of the noun pylorus) (Marshall and Goodwin, 1987).

*Helicobacter pylori* was listed in the taxonomy of gastric microorganisms in the new genus *Helicobacter*. This was on the basis of motility by means of sheathed flagella, external glycocalyx production, menaquinone 6 presence and the G+C content of chromosomal DNA (Goodwin et al., 1989). There are now more than 35 other species listed in the genus *Helicobacter*. Moreover, several new human and animal pathogenic species have been discovered and are awaiting formal nomenclature assignation (Frank et al., 2015, Shen et al., 2017). These species are subdivided into two major groups in terms of their organ specificity, i.e. gastric helicobacters and non-gastric or enterohepatic helicobacters (Kusters et al., 2006).

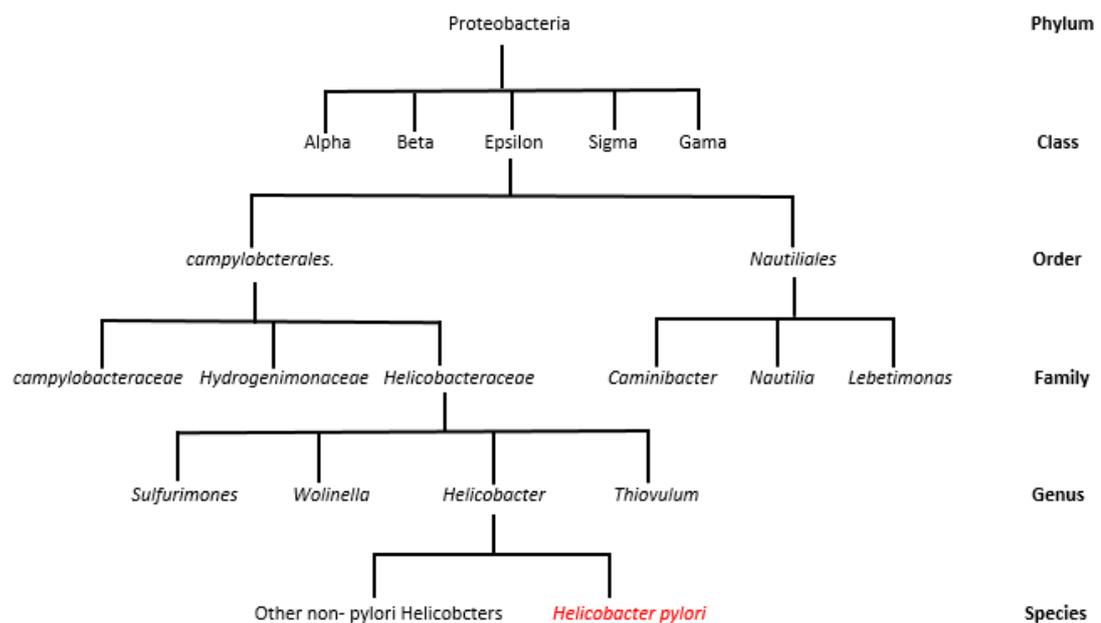
The genus *Helicobacter* is a member of *Helicobacteraceae* family in along with *Wolinella*, *Sulfurimones* and *Thiovulum*. The former two genera have been intensively studied as they are also hostile environment-living microorganisms that colonise the

host organism. In the case of *Wolinella* a commensal interaction occurs, whereby infected cattle are asymptomatic in spite of harbouring this bacteria (Baar et al., 2003).

This family is classified with *Campylobacteraceae* and *Hydrogenimonaceae* under the order *Campylobacterales*. *Campylobacter jejuni* is another highly investigated species in this order due to their host-interaction dependency and resulting disease association. Infection may cause humans acute food-borne gastritis, which can develop into Guillain-Barré syndrome (Nachamkin et al., 1998).

*Nautiliales* (genera *Nautilia*, *Caminibacter* and *Lebetimonas*) are one of the main orders in the Epsilon proteobacteria class along with the *Campylobacterales*. Generally, the members belonging to this order are marine and terrestrial environmental bacteria, which have a key roles in the nitrogen, carbon and sulphur cycles (Campbell et al., 2006).

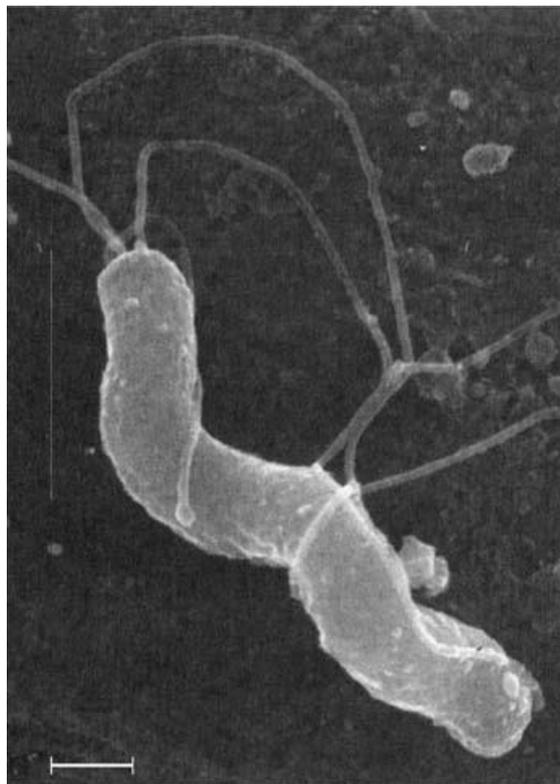
Alpha, beta, gamma, sigma and epsilon classes are identified using 16S rRNA analysis (Stackebrandt et al., 1988) these are the members of the major phylum of the bacteria kingdom known as proteobacteria “the purple bacteria and their relatives” (Woese, 1987, Garrity et al., 2003). Figure 1.1 shows a schematic representation of *H.pylori* classification and the relationships between classes, orders and families of the phylum Proteobacteria.



**Figure 1.1** Schematic representation of *H. pylori* classification and the relationships between proteobacteria classes, orders, families and genera. For clarification the species *H. pylori* is shown in red.

## 1.2.2 Morphology and growth requirements

*H. pylori* (figure 1.2) is a Gram-negative bacteria, seen as curved or S-shaped spiral rods of 0.5-1  $\mu\text{m}$  width and 2-4  $\mu\text{m}$  length showing 1-3 turns under *in vivo* microscopic observations. The bacteria are non-spore forming when grown on blood agar. Other shapes have been reported during *in vitro* culturing, and occasionally *in vivo* growth, including coccoid and straightened forms (Owen, 1998). The coccoid form might be due to prolonged growth or antibiotic treatment effect (Kusters et al., 1997). *H. pylori* is a motile bacteria, using up to six unipolar sheathed flagella (Paul et al., 2000). Each flagellum consists of basal body, hook and filament. The latter extends 3-5  $\mu\text{m}$  from the cell with a bulb or club-like thickening that is often seen at the tip of the flagellar filament (Geis et al., 1989).



**Figure 1.2** Electron microscopy image of *Helicobacter pylori*, Bar is 0.5  $\mu\text{m}$ , taken from (Atherton, 2006)

The morphology of *H. pylori* colonies on solid media are circular, convex and translucent with 1-2 mm diameter. The colonies take 3-5 days to appear, with a slight greyish colour from haemolysis, when grown at 37°C on blood-agar supplemented media such as brain heart infusion. *H. pylori* is microaerophilic; it grows optimally in an atmosphere of 5%

oxygen and 5-10% CO<sub>2</sub> (Owen, 1998). However, other studies conclude that *H. pylori* is a capnophilic and grow equally as well as aerobic and microaerobic bacteria at high bacterial densities *in vitro* growth conditions. At low cell concentrations the bacteria behave as oxygen-sensitive microaerophiles. These *in vitro* characteristics suggest that *H. pylori* physiologically adapts to change in *in vivo* environment.(Bury-Moné et al., 2006).

High amounts of L-proline, D- with L-alanine isomers and L-serine were identified in *H. pylori* cells, and these amino acids were predominant in samples of human gastric juice. According to that, *H. pylori* were hypothesised to utilize L-proline, D- and L-alanine and L-serine as important energy sources in its habitat of the mucous layer of the stomach (Nagata et al., 2003).

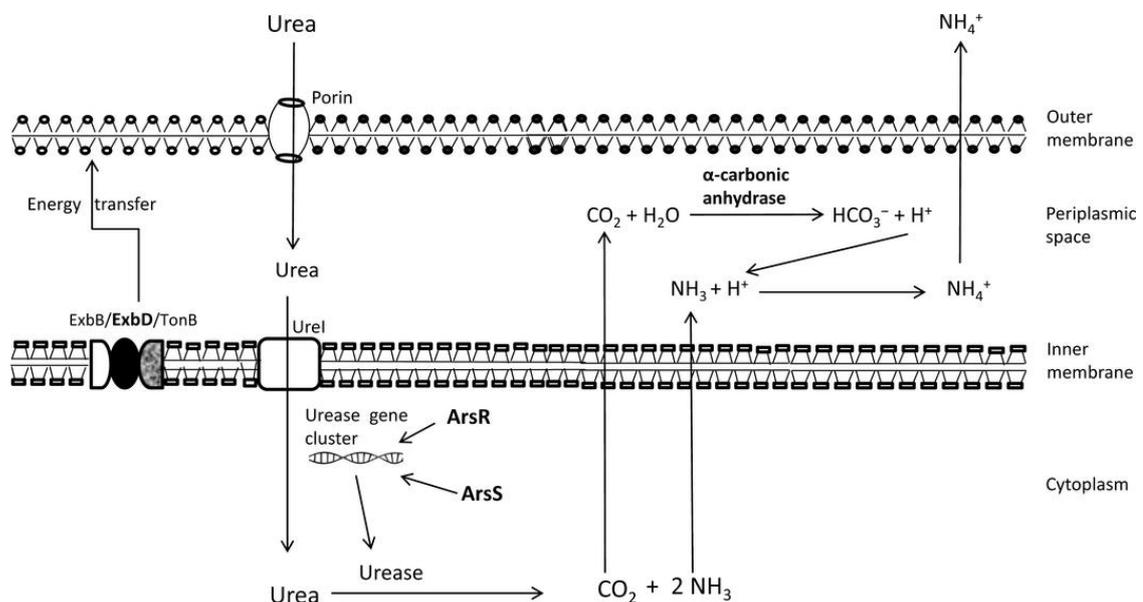
### **1.2.3 Stomach colonisation and acid acclimation**

This spiral shaped and motile bacteria inhabit and colonise the human gastric mucosa. The bacteria can persist in the stomach's acidic environment by means of highly organised acid acclimation mechanisms (described in figure 1.3). These mechanisms include the multi subunit enzyme, urease and carbonic anhydrase, together these enable *H. pylori* to neutralize gastric acidity and reduce periplasmic pH to approximately 6.1. The key processes is the hydrolysis of urea (imported to the cytoplasm) to ammonia and carbon dioxide (Ansari and Yamaoka, 2017).

As a gastric pathogen the survival of free *H. pylori* in the stomach environment is important for the early stages of infection prior to colonisation to the gastric mucosa. When the bacteria are buried in the gastric mucosa, it is protected against the strong acid of the stomach. The bacteria penetrate this layer toward the epithelial cells where the nutrients are available and are protected from the host immune system (Martínez et al., 2016).

The urea-lytic activity by urease enables *H. pylori* to neutralize the acidic pH of the surrounding microenvironment in the gastric mucosa. This leads to modification of its consistency from visco-elastic to less gel-like which provide favourable environment and allows motility through the mucosa (Celli et al., 2009). The helical shape of *H. pylori*

is also essential for adapting to the different environments. This characteristic facilitate the rapid movement of the bacteria inside the less acidic mucus layer to escape the highly acidic regions (Sycuro et al., 2010). In addition to the helical shape, the active motility system, unipolar flagella, helps orally ingested *H. pylori* to penetrate and migrate efficiently toward the higher pH regions of the gastric mucosa. Furthermore gastric epithelial cells are continuously secreting signalling chemicals into the lumen, e.g. urea, sodium carbonate and potassium carbonate, which would act as markers to attract *H. pylori* toward their higher pH (Nakamura et al., 1998a).



**Figure 1.3** Acid acclimation mechanisms in *H. pylori*: The outer membrane contains porins permeable to urea. Energy is required for the active transport of urea. This energy is provided by ExbD a part of a three proteins complex, ExbB/ExbD/TonB, located in the inner membrane (IM). ArsR and ArsS proteins regulate the expression of urease cluster of genes and increase its activity during high acid exposure. High levels of gastric acidity,  $\text{pH} < 6$ , opens the pH-gated urea channel, Urel, and trigger urea transportation to the cytoplasm. Urea stimulates intracellular urease activity to produce ammonia ( $\text{NH}_3$ ) and  $\text{H}_2\text{CO}_3$ . The latter is in equilibrium with  $\text{H}_2\text{O}$  and  $\text{CO}_2$ . These gases release readily into the periplasmic space through the inner membrane.  $\text{CO}_2$  act as a substrate for inner membrane anchored, periplasmic  $\alpha$ -carbonic anhydrase ( $\alpha$ -CA) in the presence of water to produce  $\text{HCO}_3^-$  and  $\text{H}^+$ .  $\text{HCO}_3^- / \text{CO}_2$  couple provides buffering system maintaining periplasmic pH at  $\sim 6.1$ , in addition,  $\text{NH}_3$  neutralises the protons produced by  $\alpha$ -CA and other entering protons by forming  $\text{NH}_4^+$ . Similarly, the  $\text{NH}_3$  that is released into the periplasm can exit the outer membrane to alkalis the surrounding medium. This system maintains periplasmic pH values much higher than the medium (Ansari and Yamaoka, 2017).

There is little published data on the effect of pH on the chemotactic behaviour of *H. pylori*. Nonetheless, previous research established that the optimum chemotactic activity occurred in a pH range of 5.5-6.5, but no apparent *in vitro* chemotaxis was reported with in acidic pH of 3. This suggests an effect of pH on the chemotactic response of *H. pylori* (Tadjrobehkar and Abdollahi, 2015). In *H. pylori* four chemoreceptors, TlpA, TlpB, TlpC and TlpD, are essential for sensing the chemical attractants found in the environment. These modulate the unipolar flagella induce movement in response to these chemo attractants. The presence of chemoreceptors facilitates movement in the presence of chemo attractants (Rader et al., 2011).

In summary, the breakdown of urea to  $\text{NH}_3$  and  $\text{CO}_2$  provides the means for acid-resistance, by maintaining the buffering activity of the periplasm against gastric acidity. Furthermore, chemical sensing in addition to the helical shape and the tuft of unipolar flagella are of importance to *H. pylori* as they provide screw-like active motility to swim toward the higher pH environment in the stomach. Together, these mechanisms have enabled *H. pylori* to colonise the gut of more than half of the world population (Ansari and Yamaoka, 2017). Often, this presence might be asymptomatic or might be associated with chronic inflammations and more dangerous diseases.

### 1.3 *H. pylori* diseases and treatment

*H. pylori* was the subject of research to explain its coexistence with humans and its ancestral roots (Leja et al., 2016). Simulations indicated that *H. pylori* spread from East Africa 58,000 years ago (Linz et al., 2007). More recent work suggests this ancient microorganism may have cohabited with humans for 100,000 years (Moodley et al., 2012). This study also noted genetic variability among the *H. pylori* variants that are distributed in different geographical areas (Falush et al., 2003, Breurec et al., 2011). However, whole genome analysis studies of 60 *H. pylori* strains distributed worldwide indicated that the association of this microorganism with humans may be even more ancient. The study suggested that this association began in the early stages in the development of human species (Montano et al., 2015).

The following sections will deal with epidemiology, transmission, disease spectrum, therapy regimens and emergence of antibiotic resistance in *H. pylori*.

#### 1.3.1 Epidemiology

In recent years, there has been an increasing amount of literature on epidemiology of *H. pylori*, with analysis of correlation of infection rates with many variables. Some reviews reported that *H. pylori* infections are more common in developing countries compared to developed countries (Linz et al., 2007), others conclude that individuals acquire infection during childhood (Malaty et al., 2002). Other studies show the prevalence of *H. pylori* infection is related to socioeconomic factors including poverty, level of education, hygiene conditions and household overcrowding (Eusebi et al., 2014). A systematic meta review of the literature was undertaken to describe the prevalence of *H. pylori* infections in different countries and periods (Peleteiro et al., 2014). Only reviews evaluating samples with national coverage were included. Thirty-seven studies were identified in 22 countries around the world. Five American, six Asian, ten European and one Australian. The studies published between 1995 and 2013 cover *H. pylori* prevalence data from 1968 to 2011. The majority of these reports were conducted to estimate prevalence in samples with subject median ages ranging from around 20 to 60 years. The prevalence estimates from samples analysed mostly in the

late 1990s to the early 2000s were higher among central and South American countries, with up to 70%-90% in Mexico for 20 to 60 year olds respectively. Another high prevalence area was in Asia, in 1998 the prevalence level in the Republic of Korea was 50% at around 20 years and 70% at around 60 years. However, a lower level of prevalence was reported in the studies conducted in the United States of America in 1999-2000 with 20% and 40% prevalence among these younger adults and older ages respectively. In Australia and most of European countries, lower prevalence estimates were also observed for the whole sample. More recent surveys showed lower prevalence estimates, in the Czech Republic, between 2001 and 2011, the prevalence declined from 30 to 10% in ages around 20 years and from 60 to 40% in ages around 60 years. The prevalence in the Republic of Korea also decreased between 1998 and 2011, from 50 to 20% in young adults and from 70 to 60% in older subjects.

There are no African studies covered in Peleteiro et al's review. Nonetheless, prevalence estimates of *H. pylori* in 2011 and 2013 were reported in studies conducted in Rwanda, Ethiopia, Morocco and Nigeria. The prevalence in these African countries ranged from around 70% to 80% (Eusebi et al., 2014, Mentis et al., 2015).

Another systematic review focused on *H. pylori* infection studies in the Middle East (Eshraghian, 2014). Prevalence estimations were analysed using 26 articles covering data from 1989-2013 in Iran and eight nearby countries. Two studies published in 2005 and 2009 investigated *H. pylori* infection in healthy populations in Iran. The prevalence estimation was 69 and 67.5% respectively. Among the whole countries sampled in the review, Egypt showed the highest prevalence in healthy adults. Generally, *H. pylori* prevalence estimations in most of these countries was higher than 70%.

Epidemiology of *H. pylori* was investigated for the period of April 2013 to March 2014 using literature servers such as PubMed and Medline (Eusebi et al., 2014). Several studies still show high estimated prevalence in most countries. Around one third of adults are infected in North American and North European populations. *H. pylori* prevalence is predominantly higher than 50% in South and East Europe, South America and Asia. Furthermore, in developed countries, *H. pylori* prevalence remains high among immigrants coming from developing countries. However, promises of further

decline of *H. pylori* prevalence in the future is expected as lower rates were seen in the younger populations.

### 1.3.2 Transmission

The mechanisms whereby *H. pylori* infection is transmitted remain poorly understood. This might be due to difficulties of isolating *H. pylori* from large population samples for comprehensive epidemiological transmission studies and the genetic variability between isolates (Schwarz et al., 2008). It was hypothesised that transmission takes place in early childhood and mainly through intrafamilial contact, most likely from mother to child (Konno et al., 2005). However, in populations such as Norway with apparently high hygienic standards, using both urban and rural community recruited samples, the evidence suggests transmission of this pathogenic bacteria may also start in adolescence, with the potential means of transmission associated with the use of outdoor toilets, private well water and having farm animals (Breckan et al., 2016).

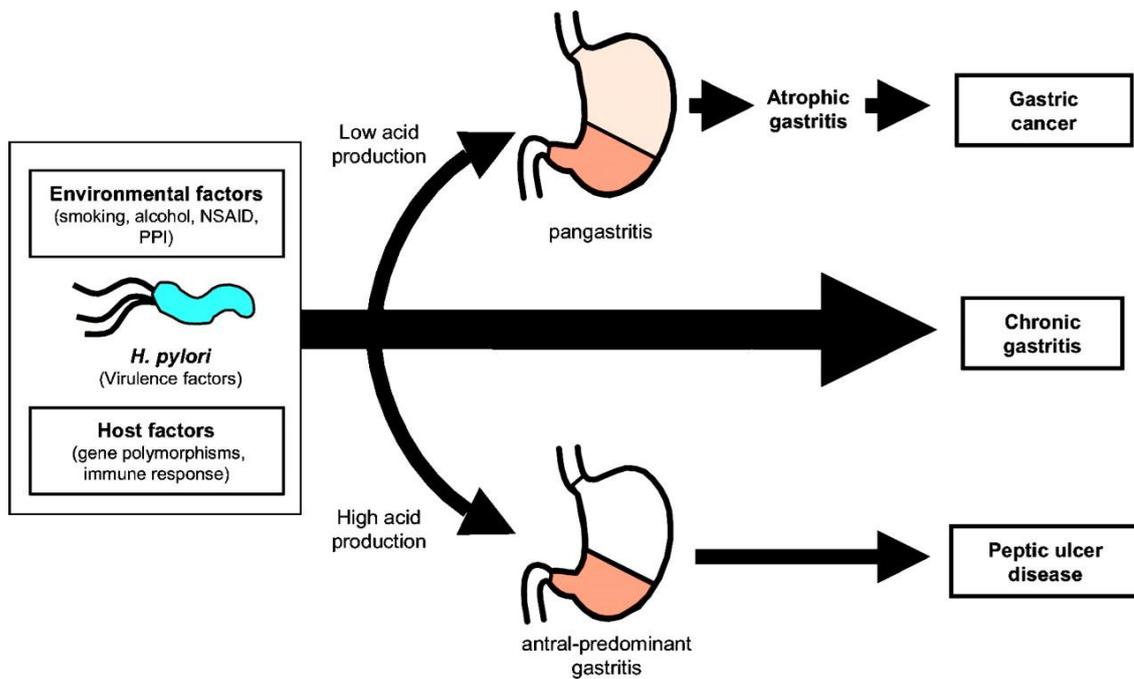
*H. pylori* is considered as narrow host pathogen as it is found almost exclusively in humans and some non-human primates (Kusters et al., 2006). No conclusive evidence for zoonotic transmission is available yet although animals have been proposed as a source of the infection (Momtaz et al., 2014). New infections are thought to occur via two main transmission routes, that is, direct contact between humans either by oral-oral, gastro-oral or faecal-oral transmission and through indirect routes via contaminated food, drinking water or animals (Khalifa et al., 2010). These routes are consistent with observations that *H. pylori* were detected in saliva, vomitus, gastric refluxate and faeces, but there is no ironclad evidence for predominant transmission via any of these products (Kusters et al., 2006). However, a recent study indicated that saliva can be both a transmitting and a re-infecting vector as similar genotypes of *H. pylori* were identified in both saliva and stomach of investigated patients (Román-Román et al., 2013). The only proven way of *H. pylori* transmission is the nosocomial infection during endoscopy. Inadequate disinfection of endoscopy has been shown to cause this iatrogenic transmission (Tytgat, 1995).

### 1.3.3 Disease spectrum

*H. pylori* colonisation is not in itself a disease, but a condition that increases the risk of developing a group of clinical disorders mainly in the upper gastrointestinal tract and possibly in the hepatobiliary tract (Kusters et al., 2006). These disorders are associated with chronic gastritis, and are implicated in more dangerous diseases such as peptic ulceration, chronic atrophic gastritis (the initial step of gastric carcinomas), and cancers of gastric mucosal lymphoid tissues, MALT lymphoma, (Nakamura et al., 1998b). Although all individuals with *H. pylori* gastric colonisation have potential to develop histologic gastritis infections, only a minority develop any apparent clinical signs of this colonisation. Lifetime risk estimations of developing gastric ulcer disease in *H. pylori* infected individuals range from 10 to 20%, whilst their risk of developing distal gastric cancer is 1 to 2%. A variety of bacterial, host and environmental factors, mostly related to the pattern and severity of gastritis, are variable factors that modulate the risk of developing these disorders in the presence of *H. pylori* infections (Kuipers, 1999, Ernst and Gold, 2000). Figure 1.4, taken from (Kusters et al., 2006), summarises the development of *H. pylori* diseases.

#### 1.3.3.1 Gastritis

When infected individuals develop gastritis, the colonisation leads to infiltration of gastric mucosal tissue in both antrum and corpus regions with mononuclear and neutrophilic cells. This chronic active disease is the first action related to *H. pylori* colonisation, and this inflammatory process specifically induces the other *H. pylori*-associated diseases namely peptic ulceration, atrophic gastritis and gastric cancer (Kusters et al., 2006). *H. pylori* in particular colonises the gastric antrum in patients with entire acid secretion, where the bacteria escape towards few acid-secreting parietal cells region. This mode of colonisation is related to antrum-predominant gastritis. Other patients with impaired acid secretion, e.g. due to use of proton pump inhibitors (PPIs) have higher distribution of bacteria in both antrum and corpus. The pathogen in the corpus is in closer contact with the mucosa which lead to corpus-predominant pangastritis (Uyterlinde et al., 1995) (Figure 1.4).



**Figure 1.4** Schematic representation of disease types and the factors developing gastric disorders in *H. pylori* infections adapted from (Kusters et al., 2006).

### 1.3.3.2 Peptic ulcer disease

Peptic ulcer disease refers to both gastric and duodenal ulcers and are defined as holes in the gastric mucosa with a minimum diameter of 0.5 cm infiltrating through the muscularis mucosa. Gastric ulcers predominantly occur along the most flattened region of the stomach, in particular this is at the transition from corpus to antrum mucosa. Duodenal ulcers occur in the area mostly exposed to gastric acid, that is the duodenal bulb (van Zanten et al., 1999). Studies conducted in the first decade after Warren and Marshall's *H. pylori* discovery highlighted a close association between gastric and duodenal ulcers and *H. pylori* infection. Up to 10-20% of *H. pylori* infected patients developed peptic ulcer disease, and the life time risk of infected subjects were estimationed as 3-4 times higher than non-infected subjects. Moreover, *H. pylori* infection was diagnosed in 60-100% of gastric ulcer patients and 90-100% of duodenal ulcer patients (Kuipers et al., 1995b). In children duodenal ulcers were mainly observed rather than gastric ulcers. Conversely, gastric ulcers were more common in adults over 40 years old (Brown, 2000). Other studies reported that gastric and duodenal regions

with high inflammation rates are significantly more subject to ulceration (van Zanten et al., 1999).

### **1.3.3.3 Atrophic gastritis, intestinal metaplasia and gastric cancer**

The development of chronic *H. pylori*-induced inflammations can ultimately lead to architectural malformation of the normal gastric mucosa, gastric gland destruction and histological replacement by fibrosis and intestinal-type epithelium. These processes occur in around half of the *H. pylori* colonised population, preliminarily in those patients and at those sites with mostly severe inflammation (Kuipers et al., 1995a). The distribution and pattern of active chronic inflammation influences the risk for atrophic gastritis. As such, the progression toward atrophy is significantly rapid in patients with decreased gastric acid secretion due to proton pump inhibitors (Kuipers et al., 1996). Gland loss and intestinal metaplasia regions expand multifocally with time. Depending on atrophy extent and severity, these regions increase the risk for gastric cancer 5 to 90 times although they do not develop to apparent symptoms (Slipponen et al., 1985). Gastric atrophy represents a starting point of gastric cancers, that is non-cardia gastric carcinoma (NCGC) and low grade B-cell MALT lymphoma, due to this *H. pylori* is designated as a global class I carcinogen in 1994 and 2012 by the International Agency for Research on Cancer IARC in the world health organization WHO (Plummer et al., 2015).

### **1.3.3.4 MALT lymphoma**

Mucosa-associated lymphoid tissue lymphoma (MALT lymphoma) is a multifocal B-cell type cancer originating in the margins of MALT, frequently of the stomach, but supposedly any other mucosal tissues can be afflicted as well (Taal et al., 1996). Normally, the gastric mucosa of healthy individuals does not contain any lymphoid tissue. However, almost always *H. pylori* colonisation can lead to the appearance of mucosa-associated lymphoid tissue (MALT). This non-Hodgkin's type of lymphoma can rarely develop due to the rise of a monoclonal population of B cells from this tissue followed by slow proliferation (Kusters et al., 2006).

### 1.3.4 Therapy regimes

Antibiotic sensitivity analysis showed that *H. pylori* is sensitive to many antibiotics *in vitro*. However, combination therapy is essential to effectively eradicate *H. pylori* infection as no single antibiotic (i.e. monotherapy), is effective alone *in vivo* due to the extreme gastric acidity which affect the antibiotic bioactivity in the stomach (Bazzoli et al., 2002).

The treatment package should include at least two antibiotics in combination with a gastric acidity controller such as proton pump inhibitors (PPI). The second Maastricht consensus report (Malfertheiner et al., 2002) introduced the concept of combination therapy that considers first-line and second line eradication therapies together. The first-line is suggested to be triple therapy including a proton pump inhibitor or ranitidine bismuth citrate in combination with clarithromycin and amoxicillin or metronidazole. Second-line therapy recommended is quadruple therapy with proton pump inhibitor, bismuth, metronidazole and tetracycline. These therapy regimes effectively cure *H. pylori* infection and prevent revival of peptic ulcer diseases, such as gastric or duodenal ulcers (Graham et al., 1992). Anti-*H. pylori* regimes are advocated as the sole initial treatment for *H. pylori*-positive gastric MALT lymphoma (Zucca et al., 2009, Asaka, 2013). In addition to this, eradication of *H. pylori* is recommended in a 'test and treat' prophylactic strategy in communities with a high incidence of gastric adenocarcinoma (Fock et al., 2009). The *H. pylori* gastritis Kyoto Consensus Report (Sugano et al., 2015) recommended that only regimes that produce at least reliable 90% eradication rate should be documented as empirical treatment.

In recent years there has been eradication failure and declining effectiveness of triple therapy regimes due to the emergence of antibiotic resistance mainly against clarithromycin, the rapid metabolism of proton pump inhibitors and poor compliance (Graham et al., 2014). Currently, updated eradication strategies including bismuth quadruple, non-bismuth quadruple (involving sequential, concomitant, hybrid and reverse therapies) and high-dose dual therapy regimes are followed to enhance eradication rate. Table 1.1 summarises the latest updated first-line treatment strategies for *H. pylori* infections (Huang et al., 2017).

Sequential therapy starts with 5 days of dual therapy with a PPI and amoxicillin followed by 5 days of triple therapy with a PPI, clarithromycin and metronidazole (Table 1.1) (Zullo et al., 2007). Another regime is concomitant therapy, which is a quadruple therapy regime containing PPI, clarithromycin, amoxicillin and metronidazole, they are all given together for 7-14 days (Table 1.1) (Essa et al., 2009).

**Table 1.1** Current classical and updated *H. pylori* therapy regimes. A; amoxicillin, C: clarithromycin, M: metronidazole, T: tetracycline, BIS: bismuth, PPI: proton pump inhibitor, bid: two times a day, qid: four times a day and SD: standard dose.

Treatment		Regime						Duration day	
		A g	C g	M g	T g	BIS g	PPI SD		
First	Standard triple therapy	1 bid	0.5 bid				1 bid	7-14	
	Concomitant therapy	1 bid	0.5 bid	0.25 qid			1 bid	7-14	
Line therapy	Sequential therapy	1 <sup>st</sup> phase	1 bid				1 bid	5	
		2 <sup>nd</sup> phase		0.5 bid	0.25 qid		1 bid	5	
	Hybrid therapy	1 <sup>st</sup> phase	1 bid				1 bid	7	
		2 <sup>nd</sup> phase	1 bid	0.5 bid	0.25 qid		1 bid	7	
	Reverse hybrid therapy	1 <sup>st</sup> phase	1 bid	0.5 bid	0.25 qid		1 bid	7	
		2 <sup>nd</sup> phase	1 bid				1 bid	7	
	Second line therapy	Bismuth-containing quadruple therapy			0.25 qid	0.5 qid	0.12 qid	1 bid	7-14
		high-dose dual therapy			0.25 qid			1 qid	14

Hybrid therapy regime components are PPI and amoxicillin as dual therapy given for the first 7 days of the course, following that quadruple therapy with a PPI, amoxicillin, clarithromycin and metronidazole are given for another 7 days (table 1.1 ) (Hsu et al., 2015). One of the drawbacks of this regime is the complexity, as drugs are switched half way through the course. This can be simplified by using reverse therapy namely reversing the drug administration sequence using quadruple therapy at the beginning followed by dual therapy. This simplifies the treatment and makes it a one-step in two-phase therapy (Table 1.1 ) (Hsu and Wu, 2016). The high-dose dual therapy is a newly developed combination of high-dose PPI and amoxicillin. This regime is superior to the classical first-line triple therapy or the alternative therapies for *H. pylori* infections (table 1.1 ) (Yang et al., 2015).

### **1.3.5 Emergence of antibiotic resistance**

Antibiotics are therapeutics used to prevent and treat bacterial diseases. However, resistance can occur when bacteria change in response to the misuse of these medicines in humans and animals. Accordingly, Antibiotic resistance can lead to longer hospital stays, higher medical costs and increased mortality.

The success of *H. pylori* eradication strategies is compromised due to the emergence of antibiotic resistance. Prevalence of resistance varies between countries and within the same country by periods of time. A potential reason behind this failure is the over use of various antibiotics in the general population, adults and children, for respiratory, dental, parasitic and gynaecological infectious diseases (Cosme et al., 2017). Prevalence of *H. pylori* antibiotic resistance are increasing significantly around the world and concordantly eradication rate are globally decreasing (figure 1.5) (Ghotaslou et al., 2015, Thung et al., 2016).

The following sections describe the prevalent antibiotic resistance against *H. pylori* treatments, particularly clarithromycin and metronidazole, around the world.

### 1.3.5.1 Clarithromycin resistance

*H. pylori* resistance against clarithromycin has been extensively studied as it is the most potent antibiotic involved in the management of the bacterial infection. In Asia, for example clarithromycin resistance increased dramatically in the Japanese population from 1.8% in 1996 to 27.1% in 2008 (Horiki et al., 2009). Similarly, in China this antibacterial resistance increased from 14.8% in 2000 to 52.6% in 2014 (Gao et al., 2010, Zhang et al., 2015). Overall, clarithromycin resistance in Asia increased from 15.2% in 2009 to 32.46% in 2014 (Ghotaslou et al., 2015).

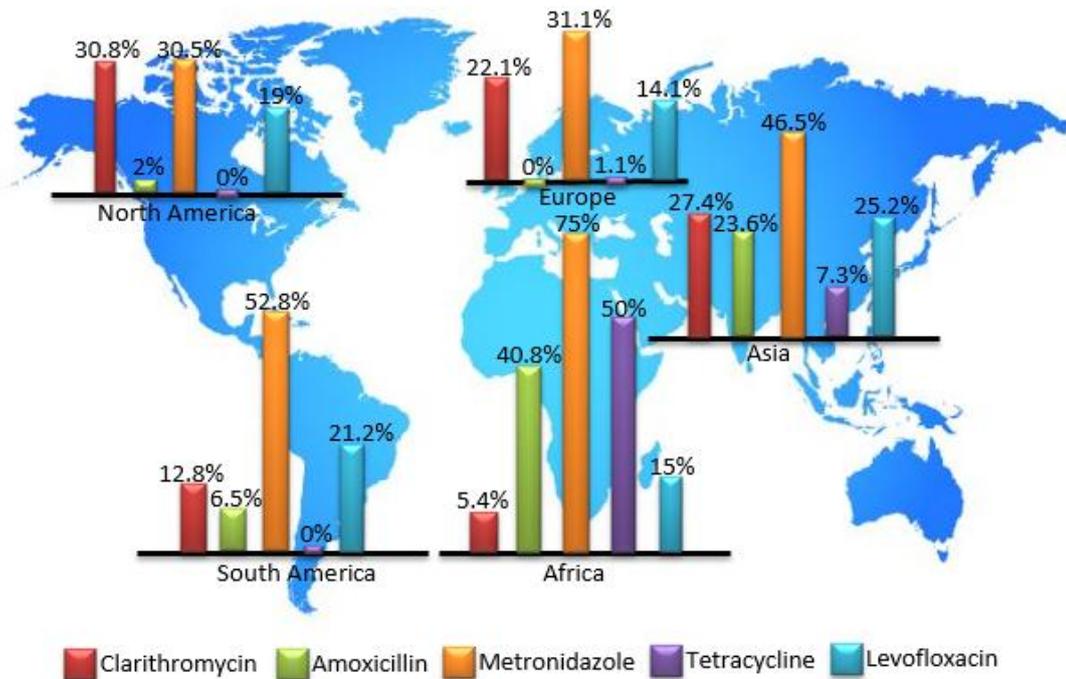
In the United States of America the prevalence of clarithromycin resistance (by mutation of the ribosome 23s subunit target of the drug (Versalovic et al., 1996) were studied among male veterans undergoing esophago-,gastro-,duodenal-, endoscopy using collected of gastric biopsies. A significant increase reported from 9.1% in 2009 to 24.2% in 2013 (Shiota et al., 2015). Moreover, a significant increase of clarithromycin resistance as high as 50% were concluded to be present in the paediatric population (Mitui et al., 2014). Overall, clarithromycin resistance rate in North America (United States, Canada and Mexico) increased to 30.8% in 2014, which is the highest rate around the world (figure 1.5) (Ghotaslou et al., 2015).

In European countries, the highest clarithromycin anti-bacterial resistance rates were reported in Spain and Portugal with 32.01% and 42.35% respectively, whilst the lowest in Norway (5.9%). Overall, European studies conducted in the six years interval reported that antibiotic resistance decreased from 36.6% to 24.3% in 2009 and 2014 respectively (figure 1.5) (Ghotaslou et al., 2015).

### 1.3.5.2 Metronidazole

The most common global antibiotic resistance found in *H. pylori* is metronidazole resistance, due to mutations in the *rdx* gene that encodes an oxygen-insensitive NADPH reductase implicated in the activation of the prodrug (Goodwin et al., 1998). Some studies emphasised that the treatment failure rate is 20% with triple therapy regimes in which metronidazole is the drug of choice (Megraud and Lamouliatte, 2003). The metronidazole resistance frequency around the world (in ascending order) is 30.5% in

North America, 31.1% in Europe, 46.5% in Asia, 52.8% in South America and 75% in Africa. Around one third of the identified *H. pylori* strains in developed countries are metronidazole resistant, and the prevalence of resistance is very high in developing countries (Ghotaslou et al., 2015).



**Figure 1.5** Anti-*H. pylori* antibiotic resistance rates in different continental areas around the world, the data adapted from (Ghotaslou et al., 2015).

The association between socioeconomic differences and this resistance can be explained by the frequency of use of metronidazole and derivatives for related infections such as parasitic, dental and gynaecological diseases (Frenck and Clemens, 2003).

Prevalence analysis indicated that metronidazole resistance has remained stable in Asia, Europe and North America but is markedly increasing in African countries with 51.3% in 2010 and 85% in 2013 (Mansour et al., 2010, Seck et al., 2013). Metronidazole resistance in 2014 has remained unchanged since 2009 in Europe. According to these analysis, metronidazole is favoured over amoxicillin in first-line therapy in patients from Asia, Europe and North America but not African patients (Ghotaslou et al., 2015).

In conclusion, the most frequently found resistance in *H. pylori* is to metronidazole and the lowest resistance is to amoxicillin. However, the most globally effective resistance is to clarithromycin which has been increasing dramatically in the recent years. Furthermore, levofloxacin resistance has significantly increased although worldwide resistance level is low. These data are alarming and hence novel therapies are required for eradication. In the immediate term it is important to develop alternative strategies to use the current drugs against *H. pylori* effectively. This may be achieved by understanding the genetics, physiology and metabolism of the pathogen.

#### **1.4 Genome and strain diversity of *H. pylori***

*H. pylori* 26695, is a virulent pathogenic strain isolated from a gastritis patient in the United Kingdom. The complete genome of the bacteria has been sequenced using whole-genome random sequencing and was published in 1997 by Tomb and his research team. Sequence analysis indicated that the pathogen has an efficient motility system, a well developed mechanism of iron scavenging and highly active DNA restriction and modification systems. The genome of this strain has a circular chromosome of 1,667,867 base pair (bp) with average guanine + cytosine (G+C) content of 39% and 1,590 predicted coding sequences. These 1,590 predicted open reading frames (ORFs), account for 91% of the chromosome, and have an average size of 945 bp (which is comparable to that in other bacteria). Three classes account for the remaining 9% non-coding regions, stable RNA (0.7%), non-coding repeats (2.3%) and intergenic sequences (6%). 1,091 of the total 1,590 ORFs were found to have counterparts in other organisms. This allowed hypothetical biological functions to be assigned to most of them, although not all have orthologues of identified function. The other 499 ORFs may be considered specific to *H. pylori* as they exhibited no database matches (Tomb et al., 1997).

Subsequently, the *H. pylori* 26695 coding sequences were re-examined by Alm and co-workers and they revised the annotations. The total number of ORFs reported in this study was 1,552, of which 1,185 have orthologues in other species, 367 are *H. pylori* specific, and 69 are specific to strain 26695 (Alm et al., 1999). A second reannotation of *H. pylori* 26695 conducted by Boneca et al. in 2003 in the light of sequences from other

strains becoming available and led to the creation of a specific database for *H. pylori*, PyloriGene (<http://genolist.pasteur.fr/PyloriGene>). From this, the percentage of hypothetical proteins decreased from 40% to 33%, and 108 coding sequences had their functions reassigned. The most recent reannotation of this strain was published in 2013, where 1573 coding sequences were reviewed from *H. pylori* 26695's genome. 1212 ORFs were identified with functions, 712 of these genes encoding proteins are metabolic and 500 non-metabolic, while 191 have new functions specific to the strain. The number of hypothetical proteins was 361 which represents approximately 23% of the genome (Resende et al., 2013).

The metabolic network reconstruction of a sequenced organism can be created from the gene functional annotation. Based on the biochemical reactions catalysed by the enzymes encoded by the annotated genes of the organism, a reconstruction allows the development of a genome-scale metabolic model. These models can be used for *in silico* simulation of the phenotype of an organism under different environmental pressures. This represent an important strategy in metabolic engineering design and the discovery of new drug targets for pathogens (Rocha et al., 2008). Two metabolic models of *H. pylori* have been published. The first model, iCS291, contains 291 genes and 388 reactions (Schilling et al., 2002). Developing this, a new model was reconstructed, ilt341 GSM/GPR, with 341 gene and 476 reactions (Thiele et al., 2005).

### **1.5 Metabolism within *H. pylori***

As stated previously, the entire genome sequencing of *H. pylori* 26695 resulted in the identification of genes responsible for virulence, replication and metabolism. This, in conjunction to the classical biochemical findings has helped to explain the pathogen's ability to colonise the human gut, and its disease association. The genome sequence also allowed the reconstruction of two genome-scale metabolic models based on experimental knowledge derived from other related species. However, this *in silico* modelling hasn't given a comprehensive understanding of *H. pylori* metabolism. This means genome sequence analysis must be followed by experimental investigation to confirm or disprove the metabolic role of the identified gene products. Conclusively, sequencing data and computational metabolic models cannot be enough to elucidate

all aspects of *H. pylori* metabolism, but experimental investigations can provide the greatest amount of understanding into the pathogen's metabolism, where hypotheses based upon genomic data and *in silico* metabolic studies can be tested (Eppinger et al., 2004). The continuation of this debate will be the main subject in the following chapters of this thesis. This will show the importance of experimental work to answer questions about two genetically related metabolic genes in *H. pylori* 26695. The following sections focus on the central metabolism of *H. pylori*.

### 1.5.1 Carbohydrate metabolism

Physiological studies of *H. pylori* provided evidence that glucose can be metabolised through both fermentative and oxidative pathways, despite being an obligate microaerophile. Furthermore, glucose is the only carbohydrate metabolised by the bacteria (Mendz and Hazell, 1993, Mendz and Hazell, 1991). The genome sequence analysis of the bacteria also suggested that glucose is the main source for carbohydrate metabolism and substrate-level phosphorylation as *H. pylori* encodes glucokinase rather than hexokinase (Tomb et al., 1997), and has a proper sodium dependent transporter only for glucose uptake (Psakis et al., 2009). This characteristic could explain the very restricted carbohydrate range utilised by *H. pylori*.

Radioactive tracer analysis and  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance spectroscopy (NMR) studies emphasised that glucose utilisation in *H. pylori* showed biphasic behaviour. The first phase starts with a slower initial period followed by a faster catabolism phase with a rate of utilisation at least an order of magnitude faster. In both phases the rate of decline in glucose levels varied between strains and depended on the growth conditions. This suggest that glucose is not a preferred metabolite but can be used when other energy substrates have been depleted (Mendz et al. 1993).

In an experimental study, enzymatic activities of oxidative and non-oxidative parts of the carbohydrate utilisation pathway in *H. pylori* were demonstrated, specifically the pentose-phosphate pathway which constitutes an important source for providing reductive biosynthetic reactions with NAD(P)H, in addition to the phosphorylated pentoses that are required for nucleotide biosynthesis (Mendz and Hazell, 1991). All the

genes coding the enzymes of the pentose phosphate pathway were verified (in the genome sequencing data of *H. pylori* 26695) except an orthologue to the gene coding 6-phosphogluconate dehydrogenase (*gnd*). This suggests the presence of a different ORF coding an enzyme that has similar function to 6-phosphogluconate dehydrogenase (figure 1.6).

In 1996 glycolytic and gluconeogenic activity were experimentally detected in *H. pylori* (Hoffman et al., 1996). Subsequently, the initial metabolic interpretation of the genetic analysis data emphasized that glycolysis-gluconeogenesis is the main metabolic axis for energy production and the initiation point of many biosynthetic pathways (Tomb et al., 1997). Figure 1.6 shows that these reciprocal pathways are shared with seven reversible reactions and are distinguished by three irreversible steps for each one (figure 1.6). There has been an argument about the completeness of glycolysis as some studies failed to detect apparent activity of some glycolytic enzymes (Chalk et al., 1994, Mendz et al., 1994a). In particular, phosphoglycerate mutase activity hasn't been observed although a gene coding its sequence was identified (Marais et al., 1999). Moreover, glycolysis lacks two key irreversible enzymes namely phosphofructokinase and pyruvate kinase (Schilling et al., 2002). The absence of detectable enzymatic activity and lack of corresponding open reading frames (ORFs) further support these findings (Tomb et al., 1997, Marais et al., 1999).

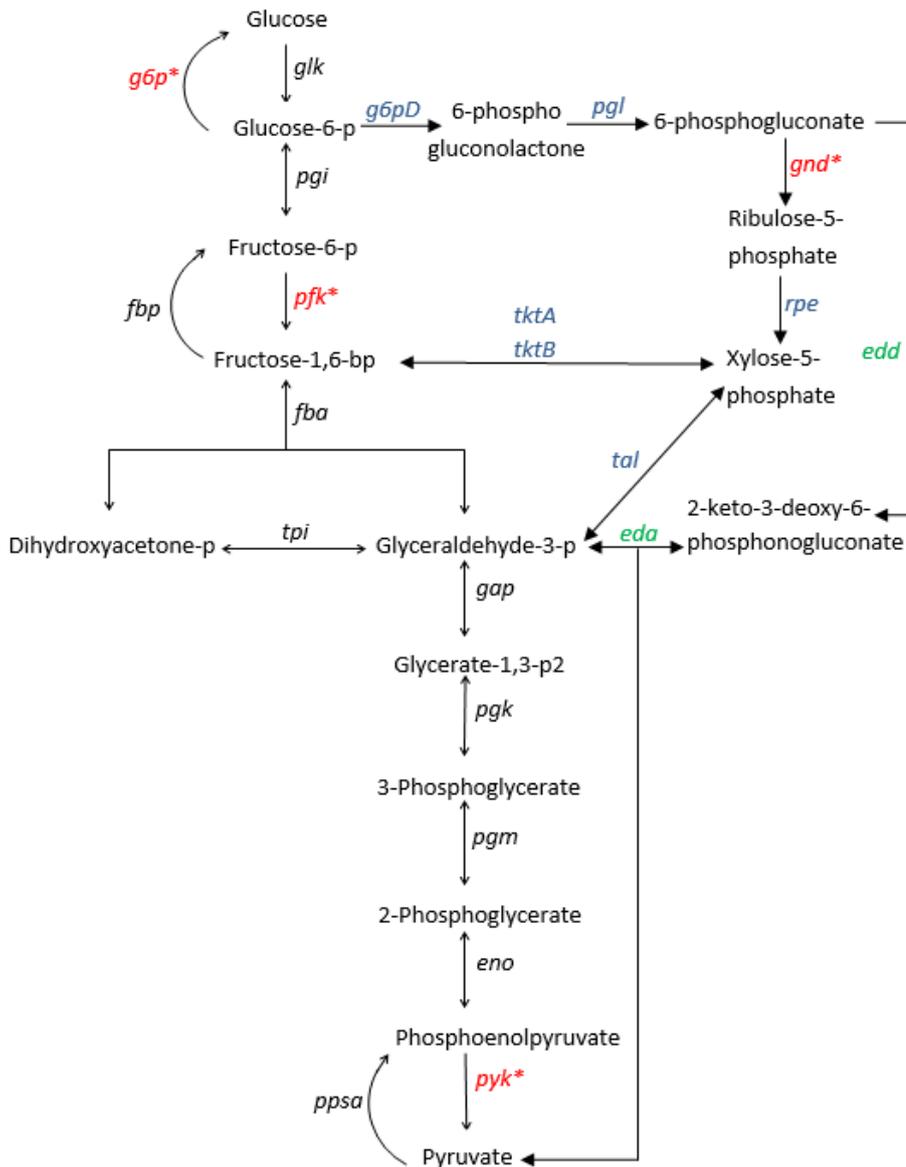
However, this defect can be substituted as glucose-6-phosphate can be utilised by the Entner-Doudoroff pathway. The steps of this pathway involve two dehydratase and aldolase catalysed reactions. These produce pyruvate and glyceraldehyde-3-phosphate from 2-keto-3-deoxygluconate-6-phosphate (figure 1.6). These enzymes were experimentally detected in the cell lysate of *H. pylori* and the open reading frames identified in the pathogen genome (Menz et al., 1994a, Tomb et al., 1997). The Entner-Doudoroff pathway produces less energy than that from glycolysis but it is considered an alternative route and is a mode for utilising aldonic acids. The pathway is inducible in *E. coli*, whereas it is constitutively active in *H. pylori* (Wanken et al., 2003).

Despite the lack of two key enzymes of glycolysis, the gluconeogenic pathway is complete in *H. pylori* including the two key role irreversible steps catalysed by

phosphoenolpyruvate synthase and fructose-1,6-bisphosphatase, in addition to the other corresponding reversible steps in glycolysis. However, no gene was identified for the enzyme responsible for catalysing the conversion of glucose,6-phosphate to glucose in the final step of gluconeogenesis that is glucose-6-phosphatase (Tomb *et al.*, 1997). This is consistent with the finding that there is no storage for glucose in *H. pylori* (Marais *et al.*, 1999). On the other hand, a gene has been detected for glucokinase (the enzyme in charge of the first irreversible reaction in glycolysis (Tomb *et al.*, 1997)).

Figure 1.6 shows the classical pathways responsible for carbohydrate metabolism namely glycolysis, gluconeogenesis, pentose phosphate and Entner-Doudoroff pathways. None of these pathways functions in complete isolation but all of them feed into each other (Elliott, 2009). For a complete picture of central metabolism understanding in *H. pylori*, the genetic analysis of glycolysis and gluconeogenesis needs to be verified by further experimental studies to investigate the presence and functions of the components of these pathways in the enteric pathogen.

At the time of writing, a number of experimental genetic, structural and mechanistic studies have been conducted to characterize some of the enzymes in glycolysis-gluconeogenesis pathways in *H. pylori*. For example biophysical and kinetic investigation indicated apparent activity of the gluconeogenic enzyme fructose-1,6-bisphosphatase (FBPase) (Ayna, 2016). Furthermore, kinetic and structural study focused on glyceraldehyde-3-phosphate dehydrogenases (GAPDHs) demonstrated that one of the two identified GAPDHs strongly prefers NADPH over NADH, this is similar to the enzyme in gluconeogenic photosynthetic organisms and archaea (Elliott, 2009). In addition to the other findings, this hypothesised that *H. pylori* might be a gluconeogenic organism. The remaining chapters in this thesis will focus on two metabolic enzymes in *H. pylori* that is the first gluconeogenic specific enzyme phosphoenolpyruvate synthase (PPSA) and the glycolytic-gluconeogenic enzyme phosphoglycerate kinase (PGK).



**Figure 1.6** Glycolysis, gluconeogenesis, pentose phosphate and Entner-Doudoroff pathways in *H. pylori*. The genes encoding for the enzymes in the pathways, shown in italic, are: Glycolysis (enzyme genes names written in black): *glk*: glucokinase, *pgi*: phosphoglucose isomerase, *pfk*: phosphofructokinase, *fba*: fructose-1,6-bisphosphate aldolase, *tpi*, triose phosphate isomerase, *gap*: glyceraldehyde-3-phosphate dehydrogenase, *pgk*: phosphoglycerate kinase, *p gm*: phosphoglycerate mutase, *eno*: enolase and *pyk*: pyruvate kinase; Gluconeogenesis (written in black): the same as glycolysis but with the exception of irreversible reactions, i.e., *ppsa*: phosphoenolpyruvate synthase, *fba*; fructose-1,6bisphosphatase and *g6p*: glucose-6 phosphate; Pentose phosphate (enzyme genes names written in blue): *g6PD*: glucose-6-phosphate dehydrogenase, *p gl*: 6-phosphogluconolactonase, *gnd*: 6-phosphogluconate dehydrogenase, *rpe*: phosphopentose isomerase and *tal*: transaldolase transketolase; Entner-Doudoroff (enzyme genes names written in green): *edd*: 6-phosphogluconate dehydratase and *eda*: 2-keto-3-deoxy-6-phosphogluconate aldolase. The enzyme genes shown in red with asterisks are not identified in the published genetic data in *H. pylori*. The figure adapted from Marais et al., 1999 and Schilling et al., 2002.

## 1.5.2 Pyruvate metabolism

As stated in the last section, pyruvate is the end product of Entner-Doudoroff and glycolytic pathways in *H. pylori*. The metabolic fate of this product was investigated in several studies. When investigated under anaerobic growth conditions, pyruvate is metabolised to lactate, ethanol and acetate whereas the end product of pyruvate metabolism under aerobic conditions is acetate (Chalk et al., 1994). As described in section 1.2.2, the true physiological conditions for optimum growth of *H. pylori* are micro aerobic with only 5% of oxygen. Cells incubated under these conditions metabolised pyruvate to acetate, lactate, succinate, formate and alanine. The presence of alanine suggests that pyruvate can play a key role in biosynthetic processes, and succinate production support the notion of pyruvate incorporation into active Krebs cycle (Mendz et al., 1994b).

The formation of acetate, ethanol and lactate emphasises the role of pyruvate in fermentative metabolism (Mendz et al., 1994b). In agreement with these observations, *H. pylori* genome sequence analysis verified the presence of the sequence encoding D-lactate dehydrogenase (which catalyses the conversion of pyruvate to lactate), and the ORF corresponding to alcohol dehydrogenase. Moreover, sequence comparison with similar genes in *Haemophilus influenza* suggests that these enzymes catalyse reactions taking place in conjunction with NADH oxidation (Tomb et al., 1997).

Pyruvate must be converted to acetyl-CoA for entry into the Krebs cycle and formation of acetate. In *H. pylori*, the oxygen sensitive and flavodoxin dependant pyruvate: acceptor oxidoreductase (PFOR) catalyses the oxidative decarboxylation of pyruvate instead of mixed-acid fermentation by the aerobic pyruvate dehydrogenase (AceEF) or the strictly anaerobic pyruvate-formate lyase (Pfl). Orthologues of this enzyme are found commonly in obligatory anaerobic bacteria such as *Clostridium* spp. PFOR is a multi-subunit enzyme and is similar to pyruvate-ferrodoxin oxidoreductases previously detected in hyperthermophilic organisms (Hughes et al., 1998). Genome analysis data showed PFOR has four subunits encoded by the *porGDAB* operon (hp 1108, hp1109, hp1110 and hp1111 respectively), consistent with these experimental observations and the microaerophily of *H. pylori* (Tomb et al., 1997, Hughes et al., 1998).

Genetic reannotation analysis revealed that hp1164, incorrectly originally annotated in Tomb's et al. paper as a thioredoxin reductase, was a member of the class of NADPH oxidase proteins. This uniquely conserved essential gene among epsilon proteobacteria was designated as flavodoxin quinone reductase (FqRB). The experimental study showed that this enzyme associated with the reversible PFOR:NADPH reductase complex system to regenerate flavodoxin, by reducing NADP (figure 1.7). This process apparently has an important role in *H. pylori* in driving CO<sub>2</sub> fixation, which is essential for replenishing pyruvate utilised by gluconeogenesis (St. Maurice et al., 2007).

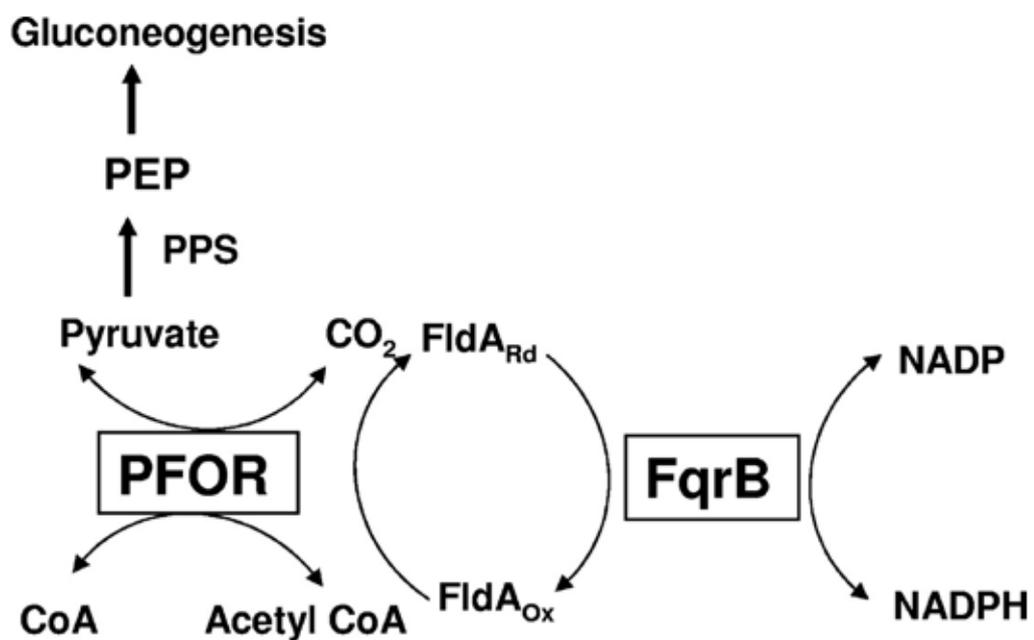


Figure 1.7 Reversible PFOR:NADPH reductase complex in *H. pylori*. PEP: phosphoenolpyruvate, PPS: phosphoenolpyruvate synthase, PFOR: pyruvate:flavodoxin oxidoreductase, FqrB: flavodoxin quinone reductase, FldA<sub>ox</sub>: flavodoxin oxidised and FldA<sub>Rd</sub>: flavodoxin reduced. Pyruvate formation (reverse reaction) may be favored due to the unique gluconeogenic nature of central metabolism in microaerophiles where only Phosphoenolpyruvate synthase (PPSA) found (as in *H. pylori*). This figure taken from (St. Maurice et al., 2007).

### 1.5.3 The Krebs cycle

Microaerophiles like *H. pylori* effectively colonise and thrive in the gastrointestinal mucosa and tend to move deeply in the crypts and near the inner epithelial cells. Accordingly, their central metabolism is highly conserved and well adapted for this low oxygen environment. Ironically, *H. pylori* utilises glucose sparingly in the stomach because the primary carbon and energy sources are amino acids and organic acids. The bacteria also lack anaplerotic reactions that connect the Krebs cycle with gluconeogenesis, for example linking oxaloacetate with phosphoenolpyruvate, hence the pathogen must rely on phosphoenolpyruvate synthase (PPSA) to satisfy gluconeogenic requirements. The latter are essential intermediates for anabolism of vitamins, nucleic acids and cell wall material (Tomb et al., 1997, Alm et al., 1999, Doig et al., 1999, Schilling et al., 2002, Nagata et al., 2003).

The Krebs cycle in *H. pylori* has been an example of how poor experimental biochemical knowledge and/or inaccurate extrapolation of genomic analysis can lead to incorrect modelling of a metabolic pathway. Initially, this cycle was hypothesised to be a non-cyclic, branched pathway, with oxidative and reductive arms. The dicarboxylic acid arm was proposed to be composed of reductive reactions directed toward the production of oxaloacetate, malate, fumarate and succinate respectively. On the other arm, tricarboxylic acid oxidative reactions proceed from oxaloacetate through citrate and isocitrate to  $\alpha$ -ketoglutarate (Pitson et al., 1999). This hypothesis was supported by genome sequence analysis as genes encoding citrate synthase (*gltA*, hp0026), aconitase (*acnB*, hp0779), isocitrate dehydrogenase (*icd*, hp0027), fumarase (*fumC*, HP1325) and fumarate reductase (*frdABC*, hp0191 to hp0193) were identified, ORFs similar to the genes encoding  $\alpha$ -ketoglutarate dehydrogenase were absent (Tomb et al., 1997). The latter result confirmed the apparent absence of  $\alpha$ -ketoglutarate dehydrogenase activity. Conversely, the existence of  $\alpha$ -ketoglutarate:acceptor oxidoreductase (OOR), which catalyses the conversion of  $\alpha$ -ketoglutarate to succinate, was experimentally identified (Hoffman et al., 1996, Pitson et al., 1999). Purification and genetic analysis studies showed that this ferredoxin -dependant enzyme is organised in a complex of four heterogeneous subunits and shown to be oxygen sensitive (Hughes et al., 1998).

Apparent activities of malate dehydrogenase and malate synthase were detected experimentally in *H. pylori* (Pitson et al., 1999), but no corresponding gene sequences encoding these enzymes were identified, suggesting that the gene sequences are different to the other related orthologues so far identified (Marais et al., 1999). The lack of a malate dehydrogenase ORF in the genome, but the apparent generation of oxaloacetate, suggested another non-classical enzyme was fulfilling this role, that is malate quinone oxidoreductase (MQO) which converts malate to oxaloacetate (Kather et al., 2000). A succinyl-CoA:acetoacetyl-CoA transferase was characterised, the gene sequenced and enzyme activity shown to convert succinyl-CoA to succinate. This reaction is dependent on the continual supply of acetoacetate and breakdown of acetoacetyl-CoA (Corthésy-Theulaz et al., 1997). Another alternative might be regeneration of acetoacetate from acetoacetyl-CoA, possibly in metabolic energy generating reactions.

In conclusion, the presence of  $\alpha$ -ketoglutarate:ferredoxin oxidoreductase, malate quinone oxidoreductase and succinyl-CoA:acetoacetyl-CoA transferase indicated that the Krebs cycle in *H. pylori* is complete, but it deviates from the standard textbook examples with these three enzymes catalysed steps (figure 1.8) (Kather et al., 2000).

Interestingly, fumarate, through fumarate reductase, may act as the terminal electron acceptor in aerobic respiration. *In vivo* experimental activity of the enzyme in *H. pylori* showed that it is a potential drug target against the bacteria (Mendz et al., 1995). The fumarate reductase operon was cloned and characterised revealing the enzyme is composed of three subunits encoding by three structural genes (Ge et al., 1997).

More recently, kinetic studies indicated that Krebs cycle in *H. pylori* is a potential molecular target of bismuth as the metallodrug is capable of binding to fumarase. Characterisation of interaction and the inhibitory effects of bismuth to fumarase showed that each fumarase monomer binds one equivalent of bismuth ( $\text{Bi}^{3+}$ ). Moreover, circular dichroism (CD) spectroscopy and kinetic analysis reported negligible secondary structure change and non-competitive inhibition behaviour (as reflected by the analysis of Lineweaver–Burk plots) respectively (Chen et al., 2012).

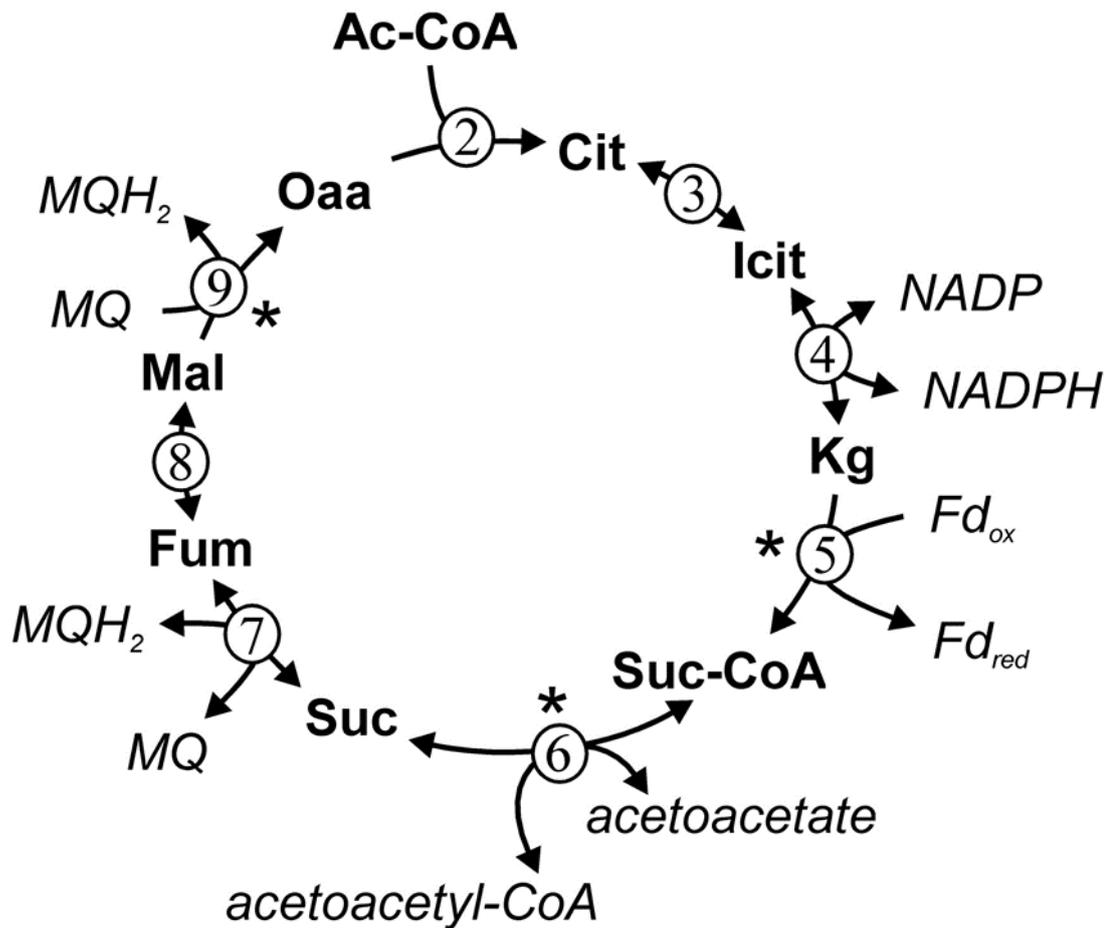
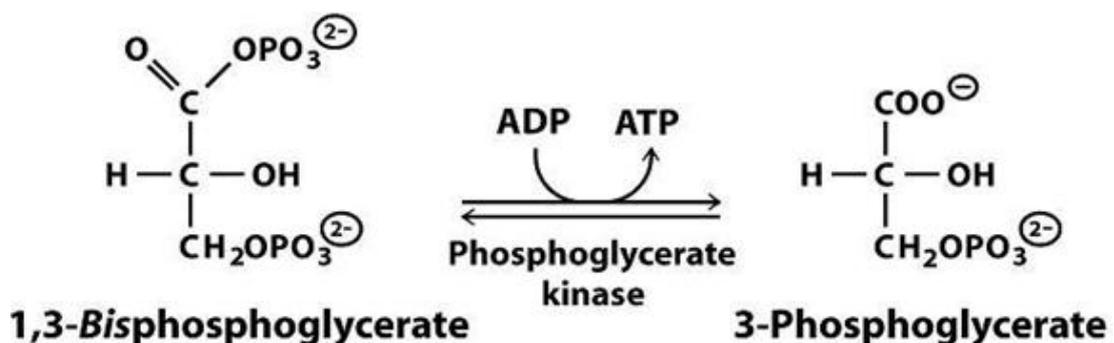


Figure 1.8 Krebs cycle in *H. pylori*, unusual enzymes denoted with asterisks. Citrate synthase (2) catalyse the condensation of acetyl-CoA (Ac-CoA) with oxaloacetate (Oaa) releasing CoA and producing citrate (Cit). This product is isomerised to isocitrate (Icit) by aconitase (3) in dehydration followed by hydration steps. The oxidative decarboxylation of isocitrate to  $\alpha$ -ketoglutarate (Kg) is stimulated by NADP-dependent isocitrate dehydrogenase (4). Further oxidative decarboxylation of  $\alpha$ -ketoglutarate to succinyl-CoA (Suc-CoA) by  $\alpha$ -ketoglutarate:ferredoxin oxidoreductase (5) catalysis, instead of the common  $\alpha$ -ketoglutarate dehydrogenase, with reduced ferredoxin byproduct. Succinate (Suc) is produced as succinyl-CoA acetoacetyl transferase (6), deviates from the standard text book examples, is able to transfer CoA to acetoacetyl. Succinate oxidation by fumarate reductase (7) leads to fumarate (Fum) formation, which is catalysed by fumarase (8) to produce the last step reactant that is malate (Mal). The oxidation of the latter to oxaloacetate is stimulated by malate:quinone oxidoreductase (9), the third deviated enzyme of *H. pylori* Krebs cycle. Figure adapted from (Kather et al., 2000).

## 1.6 Phosphoglycerate kinase and Phosphoenolpyruvate synthase

### 1.6.1 Phosphoglycerate kinase

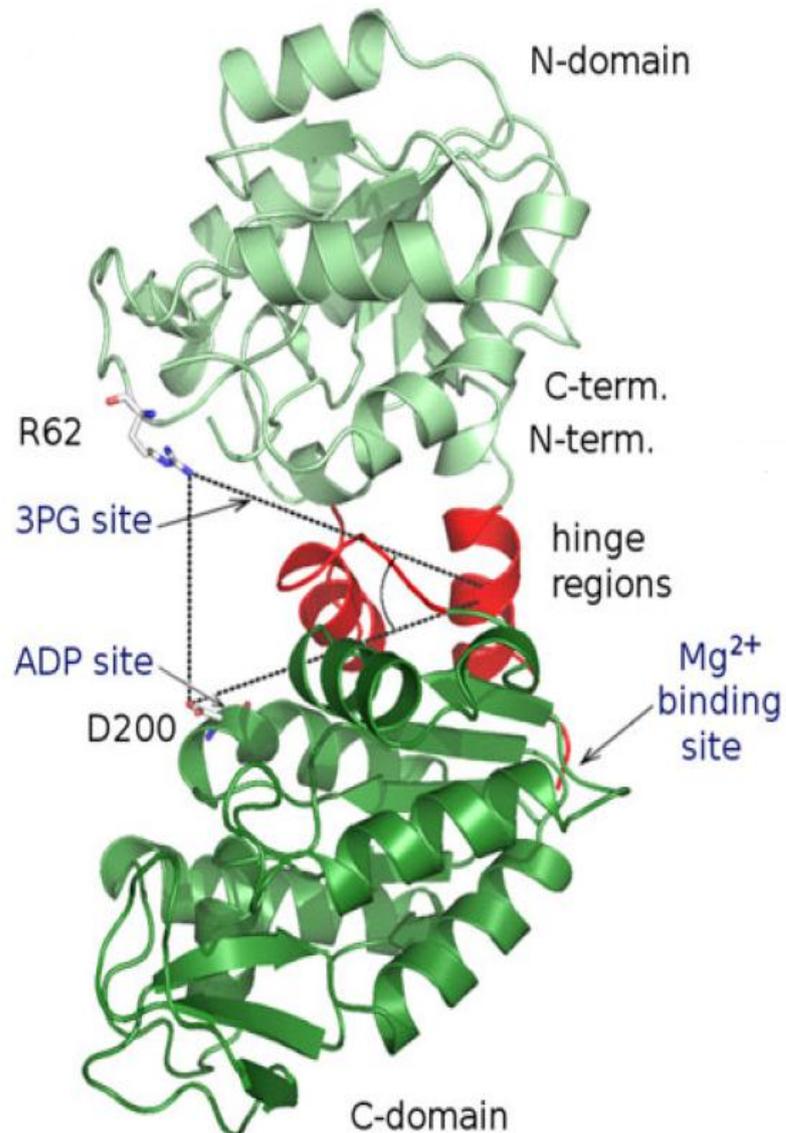
Phosphoglycerate kinase (PGK) or [ATP: 3-phosphoglycerate1-phosphotransferase, E.C.2.7.2.3] is a bi-substrate magnesium-dependant enzyme with a highly conserved amino acid sequence found in virtually all prokaryotic and eukaryotic organisms, with only one PGK encoded per genome. However, in addition to PGK1 (active in somatic cells) another sperm-specific isoform PGK2 exists in mammalian cells (Zheng et al., 2012). This enzyme catalyses the transfer of phosphoryl group by the conversion of 1,3-bis-phosphoglycerate (1,3-BPG) and ADP to 3-phosphoglycerate (3-PGA) and ATP in the glycolytic pathway, and catalyses the reverse reaction for gluconeogenesis and in the Calvin cycle of photosynthetic organisms (Figure 1.9) (Scopes, 1973). The enzyme is a monomer having two binding sites; the nucleotide binding site is located in the C-terminal domain, and binds to ATP or ADP while the other catalytic site is located in the N-terminal domain and binds 1,3-BPG or 3-PGA (Watson et al., 1982). PGK is an example of a flexible protein, with substrate induced conformation change, and comprehensive studies have documented its domain closure mechanism during catalysis (Vas et al., 2010).



**Figure 1.9** The schematic representation of the reversible reaction catalysed by PGK in which 1,3-Bisphosphoglycerate converts to 3-Phosphoglycerate. The arrow shows the conversion of ADP to ATP.

Understanding the mechanism and function of this enzyme is important as two ATP molecules are produced in this sixth reaction step of glycolysis from each 6-carbon sugar. PGK is a crucial target for disrupting anaerobic metabolism in pathogens as it is a required step in the anaerobic degradation of glucose to pyruvate (Zheng et al., 2012).

Structural studies of PGK from numerous species revealed that the enzyme is found in two main conformations; the open conformation where the C-terminal and N-terminal domains are free of substrates and far apart from each other, and the closed conformation in which the substrates bring these domains together in a hinge-bending closure which is required to facilitate the catalysis of phosphate transfer (figure 1.10) (Bernstein et al, 1997 and (Pickover et al., 1979).



**Figure 1.10** Crystal structure of *Bacillus anthracis* phosphoglycerate kinase (BA PGK). The N-domain is shown in light green, with the C-domain in dark green and the hinge region in red. The substrate (3,PGA or 1,3-BPG) binding site, nucleotide binding site (ADP or ATP) and Mg<sup>2+</sup> binding site are indicated with arrows (Taken from Zheng et al., 2012).

The N-terminal domain has a “basic patch” of residues proposed to have a regulatory rather than catalytic role as it binds and orients 1,3-BPG, and it has a binding site for anions. Two different crystal structures open forms of *Plasmodium falciparum* PGK were determined. In both structures a sulphate ion is bound in the nucleotide binding site in addition to a second sulphate ion located at the basic patch in one of the crystal forms (Smith et al., 2011).

### 1.6.2 PGK as drug target

Enzymatic and crystallographic studies have shown that PGK is a potential drug target in parasitic Trypanosomid protozoa (Verlinde et al., 2001) and (Kotsikorou et al., 2006) in spite of being a conserved enzyme across virtually all eukaryotic and prokaryotic organisms. The reason behind this is that there is an absence of a functional Krebs cycle or mitochondrial respiratory chain joined to ATP synthesis in trypanosomatids. This means the glycolytic pathway is the only source of ATP production in blood stream stages of the parasite’s life cycle (Opperdoes, 1987).

Targeting crucial glycolytic enzymes for ATP production (like PGK) can inhibit the pathway sufficiently to kill glycolysis-dependent organisms such as *Trypanosoma* and *Leishmania*. Due to the wide range of evolutionary variation between human and parasites (Fernandes et al., 1993), and the unusual organization of the central metabolism system in trypanosomes (Opperdoes and Borst, 1977), these can be exploited to develop PGK or other glycolytic enzymes inhibitors that specifically affect the metabolism of parasite rather than the metabolism of the host. This can leave the door open for future studies to target PGK with another key role enzyme in Krebs cycle of *H. pylori*, which deviates from the text book standard example (section 1.5.3).

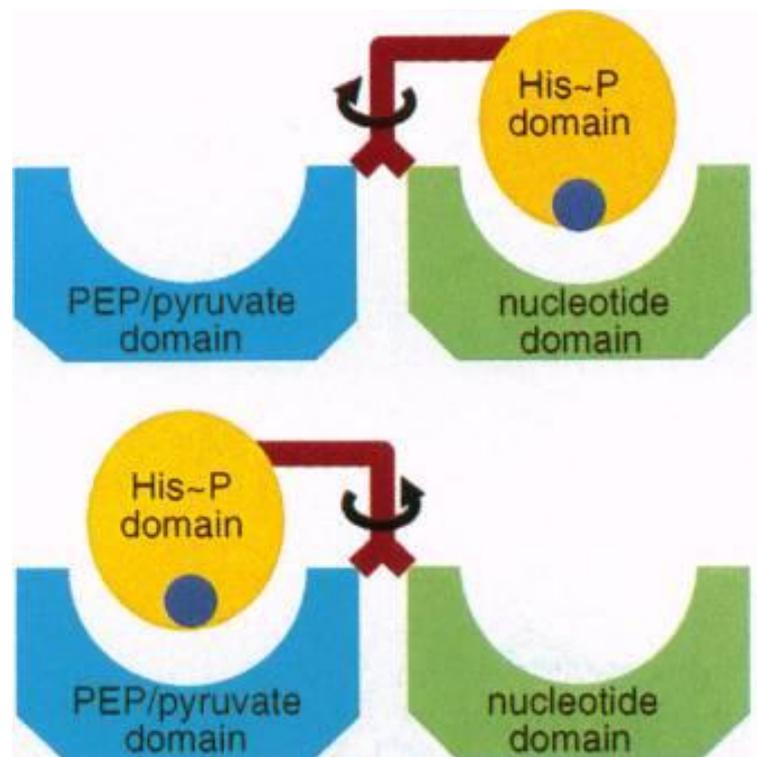
### 1.6.3 Phosphoenolpyruvate synthase

Phosphoenolpyruvate synthase (PPSA) or pyruvate water dikinase (EC 2.7.9.2) catalyses the first irreversible reaction of gluconeogenesis in the metabolic models that have been described for *H. pylori* (Marais et al., 1999 and Schilling et al., 2002). This step comprises the conversion of ATP and pyruvate to AMP, phosphoenolpyruvate and

Phosphate (Cooper and Kornberg, 1965) using the following mechanism (Cooper and Kornberg, 1967):



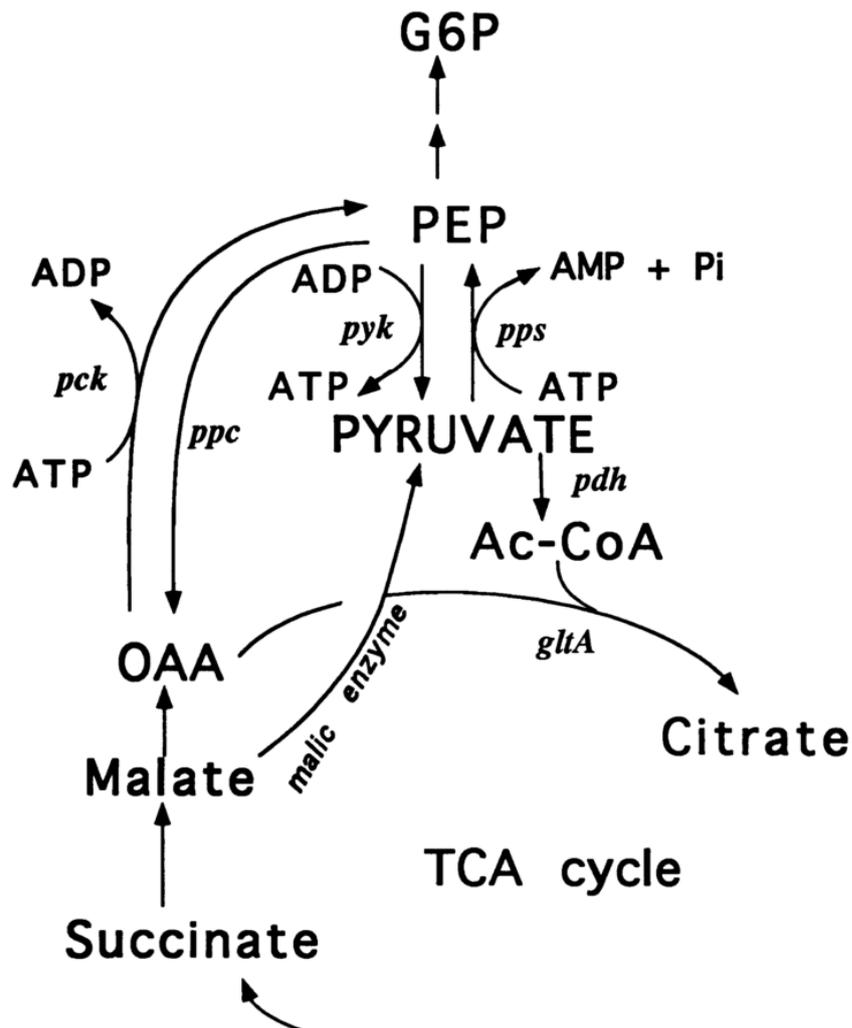
In the first step, ATP binds to the N-terminal domain of PPSA followed by hydrolysis of the  $\beta$ -phosphate group by a catalytic histidine residue in the central domain (Narindrasorasak and Bridger, 1977). This leads to the generation of AMP, inorganic phosphate ( $\text{P}_i$ ) and specific PPSA-phosphate intermediate complex. Following the formation of the phospho-enzyme, pyruvate binds to the C-terminal domain and transfers the  $\beta$ -phosphate group in a swivelling domain mechanism (figure 1.11) to form phosphoenolpyruvate (Herzberg et al., 1996).



**Figure 1.11** Simplified schematic illustration of the two step mechanisms catalysis of phosphoenolpyruvate synthase. Green box: N-terminal domain, blue box: C-terminal domain, red stick: linker peptides, yellow circle: central phospho-histidine domain, blue circle: catalytic histidine residue. The arrows demonstrate the swivelling transfer of  $\beta$ -phosphate from the N-terminal domain to the C-terminal domain (Taken from Herzberg *et al.*, 1996).

To date, there are no available structural studies on PPSA to verify this mechanism in the three dimensional context. However, as of 2017 there is only one available PPSA structure (2OLS), belonging to the pathogenic bacteria *Neisseria meningitides*, in the protein data bank (PDB) but the study has yet to be published.

There is a growing body of genetic and kinetic literature regarding the understanding of the mechanism and function of bacterial and archaeal PPSA (Imanaka et al., 2006) and (McCormick and Jakeman, 2015). Some of these studies show that the reactions catalysed by PPSA and phosphoenolpyruvate carboxykinase (PCKA) are controlling gluconeogenesis, carbohydrate synthesis and the intermediate supply of carbon to the citric acid cycle, when *H. pylori* is grown in minimal medium supplemented with succinate and pyruvate (Chao et al., 1993). Another study highlighted the essentiality of this enzyme for the growth of some enteric bacteria that have functional PPSA such as *E. coli* and *Salmonella typhimurium* when grown in media providing three-carbon compounds like pyruvate, lactate or alanine, as sole carbon sources. Nevertheless, when these bacteria are grown on six-carbon sources such as glucose, PPSA undergoes catabolite repression (Cooper and Kornberg, 1965, Smyer, 1989). Consequently, pyruvate kinase (PYK) and/or phosphoenolpyruvate carboxylase (PPC) will act to convert phosphoenolpyruvate to pyruvate and oxaloacetate (OXA) respectively (Figure1.12). In *H. pylori*, PPSA can play a key role in the metabolism as a potential unique preliminary gluconeogenic enzyme as there are no corresponding ORFs for PCKA, PYK and PPC in the pathogen. In the other microorganisms and/or mammals which lack ORF for PPSA, PCKA is one of the specific rate-controlling enzymes for gluconeogenesis, in addition to fructose-1-6-bisphosphatase and glucose-6-phosphatase (Pilkis and Granner, 1992). Understanding the essentiality of hpPPSA will open the door to identify their importance for the metabolism of this pathogen.

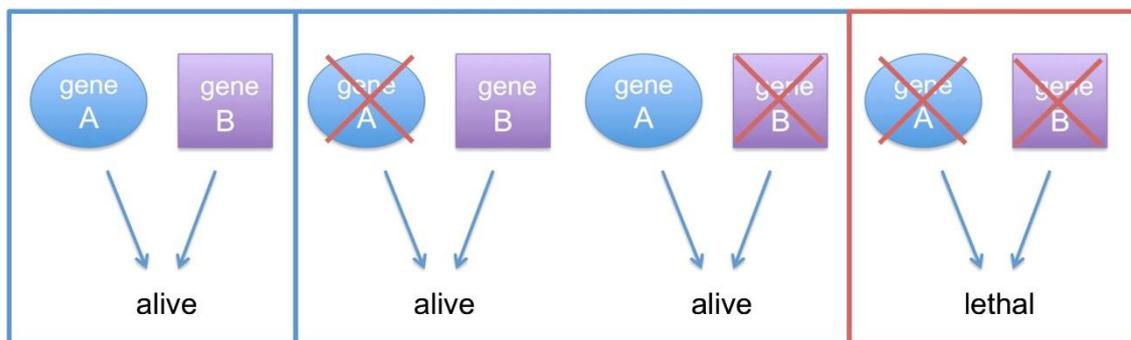


**Figure 1.12** Schematic of parts of glycolytic and Krebs cycle (tricarboxylic acid cycle (TCA) pathways in *E. coli*. The genes encoding for the enzymes, shown in italic, are: *pps*, phosphoenolpyruvate synthase; *pyk*, pyruvate kinase; *pck*, phosphoenolpyruvate carboxykinase; *ppc*, phosphoenolpyruvate carboxylase; *pdh*, pyruvate dehydrogenase; *gltA*, citrate synthase. Abbreviations: G6P, glucose 6-phosphate; Ac-CoA, acetyl coenzyme A; OAA, oxaloacetate; TCA, tricarboxylic acid (Taken from (Chao et al., 1993).

### 1.6.4 Gene essentiality of *ppsA*

A number of experimental and computational genetic essentiality studies on *H. pylori* have been carried out. These studies employed different mutational strategies such as genome prioritization and allelic replacement (Chalker et al., 2001), random mutagenesis and loop amplification (RMLA) (Jenks et al., 2001), microarray tracking of global transposon mutagenesis (MATT) (Salama et al., 2004) and metabolic modelling through computational deletion studies (Thiele et al., 2005, Schilling et al., 2002) was used to identify the essential and/or nonessential genes in the pathogenic bacteria.

With single enzyme-coding gene mutation studies, *ppsa* was frequently designated a nonessential gene for *H. pylori* viability, like the majority of metabolic genes, in spite of playing a key role in its metabolism (Jenks et al., 2001, Salama et al., 2004). This can be explained as the metabolic system is protected against such mutations by plasticity and redundancy, these strategies give the organism the ability to control phenotypic variation during environmental changes or upon malfunction of some of its parts (Güell et al., 2014). Nevertheless, the organism can fail to survive if synthetic lethals (SL), or double deletion mutants occur, which can be defined as the prohibition of growth by simultaneous deletion of pairs of nonessential genes which become essential for biomass production (Suthers et al., 2009) (figure 1.13).



**Figure 1.13** Schematic representation of synthetic lethality. A pair of genes are synthetic lethal only when their simultaneous inactivation results in cellular death. In this example, deletion of either gene A or gene B does not affect viability whereas inactivation of both simultaneously is lethal (adapted from Nijman, 2011).

Computational models have suggested that *H. pylori* would fail to survive if a *ppsa* deletion mutation was combined with other glycolytic nonessential gene knock-outs, including phosphoglycerate kinase (*pgk*), phosphoglycerate mutase (*pgm*), enolase (*eno*) or glyceraldehyde-3-phosphate dehydrogenase (*gap*). Assimilation of the *H. pylori* PPSA reaction in rich medium verified that these double deletion mutants became essential for biomass production by affecting phosphatidyl ethanolamine (PE) and phosphatidyl serine synthesis (Thiele *et al.*, 2005). These phospholipids which are in the *H. pylori* cell membrane are reported as one of the essential constituents for their biomass generation and maintenance (Schilling *et al.*, 2002). This suggests the potential

to use *ppsa* SL combinations as an anti-infective target for drug development as there is no corresponding orthologue of *ppsa* in the human genome.

## 1.7 Aims of this study

The previous sections have outlined the characteristics of the enteric pathogen *H. pylori*. An overview has been presented as how infection is accomplished in the extreme acidic environment of the human stomach and the disease spectrum associated with infection. Therapy regimes followed for eradication have been summarised in addition to a description of the prevalence of antibiotic resistance around the world. The remainder of the chapter addressed the central metabolism of *H. pylori*, highlighting classical experimental findings, interpretation of genome sequencing data and reconstruction metabolic models analysis. These studies deduced the key roles of enzymes in glycolysis, gluconeogenesis and the pentose phosphate pathway. Although these data were useful in verifying some experimental findings, they have also shown the need for further experimental investigations of the components of these pathways.

A number of structural and/or kinetics studies have been conducted to characterize some of the enzymes in glycolysis-gluconeogenesis pathways in *H. pylori*, namely triose phosphate isomerase (Chu et al., 2008), fructose biphosphate aldolase (Fonvielle et al., 2008), glyceraldehyde 3-phosphate dehydrogenase (Elliott, 2009) and fructose-1,6-bisphosphatase (Ayna, 2016). This study will aim to investigate and understand the reactions of *H. pylori* catalysed by the preliminary gluconeogenic enzyme phosphoenolpyruvate synthase (hpPPSA) and the glycolytic-gluconeogenic enzyme phosphoglycerate kinase (hpPGK) that together show possible synthetic lethality in the hypothetical double-deletion mutational studies of *H. pylori*. This means characterisation of these enzymes and verifying theoretical mutational findings in this pathogen.

The work described in this thesis involve two main aims. The first aim of this project was to investigate the essentiality of *ppsa* and *pgk* genes that are hypothesised as conditionally essential metabolic genes in the genetic reconstruction metabolic modelling analysis of *H. pylori* (Thiele et al., 2005). This aim was pursued by the

construction of *ppsA* and *pgk* mutants, through experimental genetic knock out strategies. Should the results show that each of these genes is nonessential for viability, both of them will be knocked out in *H. pylori* to check whether the pathogen will fail to survive or not.

The second aim of the study was to determine the structure and function of hpPPSA and hpPGK. This was pursued by the cloning of the open reading frames of the enzymes in to suitable expression vectors and enzyme over expression. The purified enzymes allowed kinetics and biophysical enzymatic characterisation. Subsequently, X-ray crystallography and/or homology modelling was used to determine or predict the three dimensional structure of the enzymes using the in-house crystallogensis robotics and local or synchrotron X-ray facilities, in addition to structural bioinformatics servers.

Finally, this project will contribute to understanding of glycolysis-gluconeogenesis in *H. pylori*. Should the genetic essentiality investigation, enzymatic characterisation and structural studies indicate that hpPPSA and hpPGK are essential for *H. pylori*, this may indicate possible leads for structure based drug design.

## **Chapter 2. Materials and methods**

### **2.1 Materials**

#### **2.1.1 Chemicals and reagents**

All chemicals were of analytical grade where possible, except where stated and were purchased from Sigma™. Molecular biology kits for extraction and purification of plasmid DNA were obtained from Omega, Bio-Tek. Inc, USA. All buffers and solutions were prepared with distilled water. Restriction enzymes were purchased from NEB (New England Biolabs Ltd, Hitchin, UK).

#### **2.1.2 Chromatographic media and membranes**

Metal chelating resin, Ni-NTA Agarose, was purchased from Qiagen™, Hi prep 16/60 Sephacryl S-200 HR size exclusion chromatography column was bought from GE Healthcare-USA. Viva-spin™ columns for 15 ml, 5 ml and 500 µl sample volumes were purchased from Sartorius and had a molecular weight cut-off of 10 kDa. Dialysis tubing was from Scientific Laboratory Suppliers and was pre-treated by boiling in ddH<sub>2</sub>O prior to use.

#### **2.1.3 Crystallisation accessories**

Sitting drop vapour diffusion MRC plates purchased from Molecular Dimensions were used for crystallisation. Sparse matrix crystallisation screens used within the study were: JCSG-plus™, MORPHEUS, MIDAS, NR-LBD, PACT premier from Molecular Dimensions-ltd; Cryo I and II, AmSO<sub>4</sub> Suite from Qiagen-USA and Wizard™ I and II were purchased from Emerald BioSystems-Washington-USA. Sitting drop protein crystallisation screening was carried out using the robotic system Mosquito® Crystal from ttplabtech. Litho™ loops and “mesh loops” size ranging from 0.05 mm to 0.8 mm were bought from Molecular Dimensions Ltd in order to mount the crystals for data collection either using the in house X-ray crystallography facility or at Diamond Light Source. Cryo pucks, Cryo Puck Tong, puck wands, cryo vials and magnetic caps all were from Hampton Research Ltd.

## 2.2 Bacterial culture conditions and storage

*E. coli* DH5 $\alpha$  was grown on Luria-Bertani media. Luria-Bertani Broth (LB) was prepared with 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl. Luria-Bertani Agar (LA) was prepared by the addition of 1.5% (w/v) agar to LB. LA plates were supplemented with ampicillin, chloramphenicol or kanamycin (Table 2.1) during screening for final mutants constructs. LB or LA plates with *E. coli* DH5 $\alpha$  were normally grown overnight at 37°C under aerobic conditions with shaking or static respectively.

*H. pylori* ATCC® 700392™ (26695) (USA) was grown at 37°C, under microaerobic conditions of 85% N<sub>2</sub>, 10% CO<sub>2</sub> and 5% O<sub>2</sub>. Brain Heart Infusion agar (BHIA) (Oxoid-Basingstoke- UK) was predominantly supplemented with 7% saponin-lysed horse blood, 5  $\mu$ g/ml trimethoprim and 10  $\mu$ g/ml vancomycin (which specifically selected for this pathogenic bacteria as standard antibiotics to avoid contamination). For *H. pylori* mutant selection, chloramphenicol or kanamycin were added to the growth media. *H. pylori* cells were normally sub-cultured by swabbing onto fresh BHIA plates after two to three days of incubation under the microaerobic conditions. After sub-culture, colonies were usually visible and they were ready to be used in experiments. Before use, all antibiotic stock solutions prepared with ddH<sub>2</sub>O or 50% ethanol were filter-sterilized with Millipore filters (0.22  $\mu$ M pore, Gelman Sciences). All these solutions were stored at 4°C in the dark until used.

Glycerol stocks of *H. pylori* (table 2.2) stored at -80 °C were prepared by growing the bacteria on BHIA plates supplemented with trimethoprim and vancomycin and then incubated at 37 °C for 2-3 days under microaerobic conditions. After growth, the cells were scraped from the surface of the plates in 2 mL of Brain Heart Infusion Broth (BHIB) and transferred to 1.5 ml micro-centrifuge tubes. Afterwards, cells were pelleted at 1100 xg in an Eppendorf MiniSpin Plus centrifuge for 5 minutes at room temperature. Pelleted cells re-suspended in 1 ml of BHIB and aliquots of 750  $\mu$ l of the cell suspension was transferred into cryo-tubes, to which equal volumes of 50% (v/v) glycerol was added, snap frozen on dry-ice and stored at -80 °C. *E. coli* DH5 $\alpha$  Glycerol stocks were prepared by growing cells in 5 ml of Luria-Bertani Broth (LB) overnight at 37 °C with shaking at 250 rpm. Cells were centrifuged at 3200 xg (Eppendorf centrifuge 5810R) for

20 minutes at room temperature. Then cells were re-suspended with 1.5 ml of LB and 750 µl of the cell suspension was transferred into cryo-tubes, to which equal volumes of 50% (v/v) glycerol was added, snap frozen on dry-ice and stored at -80 °C.

To subculture glycerol stocks, stored vials were put on dry ice and then cells were plated on the media of choice plates by scraping off cells with stick end of sterilised swab before being spread on to the solid media, and then incubated under the optimum growth conditions with appropriate antibiotics.

Table 2.1 Antibiotics used for strains and/or plasmids selections (purchased from Sigma).

Antibiotic	Solvent	Stock concentration mg/ml	Final concentration µg/ml
<b>ampicillin</b>	ddH <sub>2</sub> O	100	100
<b>chloramphenicol</b>	50% ethanol	20	20
<b>kanamycin</b>	ddH <sub>2</sub> O	50	50
<b>trimethoprim</b>	50% ethanol	5	5
<b>vancomycin</b>	ddH <sub>2</sub> O	10	10

Table 2.2 Bacterial strains used in this study

Bacterial strains	characteristics	Reference
<b><i>E.coli</i> DH5α</b>	Cloning host strain	Invitrogen (Paisley-UK)
<b><i>E.coli</i> Rosetta DE3</b>	hpPPSA and hpPGK protein expression	Novagen
<b><i>H. pylori</i> ATCC® 700392™ (26695)</b>	wild type (wt)	American Type Culture Collection-USA

### 2.3 Chromosomal DNA extraction

The in house protocol (lab 121, department of Genetics and Genome Biology, University of Leicester) was used to prepare the chromosomal DNA of *H. pylori*. The bacteria were grown to confluence on BHIA-5%horse blood plates for 2-3 days, then the growth was scraped off into 2 ml of BHIB. The cell suspension was transferred into a micro-centrifuge tube and centrifuged at 11300 xg for 2 min. Pelleted cells were re-suspended in 1 ml of TE buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA). Cells were pelleted by centrifugation at 11300 xg for 2 min in a micro-centrifuge. The supernatant was carefully discarded and the cells re-suspended again in 400 µl of TE buffer. Possibly,

it is necessary to dilute cells at this point depending on the growth of the bacteria. 10 µl of lysozyme, 10 mg/ml, and 10 µl of RNase, 10mg/ml, was added to the cell suspension which was then incubated at 37 °C for 30 minutes. 75 µl of 10% SDS and 10 µl of proteinase K was added and the suspension after that mixed well and then incubated at 65 °C for 30-60 minutes, or until the solution became clear. The rest of the protocol's steps were conducted by using the Zymo Research Genomic DNA Clean & Concentrator™-25 Kit (Zymo Research Inc. - USA) according to the manufacturer's instructions. The concentration of extracted chromosomal DNA was quantified either by comparison to a DNA standard following agarose gel electrophoresis or by Nano Drop 2000C (Thermo Scientific, UK). The purified DNA was stored at -20 °C for future use.

## 2.4 Plasmid DNA Extraction

For plasmid DNA extraction, *E. coli* DH5α was grown in LB overnight at 37 °C with appropriate antibiotic supplementation. The cell suspension was used to extract plasmid DNA using a plasmid-DNA preparation kit (Omega, Bio-Tek. Inc, USA) according to the manufacturer's instructions. The concentration of extracted plasmid DNA was quantified either by Nano Drop 2000C (Thermo Scientific, UK) or by comparison to a DNA standard following agarose gel electrophoresis. Purified plasmids were stored at -20 °C for future experiments.

## 2.5 Molecular Biology and cloning methods

### 2.5.1 Primer design

The sequences of all open reading frames (ORFs) were taken from genetic databases of *H. pylori* 26695 in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>) (Kanehisa and Goto, 2000). Primers for mutational study including cloning and inverse PCR were designed using Clone Manager 9.0 software (version 9, Scientific and Educational software, 2006) by adding restriction enzyme (RE) sequences (underlined sequences in Table 2.3) to the primers of the flanking regions of the genes encoding phosphoglycerate kinase (hp1345) (*pgk*), phosphoenolpyruvate synthase (hp0121) (*ppsA*) and arginase (hp1399) (*arg*) genes.

For cloning into overexpression vectors, primers were designed using Oligo Calc software (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) (Kibbe, 2007). As ORFs of *pgk* and *ppsa* were constructed in the Protein Expression Laboratory (PROTEX) (Department of Molecular and Cell Biology-University of Leicester), the primers for overexpression vectors were designed in accordance with the guidance provided by PROTEX by adding homology regions (underlined sequences in Table 2.2) to forward (F) and reverse(R) primers. All the designed primers in this study were obtained from Eurofins MWG Operon-Germany.

## **2.5.2 Gene amplification by polymerase chain reaction (PCR)**

The ORFs were amplified by Polymerase Chain Reaction (PCR) from the genomic DNA of the gastric pathogen *H. pylori* 26695 as a template. Formation of PCR products were checked using agarose gel electrophoresis.

### **2.5.2.1 PCR for cloning in pUC19**

Prior to cloning *pgk* and *ppsa* genes (with flanking regions for homologous recombination upstream and downstream each gene), Phusion high-fidelity DNA polymerase kit (New England, Biolabs-UK) was used to amplify these genes in an Eppendorf Mastercycler (Eppendorf, Scientific Support Inc., UK). This procedure included standard reaction steps provided by the manufacturers and cycling conditions of 5 minutes for the activation of polymerase at 98 °C, followed by 29 cycles of: 15 seconds of denaturation at 98 °C, 30 seconds for annealing and extension time of 30 seconds per 1 kb of DNA at 72 °C. The annealing temperature was varied according to the primers used. However, with PCR for testing primer specificity and screening for recombinant clones, Taq DNA polymerase (New England, Biolabs®inc., UK) was used according to the guidance of the manufacturers.

The reaction volumes (20 µl) were used with a final concentration of 1x reaction buffer containing 1.5 mM MgCl<sub>2</sub>; 10 pmol dNTP; 10 pmol primers; DNA template (1-30) ng chromosomal DNA or 1-20 ng plasmid DNA) and autoclaved ddH<sub>2</sub>O was used to standardise volumes to 20 µl.

**Table 2.3** Forward and reverse primers used in this study.

Primer name	Sequence (5'-3')	Restriction enzyme
<b><i>pgk</i>, <i>ppsa</i> and <i>arg</i> cloning primers (restriction sites are underlined)</b>		
hp1345-F	AAGCGT <u>CGACA</u> AAATATCATCATGCGTAGA	<i>SalI</i>
hp1345-R	GCCAGGTAC <u>CTAGTCTT</u> ACCGCATTTAG	<i>KpnI</i>
hp0121-F	CAAAGT <u>CGACT</u> CATGGAACTAATACGAA	<i>SalI</i>
hp0121-R	CCAGGGTAC <u>CGAGCTT</u> AAAATGGTTTAAA	<i>KpnI</i>
hp1399-F	AATCGT <u>CGACA</u> TAAACCCCTAGCATGACCCCTTACCCATA	<i>SalI</i>
hp1399-R	ATCGGGTAC <u>CGGTTT</u> AAATTTTTCAAGTAAAATCACACCAAT A	<i>KpnI</i>
<b>pUC19 primers</b>		
M13-F	GCCAGGGTTTTCCAGTCACGA	-
M13-R	GAGCGGATAACAATTTACACAGG	-
<b><i>ppsa</i> and <i>arg</i> inverse PCR primers (restriction sites are underlined)</b>		
Inv.hp0121-F	TGTGGGAT <u>CCTAGCGTT</u> AGTGAGGGAAGA	<i>BamHI</i>
Inv.hp0121-R	AGCCGGAT <u>CCC</u> ACTTTAATACCAATAGGC	<i>BamHI</i>
Inv.hp1399-F	ATCGGGAT <u>CCCCGAT</u> TAAAAGCGGTTGAGGTAACCGAATA CAACCCACGG	<i>BamHI</i>
Inv.hp1399-R	ATCGGGAT <u>CCC</u> CTTTAATCACATCGCCATGCGTGGCGCTTA AAGCCTCTC	<i>BamHI</i>
<b>Kanamycin primers (restriction sites are underlined)</b>		
Pro-Kan-F	TAGGGGAT <u>CCTATTGACA</u> ATACTGATAAGATAATATATAA	<i>BamHI</i>
Kan-R	TGACGGAT <u>CC</u> ACTAAAACAATTCATCCAGTAAA	<i>BamHI</i>
<b>Chloramphenicol (restriction sites are underlined)</b>		
CAT-F	TCTTGGAT <u>CCCGGAGGATA</u> AATGATGCAATTCACAAAGATT GATATAAAT	<i>BamHI</i>
Pro-CAT-F	TTACGGAT <u>CC</u> CATGGATTGAAAAGTGGATAGATTTATGAT	<i>BamHI</i>
CAT-R	AATTGGAT <u>CC</u> ATTTATTCAGCAAGTCTTGTAATTCATC	<i>BamHI</i>
<b>Clarification of mutant primers</b>		
Extra-hp1345-F	CGAACAGGTTTTGAATGTTGTCTAGGGAGTTGA	-
Extra-hp1345-R	CTAACGCTTCTTGACGACTAATGCTAGCGCTCCT	-
Extra-hp1399-F	CAATCCCTGGCATGTTCCGGCACGCCATA	-
Extra-hp1399-R	CTGAACCCCCAGCTCCAAAGCCACGGATT	-
<b>Overexpression primers (homology regions are underlined)</b>		
HPpgk-F	<u>TACTCCAATCCATGTTAGCT</u> AAAATGTCGTTTATGCAAAT GTAAA	-
HPpgk -R	<u>TATCCACCTTTACTGTCATTA</u> ATGGCGTTTATCCAAAACCTC AAAACA	-
HPppsa -F	<u>TACTCCAATCCATGGT</u> GCGATATATCAAGTTTTTCAA	-
HPppsa-R	<u>TATCCACCTTTACTGTCAAT</u> GCATAGTTAAGCCATGATCTTT	-

### 2.5.2.2 Colony PCR

A colony PCR strategy was performed to verify the presence of a correct recombinant plasmid, or mutated gene with antibiotic resistance cassette integrated in the chromosome of *H. pylori*, directly from transformants isolated on selective solid media. Colonies were picked by toothpick and re-suspended in 200 µl of sterile ddH<sub>2</sub>O in a 500 µl PCR tube. The same toothpick was used to streak these individual clones on the surface of selective solid media and the plate was incubated at 37 °C to provide a stock of each clone. The cell suspension was heated at 98 °C for 5 minutes then cells were centrifuged at 1100 xg for 3 min. The prescribed amount of the supernatant (according to the PCR kit manufacturer instructions) was used as template DNA in a PCR reaction. This reaction was performed using the same conditions described in section 2.5.2.1.

### 2.5.2.3 PCR for cloning in overexpression vectors

As a preliminary step before overexpression of *pgk* and *ppsa*, PCR amplifications by the KOD Hot Start kit (Novagen) were used according to manufacturer's instructions in a final reaction volume of 25 µl. The PCR amplifications were done in a thermal cycler PTC-100 (MJ Research). Different cycling conditions (described in the following subsections) were used if further optimisation was required for obtaining the *ppsa* PCR product.

#### 2.5.2.3.1 *pgk*

A time of 4 minutes was used for the activation of polymerase at 95 °C, after 30 cycles of denaturation for 45 seconds at 95 °C, 1 minute annealing at 55 °C and 90 seconds at 72 °C for extension.

#### 2.5.2.3.2 *ppsa*

2 minutes for the activation of polymerase at 94 °C, after those 30 cycles of denaturation for 15 seconds at 94 °C, 30 seconds annealing at 55 °C and 1 minute at 68 °C for extension.

### 2.5.3 Agarose gel electrophoresis

The plasmid DNA and PCR products were analysed by agarose gel electrophoresis using 1.2% (w/v) agarose gel dissolved in Tris-Acetate-EDTA (TAE), pH 8. DNA was prepared for analysis by mixing 5 µl aliquots of the sample with 1 µl 6x bromophenol loading buffer (contains 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol FF). DNA-agarose gels were stained with 0.5 µg/ml ethidium bromide (EtBr) and visualised under UV transilluminator. The marker, 1 kb DNA Hyperladder 1 (Bioline), was used as a marker to identify the size of the DNA. The sample DNA and markers were loaded in the wells of the gel and electrophoresis was carried out for 30-45 minutes at 100 V.

### 2.5.4 Enzymatic modification of DNA

All DNA digestion protocols were carried out using restriction enzymes supplied from NEB (New England Biolab-UK). Usage was in accordance with the instructions provided by the manufacturers. The protocol involved using 50 µl total reaction mixture containing 5 µl of DNA sample, restriction enzyme (1.25 µl *Bam*HI, 1.25 µl *Eco*RI 2.5 µl *bg*II 2.5 µl of *Sa*II or 5 µl of *Kpn*I), 5 µl buffer, 0.5 µl bovine serum albumin (BSA) and molecular biology grade water (ddH<sub>2</sub>O). The reaction mixture was incubated in a water bath for 3 hours at 37 °C. The restricted DNA was then purified according to the guidelines provided by E.Z.N.A cycle pure kit manufacturers (Omega-biotek-USA).

### 2.5.5 DNA dephosphorylation and ligation

To prevent self-ligation of the vector and to enhance the success of desired ligation the vector was dephosphorylated. The dephosphorylation reaction was catalysed by shrimp alkaline phosphatase (SAP, Roche Diagnostics-Germany). The total volume was 50 µl and consisted of 25 µl vector, 5 µl dephosphorylation buffer, 2.5 µl phosphatase enzyme and the volume completed to 50 µl with ddH<sub>2</sub>O. The reaction mixture incubated in a water bath for 1 hour at 37 °C. After incubation for dephosphorylation, the pUC19 vector was purified by E.Z.N.A cycle pure kit.

Finally, a ligation reaction was carried out using a 3:1 molecular ratio of insert to vector respectively. The reaction mixture consisted of 5 µl of 10X ligation buffer, 1 µl of T4 DNA ligase, DNA of vector and insert and the volume completed to 20 µl with ddH<sub>2</sub>O. The mixture was incubated overnight at 16 °C.

### **2.5.6 Ethanol precipitation**

Ligation products were purified by ethanol precipitation to remove salts and to concentrate DNA. Generally, DNA was precipitated using 1/10<sup>th</sup> volume of 3M sodium acetate, 2.5 volume of 100% cold ethanol and 10 mM of tRNA. After 10 minutes of incubation on ice, the mixture was centrifuged at 11300 xg for 10 minutes to precipitate the DNA. The pellet was re-suspended in 500 µl of 70% ethanol and re-centrifuged for 2 minutes at 11300 xg. The supernatant was decanted and the remaining ethanol was aspirated. The DNA pellet re-suspended with 20 µl of ddH<sub>2</sub>O after aspiration and it was stored at -20 °C for future use.

### **2.5.7 Transformation by electroporation**

#### **2.5.7.1 Preparation of electro-competent cells**

An overnight culture of *E. coli* DH5α (1 ml) was grown on in 100 ml Luria-Bertani broth medium (LB) at 37 °C in a shaking incubator until the Optical Density (OD<sub>600</sub>) of the cell suspension reached 0.4-0.6. The culture was then centrifuged at 3202 xg for 15 minutes at 4 °C to harvest the cells. The cells were washed twice with 100 ml and 50 ml of ice-cold sterile ddH<sub>2</sub>O and re-centrifuged under the same conditions. A further washing step was carried out with 20 ml of ice-cold 10% (v/v) glycerol. Finally, the cell pellet was resuspended in 1 ml of 10% (v/v) glycerol prior to proceeding to transformation (by electroporation) experiments.

#### **2.5.7.2 Electroporation**

Electrical pulses in low ionic strength environment were carried out to improve DNA uptake by electrocompetent cells. This was performed in pre-chilled electro cuvettes using a BioRad GenePulser (BioRad-UK) which was set at 1.5 kV and 1000 Ω. 5 µl of the

purified ligation product (explained in Section 2.5.5) was mixed with 50 µl of *E. coli* DH5α competent cells (described in Section 2.5.7.1) in the electro cuvette. The cells were pulsed once after putting the cuvette in the electroporation machine with a resultant time constant of around 23 millisecond (ms). After electroporation cells were recovered by flushing out in 1 ml of SOC medium (2% (w/v) Bacto-tryptone, 5% (w/v) Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgSO<sub>4</sub> and 20 mM glucose), and then incubated for 1 hour in a water bath at 37 °C. 0.1 ml of the cell suspension was spread out on Luria-Bertani agar (LA) selective medium plates supplemented with 100 µg/ml ampicillin, X-gal (30 µg/ml) and IPTG (40 µg/ml). The plates were incubated overnight at 37 °C for antibiotic resistance selection and white/blue colonies for recombinants.

### **2.5.8 Verification of gene recombination in pUC19**

The presence of recombinant plasmids, with inserted gene, was screened using blue-white screening whereby the transformed colonies with plasmid containing inserts appear in white, as inhibition by insertion disrupt *lacZ* gene function in Puc19, and the colonies which do not carry the recombinant gene will be blue.

### **2.5.9 Plasmid miniprep (miniprep) and PCR**

Several colonies on LA agar selective medium were picked up individually using sterile tips. Each tip was used to inoculate a single colony in 5 ml of LB broth and then streaked on LA agar plates. Both media and agar were supplemented with 100 µg/ml ampicillin and incubated overnight at 37 °C with shaking. Next day the cells were spun down at 3202 g for 20 minutes to pellet the cells for a plasmid miniprep. This was performed using a DNA mini kit (OMEGA, VWR, bio-tek Inc.-USA) according to the instructions of the manufacturers. The purified potential recombinant plasmid was used as template using the PCR conditions explained in Section 2.5.2.1.

### **2.5.10 Sequencing**

The protocol of Protein and Nucleic Acid Chemistry Laboratory (PNAAC-University of Leicester) was employed for sequencing of DNA template using a DNA analyser. The sequencing reaction was carried out using the Big Dye terminator kit 3.1 followed by

PCR amplification in an Eppendorf Mastercycler. The standard reaction mixture with a total volume of 10 µl consisted of 50-125 ng plasmid DNA (purified by miniprep, described in the previous Section) or 2.5-10 ng of PCR product; 20 µM primer and Big Dye 3.1 (1/8 dilution). The cycling conditions are detailed in Table 2.4. The PCR product was cleaned up by adding a 10 µl mixture of 2 µl of 2.2% of SDS and 8 µl of ddH<sub>2</sub>O. The mixture was placed in PCR at 98 °C for 5 minutes and then at 25 °C for 10 minutes. Afterward the sequencing reaction was purified from excess unbound dye by purification through a performa column. Finally, the sequencing reaction was sent to Protein and Nucleic Acid Chemistry Laboratory (PNAAC-University of Leicester) for dideoxy nucleotide sequencing.

### 2.5.11 Cloning and sequencing in overexpression vectors

The genes encoding *pgk* and *ppsa* were cloned into the pLEICS-01 and pLEICS-02 expression vectors respectively by the PROTEX service. Following that, the entire genes sequences were checked for each construct using dideoxy nucleotide sequencing at PNAAC.

Table 2.4 Steps of sequencing PCR reaction

Step	Temperature (°C)	Time
1	95	5 minutes
2	96	10 seconds
3	50	10 seconds
4	60	4 minutes
5	Repeat steps 2-4 for 29 cycles	
6	60	5 minutes
7	End	

### 2.5.12 Natural transformation of *H. pylori* strains

*H. pylori* is among those bacterial species that are naturally transformable. This aspect facilitates mutagenesis in this pathogenic bacteria as they have the ability to uptake foreign double stranded DNA from the environment, and capable of incorporating it into their genome (Gaskin and van Vliet, 2010).

*H. pylori* were inoculated onto Columbia Agar (CA) (Oxoid, Basingstoke-UK) plates (supplemented with 7% saponin-lysed horse blood, 5 µg/ml trimethoprim and 10 µg/ml vancomycin) and grown at 37 °C for 2-3 days under microaerobic conditions of 85% N<sub>2</sub>, 10% CO<sub>2</sub> and 5% O<sub>2</sub>. Using a sterile inoculation loop, the cells were harvested from the plate and transferred as thick patch of cells, approximately 1 cm diameter, on a fresh CA plate and incubated under the same conditions for 5-8 hours. After that, approximately 1-5 µg of constructed plasmid DNA in a maximum of 10 µl sterile ddH<sub>2</sub>O was added and then left to dry for 2-3 minutes. The plate was incubated for 5-20 hours at 37 °C under microaerobic conditions. After growth, the cells were transferred with a cotton swab and spread onto the surface of a fresh CA plate supplemented with selective antibiotic(s) and incubated at 37 °C under microaerobic conditions for 3-7 days. The plates were then checked for successful mutated colonies (Gaskin and van Vliet, 2010).

## 2.6 Protein expression and purification methods

### 2.6.1 Transformation

Over expression trials were carried out by firstly transforming each plasmid into *E. coli* strain Rosetta (DE3). A slightly modified standard transformation protocol (Inoue *et al.*, 1990) was followed. Competent *E. coli* strain Rosetta (DE3) cells were used from -80 °C storage (5-100µL). After that 1µL of plasmid DNA was added into competent cells and gently stirred with tip and left on ice for 30 min, to thaw. Then the cells were incubated in a water bath at 42 °C for 45 seconds. To reduce damage to the cells, the cells were returned to ice for 2 minutes. Following that, 100 µl of 2x Yeast extract Tryptone medium (2xYT) was added and cells were incubated for 1 hour at 37°C, finally, 100 µl of the resulting culture was spreaded on pre-warmed 2xYT agar plates, supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, and grown overnight at 37°C. An overnight culture of the transformant bacteria was prepared, in sterile conditions, by inoculating single colony of *E. coli* strain Rosetta (DE3) in 10 ml of 2xYT medium (contains per litre (L)): 16 gm tryptone, 10 gm yeast extract and 5 gm NaCl) supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, and the culture was incubated overnight at 37 °C.

## 2.6.2 Overexpression optimization

Next day 150  $\mu$ l of overnight culture was diluted in 10 ml of 2xYT medium containing the same concentration of ampicillin and the over expression conditions optimised by variously growing cells at 37 °C until the Optical Density ( $OD_{600}$ ) reached 0.5-0.7, then cells were induced by the addition of 400  $\mu$ M of Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) for 2, 3, 4 hours and overnight respectively. Another culture, induced by 150  $\mu$ M IPTG, was grown overnight under the same conditions except the temperature was dropped to 20 °C until  $OD_{600}$  reached 0.5-0.7. The culture in the last tube was grown overnight at 37 °C without adding IPTG. For further confirmation the same procedure was employed using 50, 100 and 150  $\mu$ M IPTG inductions, and the temperature was dropped from 37 °C to 18 °C and 20 °C for each bacterial culture after the  $OD_{600}$  reached 0.5-0.7, in addition to the overnight treatment without IPTG at 37 °C.

After incubation the cultures were spun down at 3202 g for 10 minutes, then the pellets were lysed in 750  $\mu$ l of the following buffer: 50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM DTT, 1 complete protease inhibitor (Roche-Germany). The cell lysates then sonicated at 10 kHz (10 seconds on, 30 second off, 3 times) on ice. These suspensions were centrifuged at 18407 g for 10 minutes and the insoluble pellets were re-suspended, sonicated and centrifuged under the same conditions. Samples from all the steps were collected for SDS polyacrylamide gel electrophoresis.

## 2.6.3 Small scale purification of hpPGK

To confirm the success of overexpression of his-tagged hpPGK, 750  $\mu$ l soluble fractions of the best over expression trial were added to 200  $\mu$ l slurry Ni-NTA affinity resin (Qiagen), which was washed in a buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 40mM Imidazole and 1 mM DTT. Then the mixtures were centrifuged for 3 minutes at 1500 g, after that they were washed four times with the same buffer. Finally, the tagged protein was eluted with 1.5 ml buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM Imidazole pH 7.5 and 1 mM DTT). Samples were taken from each step to assess the success of overexpression by SDS polyacrylamide gel electrophoresis.

#### 2.6.4 hpPGK expression and purification

The cells were grown at 37 °C in 750 ml of 2xYT medium supplemented with 100 µg/ml ampicillin, and then the temperature was dropped to 20 °C when the OD<sub>600</sub> reached 0.5-0.6. The cells were then induced by the addition of 100-150 µM IPTG (final concentration). Then the cells were harvested, lysed and sonicated as described in Section 2.3.3, with the exception of using larger amount of lysis buffer. The suspension was centrifuged after sonication for 30 minutes at 38700 g at 4 °C, and then the supernatant was added to Ni-NTA resin pre-washed with lysis buffer in a 50 ml Falcon tube. This binding mixture was left on a shaker in cold room for 1 hour for efficient binding, and then spun for 3 minutes at 3200 g. The pellet was washed four times with washing buffer and transferred to a 15 ml tube to be washed once again, and 12 ml of the elution buffer was added to the (His<sub>6</sub> hpPGK)-Ni-NTA beads. This suspension was left for 25 min then spin down for 2 min at 4 °C. TEV protease (1/20 TEV to protein ratio) was added to the supernatant then dialysed overnight with shaking at room temperature using dialysis tube against 1.5 l of dialysis buffer which composed of 50 mM HEPES pH 7.5, 200 mM NaCl and 1 mM DTT. The dialysed supernatant loaded to the same amount of Ni-NTA resin and the same conditions of binding, washing, elution and dialysis were followed. Finally, the flow through (His tag cleaved PGK) and dialysed supernatant (uncleaved PGK) was concentrated with Vivaspin 10 KDa MW cut off concentrator, then flash frozen in liquid Nitrogen and kept in -80 °C prior to use in crystallization and/or characterization. Samples were taken from each step for results evaluation by SDS-PAGE.

#### 2.6.5 SDS-PAGE of proteins

SDS Polyacrylamide Gel Electrophoresis of proteins (SDS-PAGE) is used as essential technique for analysis and monitoring of protein progression in terms of expression, solubility and purification. Sample preparation included mixing 10 µl of protein solution with equal amount of 2X loading dye (0.5 M Tris pH 6.8, 4.4% (w/v) SDS, 20% (v/v) glycerol, 2% (v/v) β-mercaptoethanol and 0.01% bromophenol blue). Samples were boiled at 100 °C for 5 minutes prior to loading on Run Blue pre cast protein gel, 4-20% acrylamide gradient with 17 well cassette gels (Expedeon-USA). The gel electrophoresis

process was conducted at 200 V for 35-40 minutes under reducing conditions using 20X Run blue running buffer (Expedeon-USA), which contains Tris-MOPS-SDS. After electrophoresis, the gels were stained overnight with Run blue Coomassie Brilliant Blue stain (Expedeon-USA). Next day, the gels were destained in ddH<sub>2</sub>O.

### **2.6.6 Peptide mass finger printing of hpPGK**

To verify the identity of the purified enzyme coomassie stained bands of the appropriate molecular weight from SDS-PAGE gel containing both TEV cleaved and uncleaved purified samples of hpPGK (without and with His<sub>6</sub> tag respectively) were sent to the PNAAC lab for peptide mass fingerprinting. The technique compares practical and theoretical trypsin digestion of the purified enzyme given its amino acid sequence from data bases, subsequently mass spectroscopy of each resulting peptides from the purified hpPGK is compared to interpret amino acid sequence. Finally, both mass spectroscopy results are matched to explore the identity of the investigated enzyme.

### **2.6.7 Size exclusion chromatography (SEC)**

A second purification step was applied in which SEC chromatography was performed to improve protein purity and to investigate the oligomeric state of hpPGK. Prior to the experiment, the enzyme was concentrated to 10 mg/ml using a Vivaspin concentrator with 10 kDa molecular weight cut-off. Hi prep 16/60 Sephacryl S-200 HR (GE Healthcare-USA) SEC column (mounted on AKTA purifier) was equilibrated with 2 column volume (CV) of gel filtration buffer (50 mM HEPES pH7.5, 200 mM NaCl and 1 mM DTT). The concentrated enzyme was then injected on to the column and then eluted at a flow rate of 1 ml/min. Fractions were collected in a fraction collector and samples were taken from each fraction for SDS-PAGE analysis. The fractions containing pure enzyme were pooled, concentrated and flash frozen in liquid Nitrogen and then stored at -80 °C for the next experiments. The SEC column was pre calibrated with mixture comprising standard proteins (Figure 2.1).

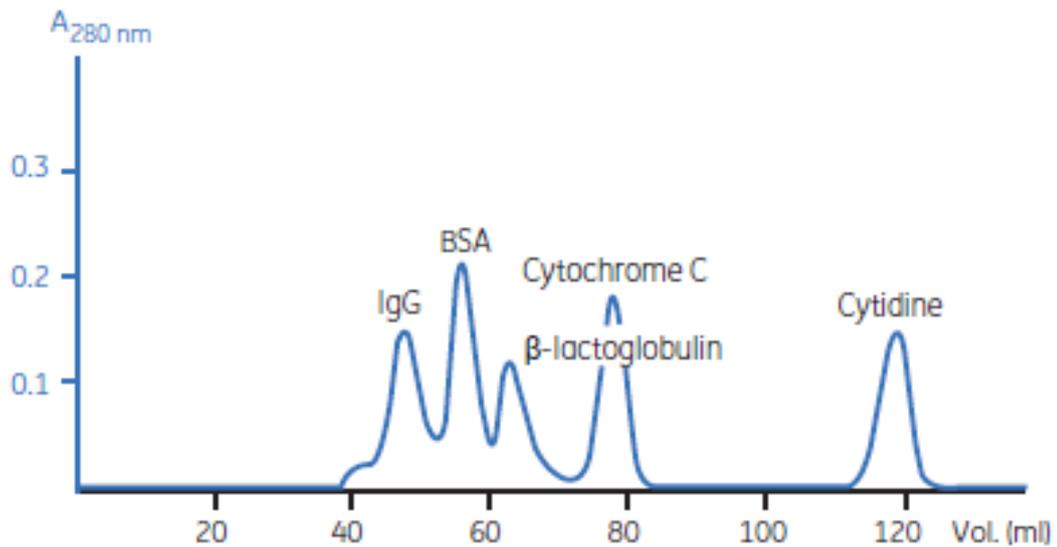
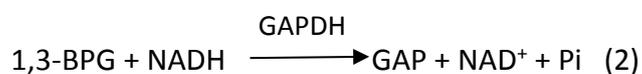
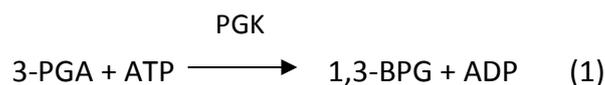


Figure 2.1 Separations of standard proteins by Size Exclusion Chromatography (SEC) using Hi prep 16/60 sephacryl S-200 HR column. The column was loaded with 500  $\mu$ l mixture comprising IgG (Mr 160 kDA), BSA (Mr 67 kDA),  $\beta$ -lactoglobulin (Mr 35 kDA), cytochrome C (Mr 12.4 kDA), and cytidine (Mr 0.24 kDA). The experiment carried out using 0.05 M sodium phosphate buffer pH 7.0 and 0.15 M NaCl, with a flow rate of 0.8 ml/min (24 cm/h). This figure adapted from GE Healthcare (USA) website [https://www.gelifesciences.com/gehcls\\_images/GELS/Related%20Content/Files/1474553092654/litdoc18106088\\_2016101216587.pdf](https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1474553092654/litdoc18106088_2016101216587.pdf).

## 2.7 Enzymatic characterization

### 2.7.1 Principle of hpPGK assay

To test the activity of hpPGK, a coupled enzyme activity assay was employed (with minor modifications) by monitoring the formation of 1,3-bisphosphoglycerate (1,3-BPG) from 3-phosphoglycerate (3-PGA) in the gluconeogenic direction (Krietsch and Bucher, 1970). The reaction steps can be explained in the following chemical equations:



PGK catalyses the ATP dependent phosphorylation of 3-PGA to 1,3-BPG, then the latter is reduced and dephosphorylated to glyceraldehyde,3-phosphate(GAP) by glyceraldehyde,3-phosphate dehydrogenase (GAPDH), simultaneously with the oxidation of NADH to NAD<sup>+</sup> which is measured spectrophotometrically at 340 nm by monitoring absorbance decrease in a time course of 4 min at 1 min intervals.

The assay reaction components in 1 ml total volume are: 40 mM tri-ethanolamine pH 7.5, 2 mM MgSO<sub>4</sub>, 2 mM ATP, 0.15 mM NADH, 2.5 mM 3-PGA and 50 µg/ml GAPDH and 100 nM of PGK. The unit of activity is defined by the amount of enzyme required for conversion of 1 µM of 3-PGA to product in 1 min at 25 °C. The reaction is started by adding PGK to all the reaction components after calibration of the spectrophotometer (PerkinElmer-USA). The data were presented in an average of triplicates as 3 repeat experiments were carried out, and the error bars were estimated by calculating the standard deviation of the results using excel and GraphPad Prism 6.0 respectively.

### **2.7.2 Circular Dichroism (CD) spectroscopy**

This experiment was carried out in a ChiraScan Circular Dichroism (CD) spectrometer (Applied Biophysics-UK) to study the secondary structure and thermal stability of hpPGK. For CD spectra, a final concentration of 0.3 mg/ml of apo N-His<sub>6</sub>hpPGK was prepared in a buffer containing 50 mM Na<sub>2</sub>HPO<sub>4</sub>- NaH<sub>2</sub>PO<sub>4</sub> PH 7.4, 200 mM NaF and 1 mM DTT. The protocol was performed at 20 °C in 0.1 cm light path cuvette with a wavelength range of 185-260 nm in the far Ultra Violet (UV) region with a bandwidth of 1 nm using a quartz cuvette with 10 mm path length. The spectra was de-convoluted by deconvolution software (CDNN 2.1) written by Dr. Gerald Böhm (Böhm, 1997). In addition, the effect of temperature on hpPGK was checked by increasing in 2 °C intervals over the range of 20-90 °C.

### **2.8 Crystallographic methods**

The following sections will deal with the crystallographic methods applied during this study. Several textbooks and reviews provide an introduction to the theory of protein X-ray crystallography. The reader is referred to the book by Bernard Rupp (Rupp, 2009) for a thorough introduction to the theory and mathematics of protein crystallography.

## 2.8.1 Protein crystallisation

To date, X-ray diffraction from high quality crystals represents the most reliable strategy to achieve detailed structural information, it has provided powerful insights into molecular mechanisms and explanations of the cellular functions of biological macromolecules and how they interact to form macromolecular complexes. Furthermore, knowledge of atomic details of the binding sites of complexes allows a rational framework step for searching for new medicines that is drug targets. For this aim, protein crystallography currently used at all levels, including target selection and identification.

Common to all molecules, the crystallisation process comprises three main steps namely nucleation, crystal growth and cessation of growth. In the first step, an adequate amount of molecules are arranged in three dimensions to form thermodynamically stable array of molecules, so called critical nucleus, which provides suitable surfaces for the crystal growth step. The next stage immediately following the nucleation stage, is controlled by the attachment of molecules to the surface of the critical nuclei and their regular assembly onto the growing crystal. Specific, highly directional and organised interactions are required for protein crystal formation. This process is appropriate for the formation of three-dimensional crystal lattice. The crystal growth stage terminates when protein molecules are sufficiently depleted from the solution, or the lattice is destabilised by deformation-induced strain or impurities contaminate the growing crystal faces. Generally, the oligomeric homogeneity in addition to chemical and conformational purity of the sample are crucial factors affecting protein crystallisability.

The whole crystal growth process can be graphically illustrated in a two-dimensional phase diagram (figure 2.2) representing the three stable states liquid, crystalline and precipitate as a function of two variables that is protein concentration and solubility. The protein solution becomes supersaturated when protein concentration exceeds its solubility limits. This zone of the diagram is composed of three regions depending on the supersaturation level; very high supersaturation ("precipitation"), where amorphous aggregates formed, intermediate supersaturation ("labile"), where

nucleation and crystal growth occur and lower supersaturation (“metastable”) where only crystal growth is supported. As these regions are depending on both kinetic and equilibrium parameters, the boundaries between them not well defined.

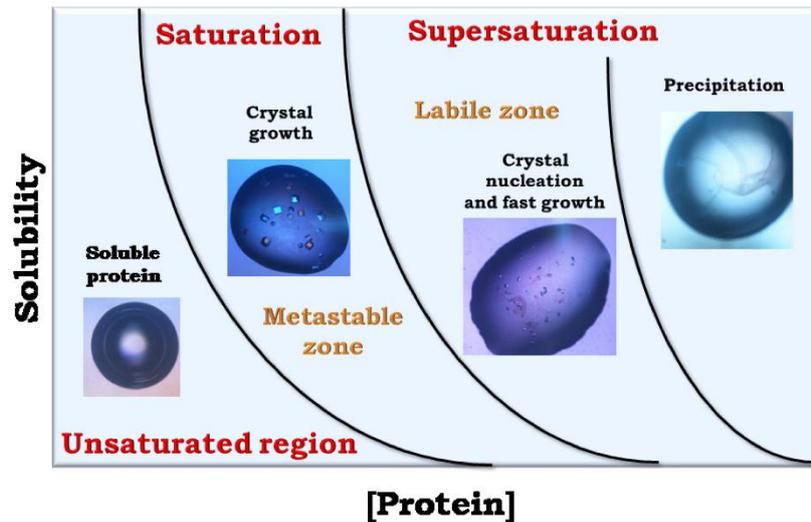


Figure 2.2 Schematic representation of the whole crystallisation process. The diagram shows protein solubility curve, as a function of protein concentration. This figure adapted from (Russo Krauss et al., 2013).

The optimum strategy to be employed is the induction of nucleation at the minimum level of supersaturation just within labile zone. After nuclei formation, protein concentration decreases gradually in the solution, driving crystallisation process toward metastable zone, where crystal growth occurs slowly. Unfortunately, it is very hard to identify these typical conditions. Therefore, the prediction of protein crystallisation conditions from its chemico-physical features is not yet possible.

### 2.8.1.1 Crystallisation trials of hpPGK

Crystallisation trials employed the sitting drop vapour-diffusion technique using Mosquito crystallisation robot (TTP Labtech-UK). This was performed in duplicate MRC 96 well sitting drop crystallisation plates using commercial available screens such as JCSG+, MORPHEUS, MIDAS, NR-LBD, PACT premier (Molecular dimensions-USA), WIZARD (Emerald Bio- Island) and AmSO<sub>4</sub> Suite (Qiagene-USA). The reservoir size in each well was 80 µl, and 100 nl of 5-15 mg/ml of apo hpPGK and liganded hpPGK with equal volume of the reservoir solutions were dispensed by the robot in humidified conditions.

The plates were then covered with transparent tape and left at a constant temperature until crystal formation which was checked regularly by optical microscope.

As always it is the case in crystallography, the first step is to produce high quality crystals. The initial conditions were optimised depending the initial crystal or precipitant formed. The conditions were generally optimized by varying pH, precipitant or protein concentration and equilibration temperature. 10% additive screen (Hampton research-USA), 10% JCSC+ screen, micro seeding and macro seeding were also used to further optimise the conditions. Generally, either the above robotic crystallisation technique or 24 well sitting drop plates were used for this purpose. The reservoir size of the later was 0.3-0.5 ml and appropriate amounts of protein and reservoir solution were mixed manually (for example 1  $\mu$ l + 1  $\mu$ l) and crystal growth was monitored each day by optical microscope.

### **2.8.1.2 Cryoprotectant selection for cryo-cooling of hpPGK crystals**

X-ray data collection was performed mainly in cryo-cooled conditions using a stream of Nitrogen at 100 K. This required the use of cryoprotectants to avoid crystal damage and ice formation. A group of compounds such as polyethylene glycol (PEG) or glycerol were explored as cryoprotectants. The optimal cryoprotectant for hpPGK was determined to be glycerol or PEG400 with a concentration between 20-25%. Incubating hpPGK crystals for 15-30 seconds was sufficient for successful cryoprotection of the crystals prior to cryo-cooling in liquid nitrogen as a preliminary step for data collection.

### **2.8.2 Data collection and data processing**

Once the crystals have been obtained, they are transferred with suitable loops into cryoprotected conditions and rapidly cryo cooled by plunging them into liquid nitrogen. Mounted crystals are then placed in steel pucks (Molecular dimensions-UK). The pucks were placed in a cryogenic storage Dewar, filled with liquid nitrogen, prior sending the crystals to Diamond Light Source (DLS) synchrotron centre (Oxford-UK). At DLS, the pucks were placed in the robotic system specified to mount the crystal within the X-ray beam accurately. The crystals were mounted and centered on a goniostat with the

crystal positioned in a stream of nitrogen gas at 100 K. This was to position and rotate (oscillate) the crystals within the X-ray beam accurately.

After crystal mounting and centering, X-ray diffraction data, as images of reflections, were collected at a range of contiguous angles about a rotation axis perpendicular to the X-ray beam. According to Bragg's law, peaks in the diffraction patterns are formed from constructive interference, rather than destructive, where the scattered X-rays are in phase. The conditions of this interference and its relationship to the crystal lattice were explained by the famous British scientists (Bragg and Bragg, 1913).

After collecting diffraction as two dimensional images, using different oscillation angles, the data need to be processed through three different stages that is indexing, integration and scaling (Rossmann and van Beek, 1999). Firstly, data indexing is required for the initial estimation of crystal symmetry and the cell parameters and orientation. This process can be done using iMOSFLM (Battye et al., 2011). This software uses the positions of spots in one or more diffraction images for the determination of crystal lattice parameters and its orientation.

The next stage is integration, during this stage the proportion of the intensity of each reflection ( $I$ ) to the uncertainties of intensity ( $\delta I$ ) is determined. The accurate sizing and placement of the measurement box is critical for high quality data as the obtained data can be negatively affected by the inclusion of any background noise during integration, the cell parameters and orientation will be continuously refined and updated. After integration, the symmetry in the data is analysed by the program POINTLESS to determine the point-group and the likely space group (Evans, 2006). Finally, a scaling program, AIMLESS, uses the determined symmetry and averages measurements of symmetry-related reflections by putting all the measurements on a common scale. This software provides the first important measures of data quality by producing many statistics (Evans and Murshudov, 2013).

Solvent content in protein crystals ranges from 27% to 65%. This can be calculated using the Matthews co-efficient and probabilities to determine the probable number of molecules per crystallographic asymmetric unit (Matthews, 1968).

### 2.8.3 The phase problem

The phases of the reflections as well as amplitudes values are required to calculate electron density maps from the diffraction data. However, as X-ray is electromagnetic radiation, it is impossible to measure the phase directly by detector. The latter can only measure the intensity which represents the square of the amplitude. Other approaches must be used to overcome this. Depending on the type of the problem, three methods of macromolecular phasing were developed that is Isomorphous Replacement, Anomalous Scattering and Molecular Replacement (MR). In the recent years, MR has become the predominantly used strategy for structure determination as the database of known structures expands. Phaser is the most widely used program for macromolecular phasing of crystal structures by MR. The phasing algorithms implemented in phaser have been developed using maximum likelihood methods, operating within reciprocal space, for the rotation and translation function searches (McCoy et al., 2007).

Generally, MR is used to solve the unknown crystal structure by using another related previously known structure model. The model should be similar to the target protein normally with at least 30% sequence identity. In principle, MR simply involves this model and a set of measured diffraction intensities. All possible orientations and positions of the model in the unknown crystal are tried to find the best matches of the predicted and the observed diffraction data. At this point the model structure is best fitted to the target structure. The calculated phases of the model are used to 'borrow' the phases for the reflections of the unknown crystal as if it were the model that had crystallised in the unknown crystal. An initial map is calculated with the experimentally observed amplitudes and these borrowed phases. Thereafter, the measured amplitudes are used to calculate difference maps that used to supply rebuilding information of the model in order that it more closely resembles the target structure. The MR problem at this point becomes a structure refinement problem (Evans and McCoy, 2008).

Given moderate resolution and high quality data, the Phenix software suite is a highly automated and comprehensive system for macromolecular determination of crystal structures that can quickly provide an initial model of protein structure without

significant manual human intervention. Auto MR Wizard is used to solve crystal structures by molecular replacement. This program provides a convenient interface to phaser MR and feeds the output results directly to the AutoBuild Wizard for automated model rebuilding (Adams et al., 2011).

#### 2.8.4 Refinement

The search model rotation and translation onto the position of the target protein model within the asymmetric unit generates MR output files that is a coordinate file (.pdb). Furthermore, the structure factor amplitudes (F) and calculated phases are written out from MR in the mtz file format. All of these parameters are combined via Fourier transform to calculate initial electron density maps of the molecule. The map quality will be influenced by errors during data collection and the data quality. Interactive map building programs, for structure refinement, are used to enhance the agreement between the atomic model and the observed data (map fitting). This process can be repeated until the best representation of the observed data is achieved in an optimum model.

Generally, on the basis of initial models the agreement is poor between the observations ( $F_o$ ) and the calculated structure factor amplitudes ( $F_c$ ) which leads to model bias. The latter can be reduced during phasing by calculating  $2F_o - F_c$  Fourier map. Moreover,  $F_o - F_c$  map is also calculated by subtracting the calculated amplitudes from the observed structure factor amplitude. This produces three potential electron densities depending on the subtraction result. Either Zero if the model is perfect, positive with the presence of some parts in the structure but missing in the model or negative where parts are absent in the structure but present in the model. These maps can be viewed and compared to the atomic model in a molecular-graphics software such as Coot (Emsley et al., 2010).

Within phenix, phenix.refine is the software which optimise the atomic model with respect to the observed reflections. An automatic run of refinement in phenix.refine predominantly comprises three main stages. Initially reading in and processing of the data, selecting the optimum refinement protocols, refinement and finally writing out a

refined model, entire statistics of refinement and electron density maps in a variety of formats (Adams et al., 2011). Water molecules are updated (added, removed and refined) during the refinement process using an automated protocol for updating the ordered solvent model (Afonine et al., 2012). After refinement during model building, the model parameters including all atoms coordinates and temperature factors in a structure are adjusted to achieve a closer agreement between the observed and calculated structure factors. The reliability factor (R factor) is an indicator for judging the success of refinement process which allows further monitoring to the subsequent changes in the model. Nonetheless, 5% of the data before refinement are randomly selected and used as a measure to ensure that the model is not over-refined and is used for the generation of  $R_{\text{free}}$ . Calculating  $R_{\text{free}}$  is similar to R factor except random selection of the observed reflections is used (Brünger, 1992).

### 2.8.5 Structure validation

Three simple but important questions regarding reported information completeness, analysis quality and structure correctness are addressed with structure validation (Spek, 2009). As computational methods were used during refinement and rebuilding processes, similar strategies can be followed to check the potential problematic pinpoint parts of the model prior to deposition and publication (Kleywegt and Jones, 1998). R and  $R_{\text{free}}$  factors are used to indicate how well the model fits within the electron density. In addition to this stereochemical checks of the model are required, this can be carried out using Root Mean Square Deviation (RMSD) from ideal values determined from very accurate small structures and Ramchandran plots of backbone dihedral angles. Commonly used data validation programs are PROCHECK (Laskowski et al., 1993), MolProbity (Davis et al., 2004) and PDB\_REDO (Joosten et al., 2014).

## 2.9 Bioinformatics

### 2.9.1 Primary sequence alignment

Multiple sequence alignment provides analysis tool to investigate conserved and non-conserved regions among structurally known proteins and the target protein. Amino

acid sequence were aligned to other sequences of known structure orthologues in the Protein Data Bank (PDB) using the Basic Alignment Search Tool (BLAST) web-server (Altschul et al., 1990). This program allows rapid detection of homologues proteins using the amino acid sequence database and the PDB (Berman et al., 2000). Prior to homology modelling, hpPPSA amino acid sequence was taken from the genetic data bases for *H. pylori* 26695 in KEGG (section 2.5.1) and submitted to protein BLAST web-server([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)) for primary sequence alignment of hpPPSA with other orthologues in the PDB.

### 2.9.2 Homology modelling of hpPPSA

In the structural genomics project, homology modelling, also termed as template-based modelling or comparative modelling, plays a central role in the prediction of protein structures. After given the target protein amino acid sequence, homology modelling process consists of four steps that is identification of the known structure homologue from the Protein Data Bank (PDB), secondary structure alignment of the query sequence with the template structure, model building based on the alignment result and model assessment and refinement (Xiang, 2006).

For hpPPSA structure modelling, the sequence was obtained from the genetic data bases server KEGG (Kanehisa and Goto, 2000). The protein sequence was submitted in the Phyre2 web portal for protein modelling (<http://www.sbg.bio.ic.ac.uk/phyre2>), prediction and analysis. This web available suite uses remote homology detection methods to build three dimensional protein models (Kelley et al., 2015). The automatic server modelled hpPPSA using *Neisseria meningitides* PPSA (nmPPSA) structure as the template obtained from the PDB (2OLS) which has 53% sequence identity with hpPPSA.

## Chapter 3. Investigation of *ppsA* and *pgk* synthetic lethality in *H. pylori* 26695

### 3.1 Introduction

Targeted mutagenesis of *H. pylori* requires the use of shuttle mutagenesis. This can be conducted by inactivation of *H. pylori* DNA cloned fragments in *E. coli* prior to be introduced into *H. pylori* by natural transformation for allelic replacement. Shuttle mutagenesis is a widely used technique in the genetic studies of *H. pylori*. After cloning in *E. coli*, this technique starts with inactivating part or all of the gene, by inserting a selectable genetic marker, in a replication vector like pUC19 which is unable to self-replicate and therefore is a suicidal plasmid for *H. pylori* (Labigne and Jenks, 2001).

Two main strategies were employed in this study, insertion-inactivation (removing very small base pairs of the gene prior to inserting antibiotic resistance cassette within the gene) and deletion-inactivation (removing significant part of the gene coding sequence followed by inserting antibiotic resistance cassette). The strategies were pursued to mutate *pgk*, *ppsA* and *rocF* in *E. coli* DH5 $\alpha$  prior to the introduction of these genes in the wild type *H. pylori* 26695 as a final stage of mutagenesis process. Insertion of an antibiotic resistance cassette into a gene may cause polar (distal) effects on the expression of genes downstream the mutated gene, especially if they are located in an operon involving several genes and transcribed from a single promoter. In this study, terminatorless antibiotic resistance cassettes were used so that no or only minor polar effects were exerted on the downstream genes (Fischer et al., 2001).

The aim of this study was to investigate the computational findings of hp0121 (*ppsA*) and hp1345 (*pgk*) synthetic lethality as they identified as conditionally essential genes in the *in silico* double deletion mutational studies of *H. pylori* 26695 (Thiele et al., 2005). This was achieved by constructing isogenic mutants of *ppsA* and *pgk* in *H. pylori* 26695. In addition, constructing isogenic mutants of the non-essential gene hp1399 (*rocF*) (Langford et al., 2006) was required as a controls. This gene encodes arginase, an enzyme that hydrolyses L-arginine to L- ornithine and urea (Mendz and Hazell, 1996).

Two objectives were pursued as described in this chapter. The steps of the first objective were the cloning of *ppsA*, *pgk* and *rocF* genes with flanking regions (figure 3.1) into pUC19 plasmid, mutating each gene by introducing an antibiotic resistance cassette into the gene and recombination of the mutated gene into the chromosome of *H. pylori* 26695. The second objective was testing synthetic lethality by creating double mutant of *ppsA* and *pgk* in *H. pylori* 26695.

Figure 3.1 illustrates the arrangement of *pgk*, *ppsA* and *rocF* in the genome of *H. pylori* 26695. *pgk* is found in a putative operon containing genes involved in metabolism, transport and membrane integration. However, *ppsA* and *rocF* are predicted not to be involved in an operon (Caspi et al., 2012).

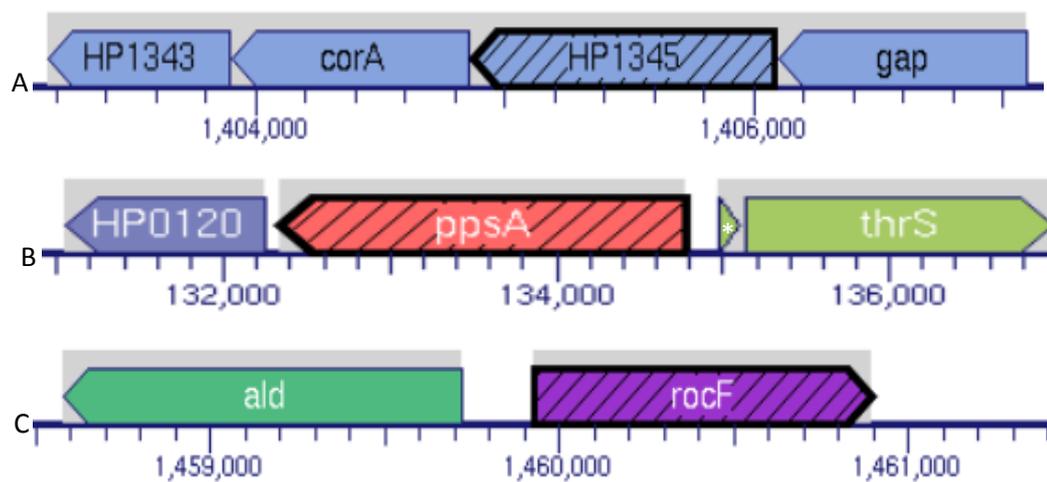


Figure 3.1 Organisation of hp1345 (*pgk*), hp0121 (*ppsA*) and hp1399 (*rocF*) in the genome of *H. pylori* 26695. (A) HP1343: conserved hypothetical integral membrane protein, *corA*: magnesium and cobalt transport protein coding gene (hp1344), HP1345: phosphoglycerate kinase coding gene (hp1345) (*pgk*), *gap*: glyceraldehyde-3-phosphate dehydrogenase coding gene (hp1346) (*gap*); (B) HP0120: hypothetical protein, *ppsA*: phosphoenolpyruvate synthase coding gene (hp0121) (*ppsA*), white asterisk: *H. pylori* predicted coding region (hp0122), *thrS*: threonyl-tRNA synthetase coding gene (hp0123) (*thrS*); (C) *ald*: alanine dehydrogenase coding gene (hp1398) (*ald*), *rocF*: arginase coding gene (HP1399) (*rocF*). This figure was adapted from *H. pylori* genomic context server. Gene colouring indicates predicted operon membership (Caspi et al., 2012).

## 3.2 *pgk*, *ppsA* and *rocF* cloning into pUC19

### 3.2.1 Cloning *pgk* into pUC19

The sequence of *pgk* gene was obtained from the online database KEGG (<http://www.genome.jp/kegg/>). The gene was amplified using hp1345-F and hp1345-R primer pairs containing *Sal*I and *Kpn*I restriction sites respectively (Table 2.2). The primers were designed to anneal from approximately 530 bp upstream and 360 bp downstream of the gene to facilitate crossing over between the mutant and wild type alleles in future applications. The PCR was performed as described in Section 2.5.2.1 using chromosomal DNA of *H. pylori* 26695 as a template. The success of the amplification was assessed via agarose gel electrophoresis and an expected 2140 bp DNA band was obtained (Figure 3.2).

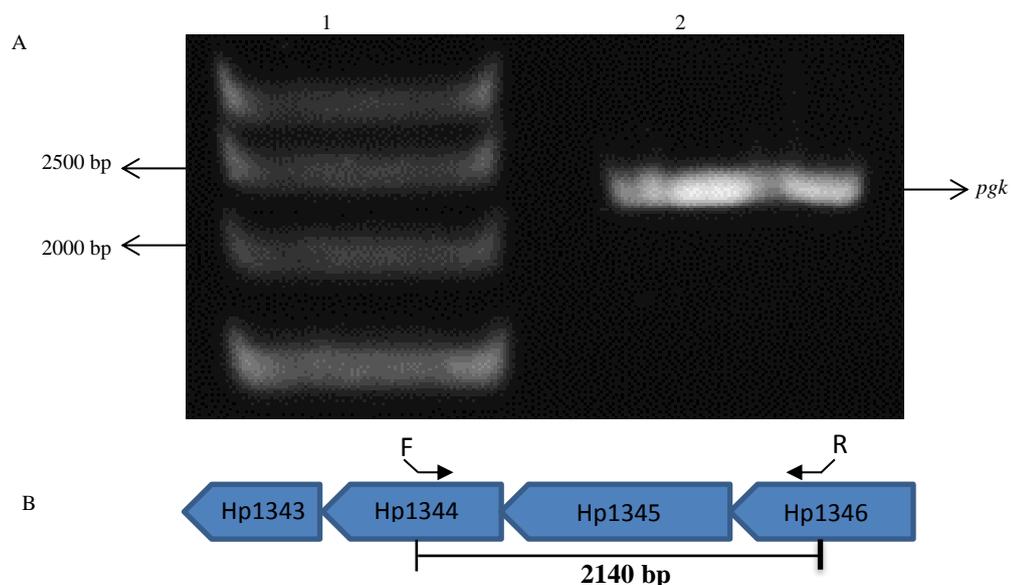


Figure 3.2 PCR amplification of *pgk* from *H. pylori* 26695. Panel A: Agarose gel electrophoresis of approximately 2140 bp PCR product of *pgk* (around 1200 bp) with flanking regions of approximately 530 and 360 nucleotides upstream and downstream the gene respectively with *Sal*I and *Kpn*I restriction sites. Lanes contain: lane 1. DNA marker (Hyper ladder (1kb)), lane 2. *pgk* PCR product (2140) after using hp1345-F and hp1345-R forward and reverse primers respectively and *H. pylori* 26695 chromosomal DNA as template. Panel B: diagrammatic representation of PCR amplification of *pgk*. Solid arrows show primer annealing positions, open arrows indicate coding sequence orientation.

The purified PCR product and pUC19 were subjected to sequential digestion by *SalI* and *KpnI* as described in section 2.5.4 and the products were evaluated by agarose gel electrophoresis (Figure 3.3). The *pgk* amplicon (insert) and the restricted dephosphorylated pUC19 were ligated and then purified by ethanol precipitation. The purified ligation reaction was transformed in *E. coli* DH5 $\alpha$  as described in Section 2.5.7. The putative recombinant transformed cells (white colonies formed due to interruption of *lacZ* gene in pUC19) were picked and inoculated in LB broths supplemented with ampicillin and grown overnight at 37 °C. The cells of these cultures were harvested for plasmid miniprep purification as described in Section 2.5.9. The successful insertion of *pgk* in pUC19 was verified by colony PCR using hp1345-F and hp1345-R primers (Figure 3.4). The presence of amplicon within the construct was further verified by sequencing using pUC19 specific primers (M13F and M13R) (data not shown). The resultant plasmid was designated as Puc19::*pgk*.

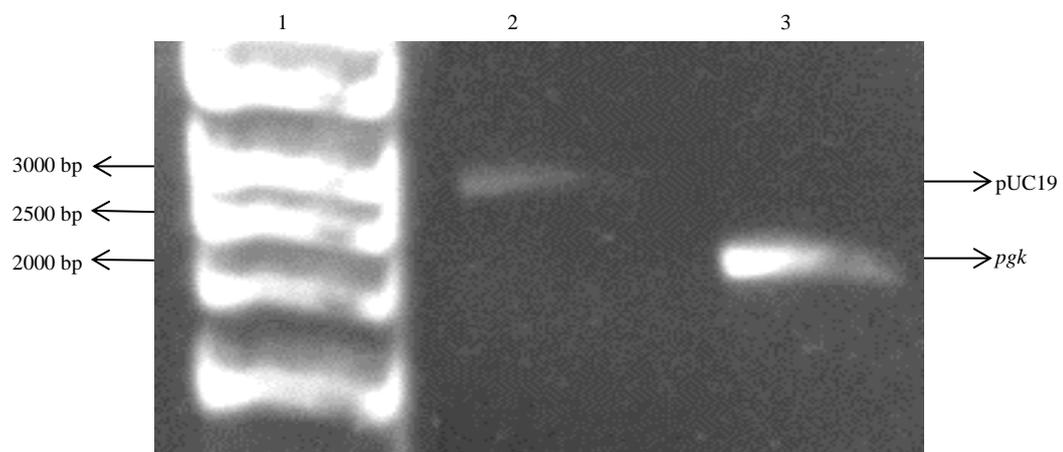


Figure 3.3 Agarose gel electrophoresis of pUC19 and *pgk* restriction with *SalI* and *KpnI* restriction enzymes. Lanes contain: lane 1. DNA marker (Hyper ladder (1kb)), lane 2. Dephosphorylated pUC19 cut with *SalI* and *KpnI* restriction enzymes (2665 bp), lane 2. *pgk* insert (2140 bp) cut with *SalI* and *KpnI* restriction enzymes.

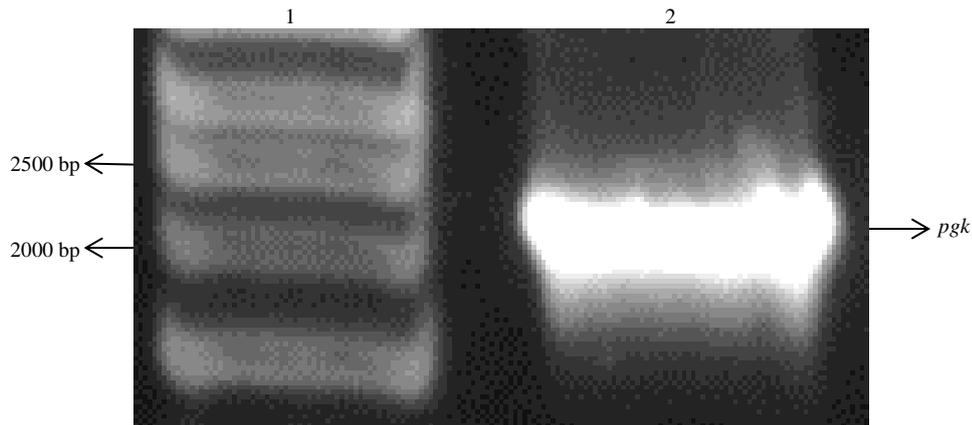


Figure 3.4 Agarose gel electrophoresis of *pgk* PCR product using recombinant pUC19 and *pgk* as template. Lanes contain: lane 1. DNA marker (Hyper ladder (1kb)), lanes 2. *Pgk* (2140 bp) amplified from recombinant *pgk* and Puc19 (as template) purified by miniprep from different selected *E.coli* DH5 $\alpha$  colonies.

### 3.2.2 Cloning *ppsA* in pUC19

Cloning and identification of *ppsA* was accomplished using the same strategy employed in the cloning of *pgk*. The primers hp0121-F and hp0121-R incorporating *SalI* and *KpnI* restriction sites respectively (Table 2.2) were amplified from chromosomal DNA of *H. pylori* 26695. These primers were designed to anneal at approximately 40 bp and 880 bp flanks on each side of 5' and 3' ends of *ppsA* respectively. The 5' flank was short to avoid repeating sequences in the upstream region which may affect the incorporation of the mutated gene to the chromosome in the next stages of mutagenesis. However, an additional 540 bp of *ppsA* was used as homology region (in total the size with the 5' flank is 580 bp) that was required for the integration of the mutated allele. The incorporation of these flanks is important for successful recombination of the mutated allele within *H. pylori* 26695 chromosome in the final stages of mutagenesis. An approximately 3380 bp DNA band in agarose gel electrophoresis verified PCR success (Figure 3.5). The plasmid (pUC19) and PCR product (*ppsA*) were restricted by *SalI* and *KpnI* restriction enzymes as described in section 2.5.4 (Figure 3.6).

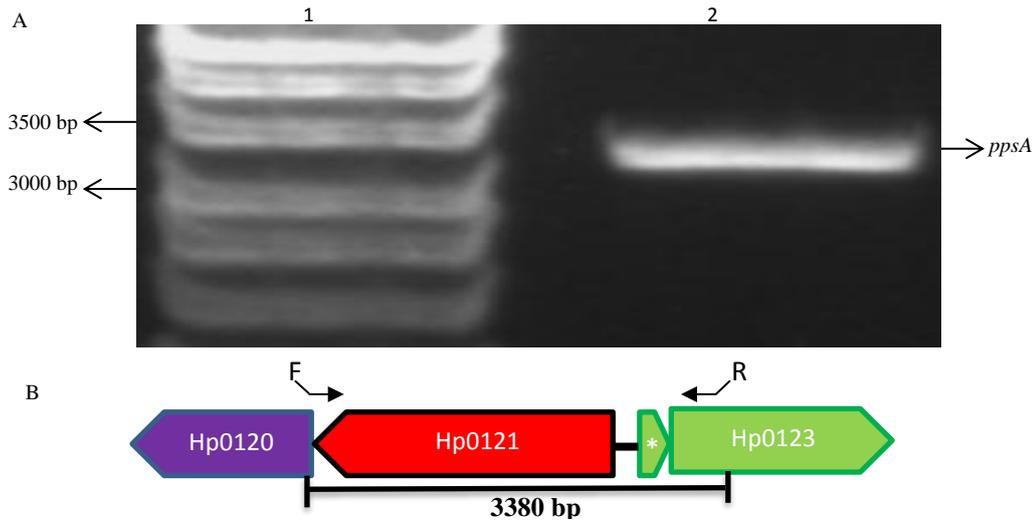


Figure 3.5 PCR amplification of *ppsA* with 40 and 880 nucleotides flanking regions on 5' and 3' ends of the gene with *SalI* and *KpnI* restriction sites respectively. Panel A: agarose gel electrophoresis of PCR amplification of *ppsA*, lanes contain: lane 1(Hyper ladder (1kb)), lane 2. *ppsA* PCR product (3380 bp) after using hp0121-F and hp0121-R forward and reverse primers and *H. pylori* 26695 chromosomal DNA as template. Panel B: diagram representation of PCR amplification of *ppsA*. Solid arrows show primer position for annealing, open arrows indicate coding sequence orientation.

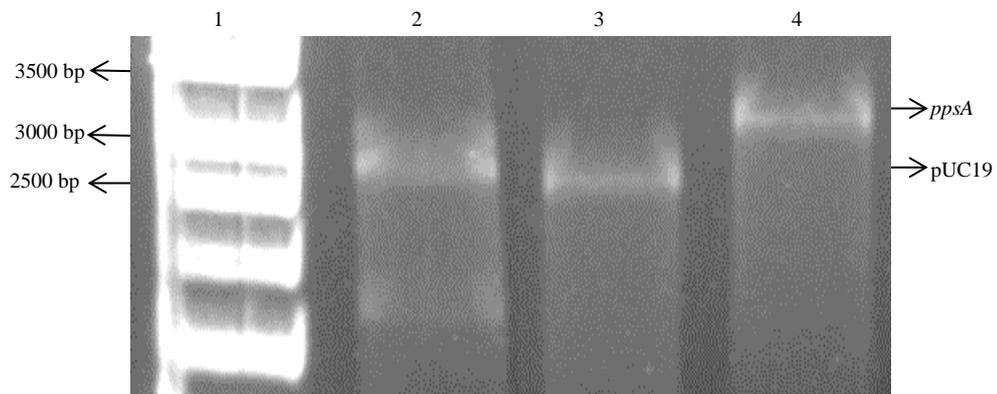


Figure 3.6 Agarose gel electrophoresis of pUC19 and *ppsA* restriction by *SalI* and *KpnI* restriction enzymes. Lanes contain: lane 1. DNA hyper ladder (1kb), lane 2. Undigested pUC19 plasmid control, lane 3. Dephosphorylated pUC19 (2665 bp) restricted with *SalI* and *KpnI* restriction enzymes, lane 4. *ppsA* insert with 40 and 880 nucleotides flanking regions on 5' and 3' ends of the gene respectively (3380 bp) digested with *SalI* and *KpnI* restriction enzymes.

After that, the vector and the insert were ligated and then transformed into *E. coli* DH5 $\alpha$  (Section 2.1.3.1). White/blue screening of the transformants colonies was done in a similar manner to *pgk* cloning in pUC19 to identify and choose successful

recombinant white colonies (containing pUC19-*lacZ* disrupted with *ppsA* construct). After carrying out a plasmid miniprep, these recombinant clones with insertion of *ppsA* fragment into pUC19 were identified by colony PCR using hp0121-F and hp0121-R primers. The constructed plasmid was used as a template to verify the presence of the amplicon with an expected size of 3380 bp (figure 3.7). A further verification of *ppsA* presence in pUC19 was then followed by Big-Dye reaction sequencing using the pUC19 primers, namely M13-F and M13-R (data not shown). Accordingly, the construct was designated as Puc19::*ppsA*.

### 3.2.3 Cloning *rocF* in pUC19

*rocF* in pUC19 was constructed in a similar way to *pgk* and *ppsA* by cloning in pUC19 and using the construct as control in the next stages of *pgk* and *ppsA* mutagenesis in *H. pylori* 26695. Hp-1399-F and hp-1399-R primers containing *SalI* and *KpnI* restriction sites respectively (table 2.2) were used to amplify *rocF* from genomic DNA of *H. pylori* 26695 with the addition of flanking regions of 700 bp upstream and 500 bp downstream of the gene. The PCR reaction produced an approximately 2190 bp DNA fragment including the coding sequence of *rocF* flanked by the additional sequences and the restriction sites at both ends (figure 3.8).

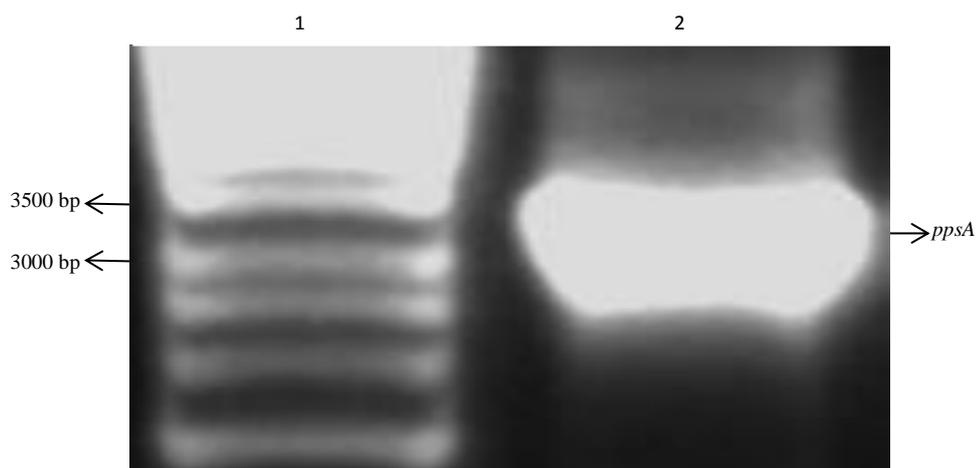


Figure 3.7 Agarose gel electrophoresis of *ppsA* colony PCR product using recombinant pUC19 and *ppsA* as template. Lanes contain: lane 1. DNA marker (Hyper ladder (1kb)), lane 2. *ppsA* (3380 bp) amplified from recombinant *ppsA* and Puc19 (as template) purified by miniprep from different selected *E.coli* DH5 $\alpha$  colonies.

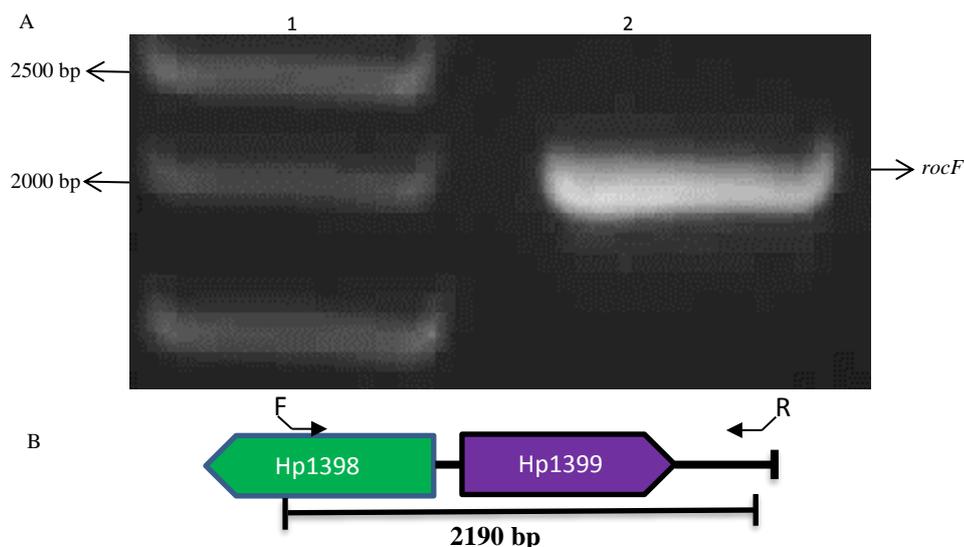


Figure 3.8 PCR amplification of *rocF* with 700 and 500 nucleotides flanking regions on 5' and 3' ends of the gene with *SalI* and *KpnI* restriction sites respectively. Panel A: Agarose gel electrophoresis of PCR amplification of *rocF*, lanes contain: lane 1(Hyper ladder (1kb)), lane 2. *rocF* PCR product (2190 bp) after using hp1399-F and hp1399-R primers respectively and *H. pylori* 26695 chromosomal DNA as template. Panel B: diagram representation of PCR amplification of *rocF*. Solid arrows show primer position of primer annealing, open arrows indicate coding sequence orientation.

In order to construct recombinant *rocF* in pUC19, both vector (pUC19) and insert (amplified *rocF*) were digested with *SalI* and *KpnI* restriction enzymes as described in section 2.5.4 (figure 3.9). The success of each stage was monitored via DNA agarose gel electrophoresis to verify the sizes of the expected fragments. The restricted *rocF* PCR product and pUC19 were ligated and transformed into *E. coli* DH5 $\alpha$  by electroporation. Similarly to *pgk* and *ppsA* cloning in pUC19, blue/white colony screening of the transformants were conducted to identify cells that contained recombinant plasmids.

Cells containing *rocF* inserted in pUC19 (*lacZ* gene disrupted white colonies) were selected and screened using colony PCR (hp1399-F and hp1399-R primers were used) after plasmid miniprep purification (figure 3.10). Further verification of the construct was carried out by sequencing using pUC19 specific primers M13-F and M13-R (data not shown). After verification, the construct was designated as pUC19::*rocF*.

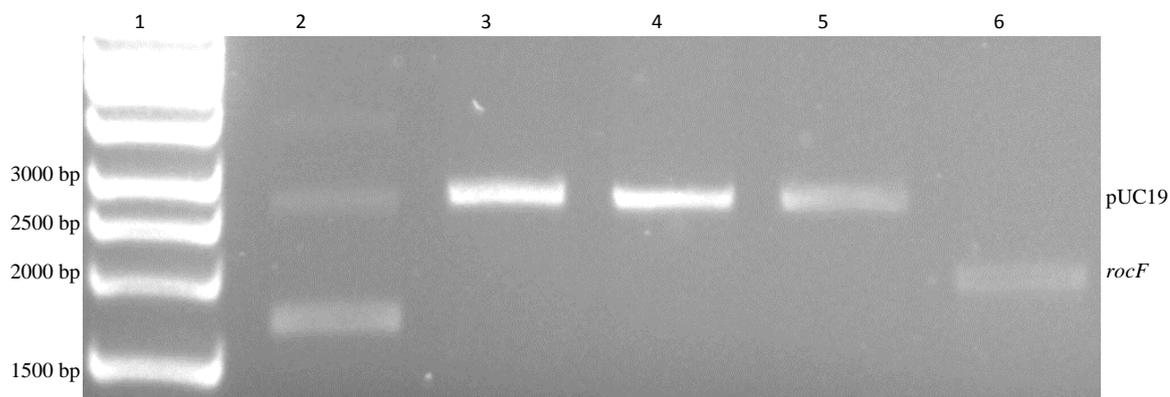


Figure 3.9 Agarose gel electrophoresis of pUC19 and *rocF* restriction by *SalI* and *KpnI* restriction enzymes. Lanes contain: lane 1. DNA hyper ladder (1kb), lane 2. pUC19 plasmid –ve control, lane 3. pUC19 +ve control restricted with *SalI*, lane 4. pUC19 +ve control restricted with *KpnI*, lane 5. Dephosphorylated pUC19 restricted with *SalI* and *KpnI* restriction enzymes, lane 6. *rocF* insert with 700 and 500 nucleotides flanking regions on 5' and 3' ends of the gene respectively (2190 bp) restricted with *SalI* and *KpnI* restriction enzymes.

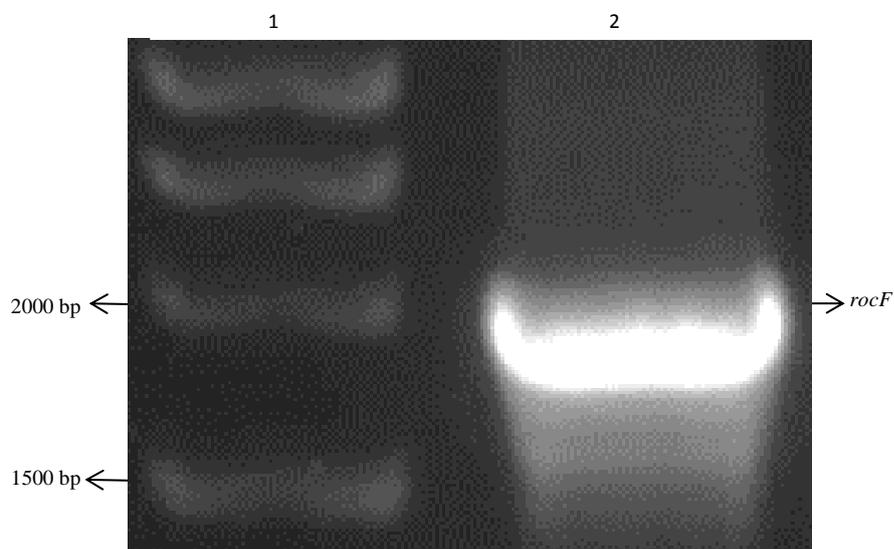


Figure 3.10 Agarose gel electrophoresis of *rocF* colony PCR product using recombinant pUC19 and *rocF* as template. Lanes contain: lane 1. DNA marker (Hyper ladder (1kb)), lane 2. *rocF* (2190 bp) amplified from recombinant *rocF* in Puc19 (as template) purified by carrying out a miniprep from different selected *E. coli* DH5 $\alpha$  transformant colonies.

At the end of this part of the study, three constructs were produced from cloning *pgk*, *ppsA* and *rocF* into pUC19 that is Puc19::*ppsA*, Puc19::*pgk* and pUC19::*rocF* respectively. These plasmids were used as templates for constructing mutated alleles within pUC19 in the next step of the study namely mutagenesis of *pgk*, *ppsA* and *rocF* in *E. coli* DH5 $\alpha$ .

### 3.3 Mutagenesis of *pgk*, *ppsA* and *rocF* in *E. coli* DH5 $\alpha$

#### 3.3.1 Insertion- inactivation of *pgk*

Insertional-inactivation mutagenesis was used to mutate *pgk*. A chloramphenicol acetyl-transferase (*cat*) gene with cognate promoter (734 bp) cassette had already been generated and used to mutate *pgk* in this study. The cassette was digested with *Bam*HI restriction enzyme. Concurrently, recombinant *pgk* in pUC19 plasmid (section 3.2.1) was digested by *Bgl*III restriction enzyme as naturally occurring site in the middle region of *pgk* gene is available (figure 3.11).

After that, the *Bam*HI restriction site tagged cassette was ligated into the *Bgl*III compatible cohesive ends of *pgk* gene into pUC19 plasmid to produce *pgk* mutated plasmid. The ligation reactions were transformed into *E. coli* DH5 $\alpha$  cells by electroporation (section 2.5.7). Positive transformant colonies were selected by plating on Luria-Bertani (LB) agar (section 2.2) supplemented with chloramphenicol (20  $\mu$ g/ml) and ampicillin (100  $\mu$ g/ml) (table 2.1). The existence and orientation of chloramphenicol resistance cassette in *pgk*-pUC19 construct was verified by colony PCR using a combination of *cat* specific primers (pro-CAT-F and CAT-R) and *pgk* cloning primers (hp1345-F and hp1345-R) (table 2.2). The orientation of the cassette was verified by using the primers pair of hp1345-F and CAT-R. As calculated with Clone Manager 9.0, the expected size of the PCR fragment was approximately 1840 bp (figure 3.12).

After identification of the cassette orientation, the construct was used as template in sequencing (section 2.5.10) using hp1345-F, hp1345-R and CAT-R to verify the insertion of *cat* within *pgk* and any potential errors within the sequence. This construct was designated as pUC19-*pgk*::*cat* (figure 3.13).

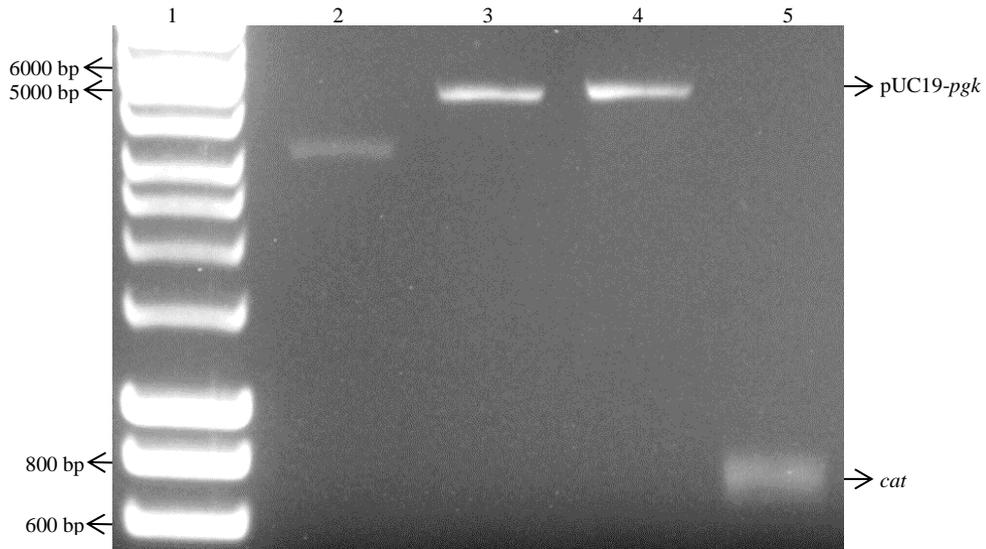


Figure 3.11 Recombinant *pgk* in pUC19 and *cat* restriction digestion by *Bgl*III and *Bam*HI respectively. Lanes contain: lane 1. Hyper-ladder DNA marker (1 kb), lane 2. pUC19-*pgk* plasmid uncut (-ve control), lanes 3 and 4. *Pgc*-pUC19 plasmid (approximately 4800 bp) restricted from the naturally occurring *Bgl*III site inside *pgk*, lane 5. *Bam*HI tagged promoter *cat* (734 bp) after cutting both ends of the gene with the restriction enzyme.

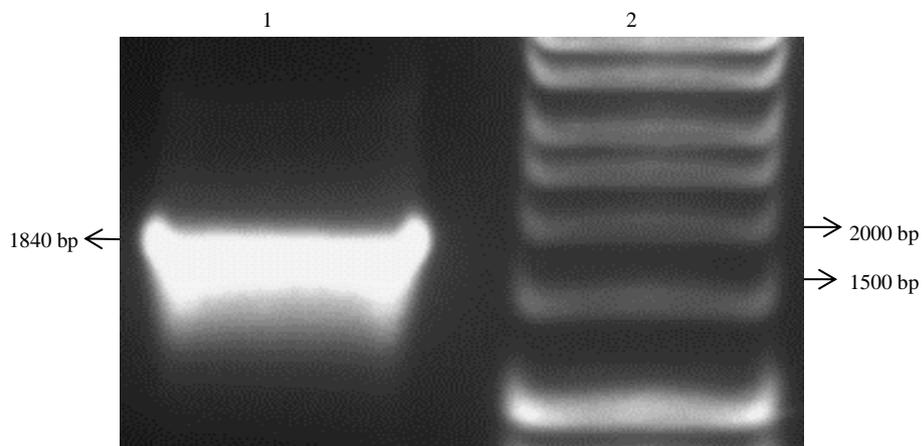


Figure 3.12 mutagenesis of *pgk* by Insertional-inactivation strategy: verification of the insertion and orientation of promoter *cat* in *pUC19-pgk* was carried out by colony PCR. Lane 1. PCR product amplified from pUC19-*pgk*::*cat* plasmid using hp1345-F and CAT-R primer pair ( predicted size 1840 bp), lane 2. Hyper-ladder DNA molecular weight marker (1kb).

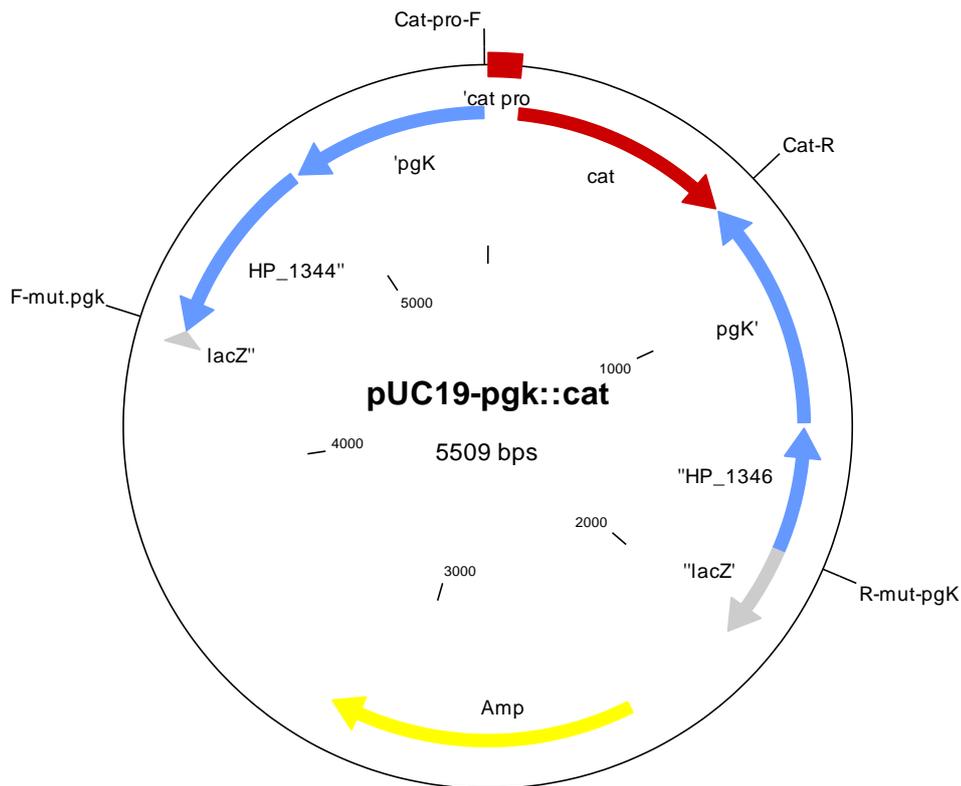


Figure 3.13 Plasmid map of pUC19-*pgk::cat*, *pgk* gene was insertionally inactivated by promoter chloramphenicol resistance cassette. The location of PCR primers used for verifying the insertion and the orientation of the cassette F-mut-*pgk* (hp1345-F) and R-mut-*pgk* (hp1345-R) are indicated in addition to *cat* primers that is Cat-pro-F and Cat-R primers pair.

### 3.3.2 Deletion-inactivation of *ppsA*

After successful production of a pUC19-*ppsA* construct, a deletion-inactivation strategy was carried out by inverse PCR mutagenesis of the construct to enable deletion of a significant proportion of the coding sequence for *ppsA* (that is approximately 1800 bp). To create this mutant, inverse PCR primer pair hp0121-invF and hp0121-invR with a 5' *Bam*HI restriction site were used (table 2.2). The pUC19-*ppsA* construct was used as a template for these primers in an inverse PCR amplification to produce a linear inverse PCR product with a size of 4290 bp (figure 3.14).

To mutate *ppsA*, a promoter kanamycin resistance cassette (*kan*) was amplified by PCR using pro-kan-F and kan-R (table 2.2) with 5' *Bam* HI restriction site ends. The *ppsA* deleted inverse PCR product and the cassette were digested with *Bam* HI (figure 3.15),

ligated and transformed into *E. coli* DH5 $\alpha$  electrocompetent cells by electroporation. LB agar plate supplemented with ampicillin (100  $\mu\text{g}/\text{ml}$ ) and kanamycin (50  $\mu\text{g}/\text{ml}$ ) was used for selecting *E. coli* colonies that contained pUC19 $\Delta ppsA::kan$  plasmid.

The presence and the orientation of the cassette within the construct was verified by colony PCR using a combination of *ppsA* cloning primers hp0121-F and hp0121-R and kanamycin primers pair pro-kan-F and kan-R. The PCR product resulting from using hp0121-F and pro-kan-F identified the orientation of the kanamycin resistance cassette within the construct. As predicted and calculated in Clone manager 9.0, the size of the PCR fragment was 1460 bp (figure 3.16).

For further conformation, the construct was sequenced using *ppsA* cloning primers (hp0121-F and hp0121-R) and kan-F primer to verify the existence of kanamycin within mutated *ppsA* and any potential errors within the sequence of the flanking regions upstream and downstream *ppsA*. This construct containing kanamycin resistance cassette inserted within deleted mutated *ppsA* was designated as pUC19 $\Delta ppsA::kan$  plasmid (figure 3. 17).

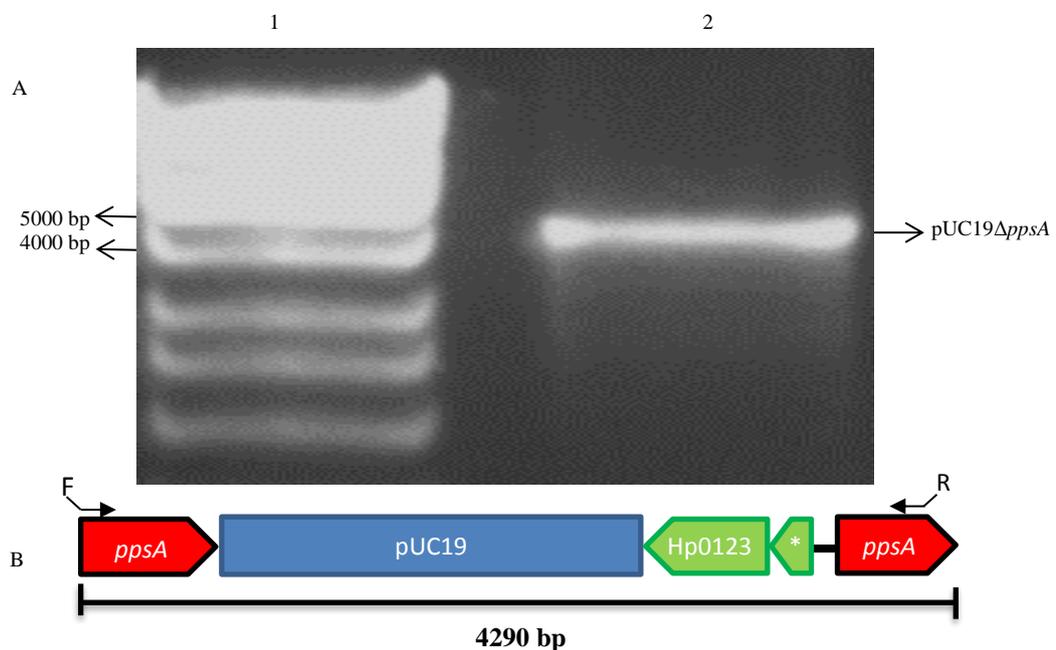


Figure 3.14 Mutagenesis of *ppsA* by inverse PCR using pUC19-*ppsA* construct as template. (Panel A) Lanes contain: lane 1. Hyper ladder DNA marker (1 kb), lane 2. Inverse PCR product from pUC19-*ppsA* plasmid (pUC19 $\Delta ppsA$ ) using hp0121-invF and Inv.hp0121-invR to produce 4290 bp amplicon. (Panel B) Schematic representation of the inverse PCR product explained in lane 2. White asterisk: hp0122.

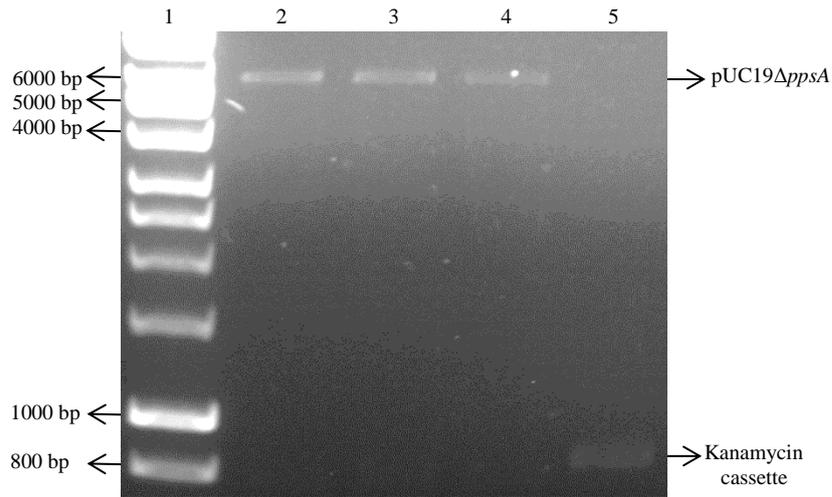


Figure 3.15 Deletion-inactivation of *ppsA* by inverse PCR mutagenesis. This figure illustrates the restriction digestion of *ppsA*-Puc19 plasmid inverse PCR product that is pUC19  $\Delta$ *ppsA*(4290) and promoterized kanamycin resistance cassette (*kan*). Lanes contain; lane 1. DNA hyper ladder marker (1 kb), lane 2-4. Inverse PCR products of *ppsA*-Puc19 construct (using Inv.hp0121-F and Inv.hp0121-R) restricted with *Bam*HI, lane 5. *Bam*HI restriction digestion of promoterized kanamycin resistance cassette (880 bp).

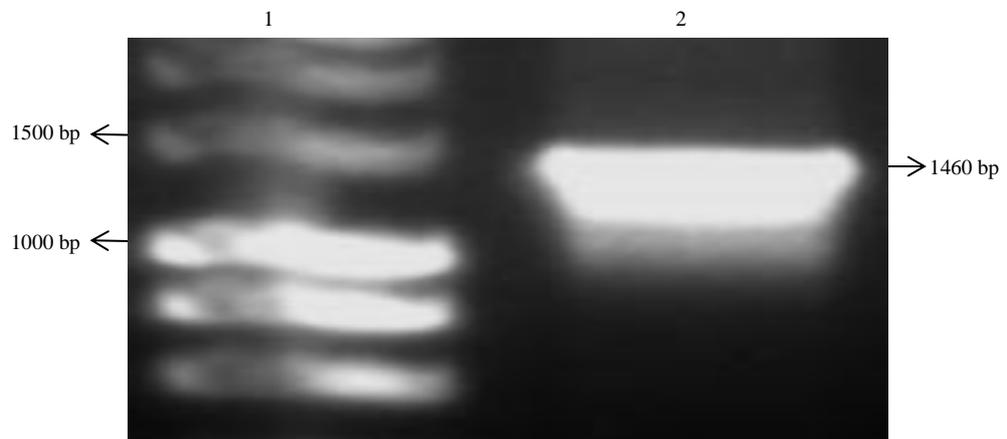


Figure 3.16 colony PCR screening of pUC19 $\Delta$ *ppsA*::*kan* plasmid for verification of kanamycin resistance cassette insertion and orientation in pUC19-*ppsA* plasmid constructed by cloning. Lanes contain: lane 1. 1kb DNA marker (hyper ladder), lane 2. Colony PCR product amplified from pUC19 $\Delta$ *ppsA*::*kan* construct as template using F-mut-*ppsA* (hp0121-F) and F-pro-*kan* (kan-F).

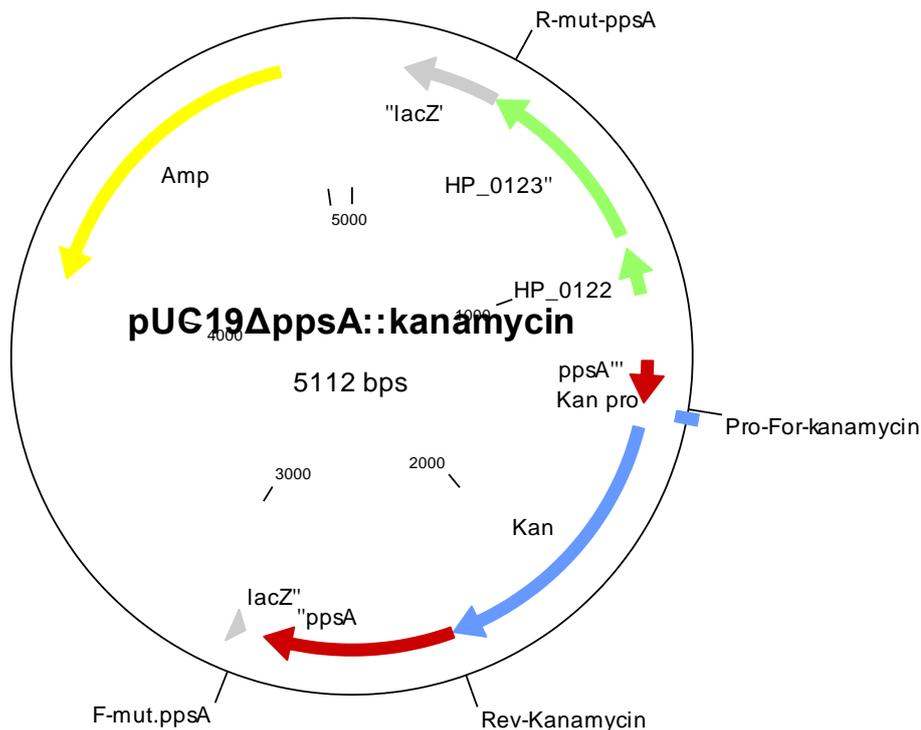


Figure 3.17 Plasmid map of pUC19Δ*ppsA*::*kan*, deletion-inactivation of *ppsA* was carried out by inverse PCR strategy and inserting promoter kanamycin resistance cassette. This figure indicate *ppsA* cloning primer pair F-mut.*ppsA* (HP0121-F) and R-mut.*ppsA* (hp0121-R) in addition to kanamycin resistance cassette primers pro-for-kanamycin (pro.kan-F) rev-kanamycin (kan-R) was used for identification the insertion and orientation of the cassette within the plasmid.

### 3.3.3 Deletion-inactivation of *rocF*

As in *ppsA* mutagenesis, deletion-inactivation strategy was followed for *rocF* mutagenesis. To achieve this goal, *rocF* was mutated by inverse PCR to enable the deletion of approximately 700 bp of this gene coding sequence. Amplification of pUC19-*rocF* template (section 3.2.3) was conducted using *rocF* inverse PCR primers (Inv.hp 1399-F and Inv.hp1399-R containing 5' *Bam*HI restriction sites to create the mutant (table 2.2)) to produce an approximately 4100 bp fragment. This PCR product consisted of pUC19 sequence and the sequences flanking *rocF* in addition to the remaining sequence of this coding sequence after deletion. The resulted DNA fragment size after inverse PCR was approximately 4150 bp (figure 3.18).

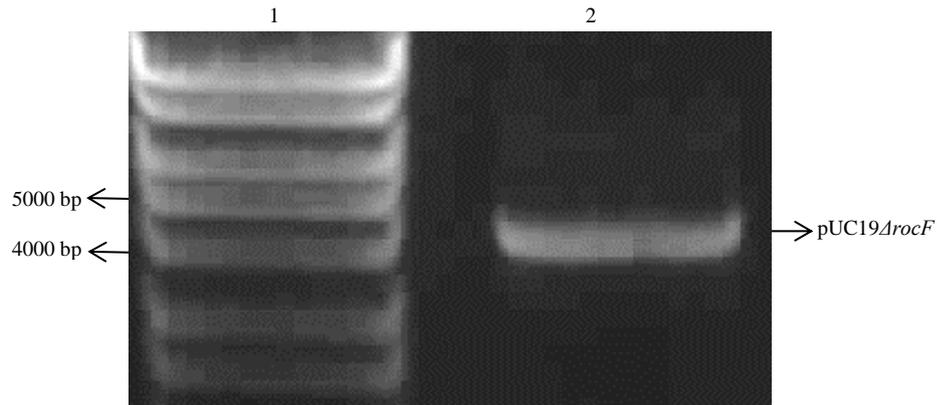


Figure 3.18 Inverse PCR deletion-inactivation mutagenesis of *rocF* using pUC19-*rocF* construct as template. Lanes contain: lane 1. Hyper ladder DNA marker (1 kb), lane 2. Inverse PCR product from pUC19-*rocF* plasmid using Inv.hp1399-F and Inv.hp1399-R to produce approximately 4150 bp PCR product.

As *rocF* was required as a control with pUC19-*pgk*::*cat* and pUC19 $\Delta$ *ppsA*::*kan* constructs during the introduction of these mutated alleles into *H. pylori* 26695, the following sections will describe the insertion of chloramphenicol and kanamycin resistance cassettes within pUC19 $\Delta$ *rocF* inverse PCR product.

### 3.3.3.1 *cat* insertion within pUC19-*rocF* inverse PCR product

Prior to the introduction of the mutated allele into *H. pylori* 26695, a promoter-less chloramphenicol resistance cassette tagged with 5' *Bam*HI restriction sites was used to create a positive control *rocF* mutated construct. A pUC19 $\Delta$ *rocF* inverse PCR product and *cat* (630 bp fragment) were digested with *Bam*HI (figure 3.19) and then ligated to form a 4800 bp plasmid. The resultant ligation (consists of *cat* insert and pUC19 $\Delta$ *rocF* backbone) was introduced into *E. coli* DH5 $\alpha$  by electroporation. After overnight growth at 37 °C, recombinant colonies containing plasmids were selected by plating on the surface of LA media supplemented with ampicillin (100  $\mu$ g/ml) and chloramphenicol (20  $\mu$ g/ml). The presence and the orientation of the insert ( $\Delta$ *rocF*::*cat*) was established by colony PCR using *rocF* cloning primers (hp1399-F and hp-1399-R) and promoter-less *cat* primers pair (CAT-F and CAT-R). The PCR product resulted from using hp1399-F and CAT-R identified the cassette orientation and the expected size estimated by clone manager 9.0 (approximately 1475 bp) (figure 3.20).

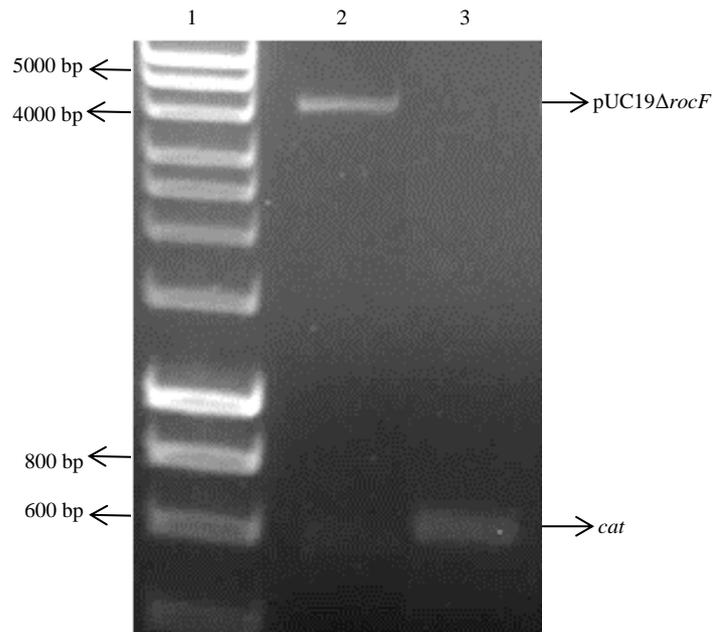


Figure 3.19 Inverse PCR mutagenesis (Deletion-inactivation strategy) of *rocF*. The figure shows *Bam*HI restriction digestion of inverse PCR product resulted from *rocF*-Puc19 plasmid and promoter-less chloramphenicol resistance cassette. Lanes contain; lane 1. DNA hyper ladder marker (1 kb), lane 2. Inverse PCR product of *rocF*-Puc19 construct (4800 bp) (using Inv.hp0121-F and Inv.hp0121-R) restricted with *Bam*HI, lane 3. *Bam*HI restriction digestion of promoter-less chloramphenicol resistance cassette (655 bp).

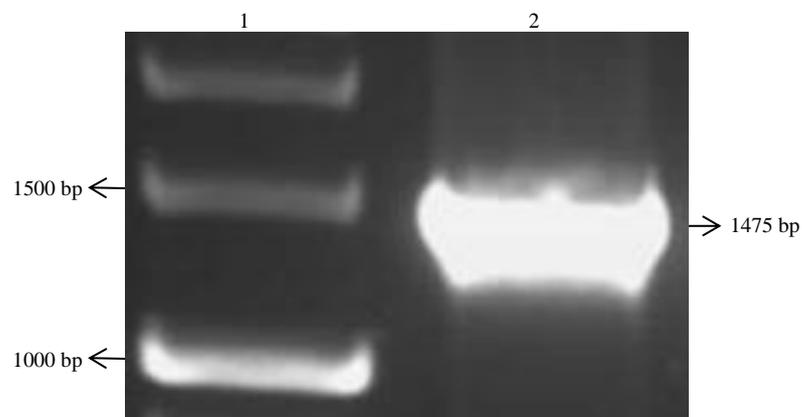


Figure 3.20 mutagenesis of *rocF* by Deletion-inactivation strategy: identification of the insertion and orientation of promoter-less *cat* in *pUC19ΔrocF* inverse PCR product was carried out by colony PCR. Lane 1. Hyper-ladder DNA molecular weight marker (1kb), lane 2. PCR product amplified from *pUC19ΔrocF::cat* plasmid using hp1399-F and CAT-R primer pair (predicted size 1475 bp)

Finally, the insert of a plasmid recombinant clone was sequenced using hp1399-F, hp1399-R, CAT-F and CAT-R primers. Sequencing results verified that there are no sequence errors in the upstream and downstream flanking regions. Furthermore, the existence of chloramphenicol resistance cassette within the mutated *rocF* allele was verified. The construct containing mutated *rocF* and promoter-less *cat* was designated as pUC19 $\Delta$ *rocF*::*cat* (figure 3.21).

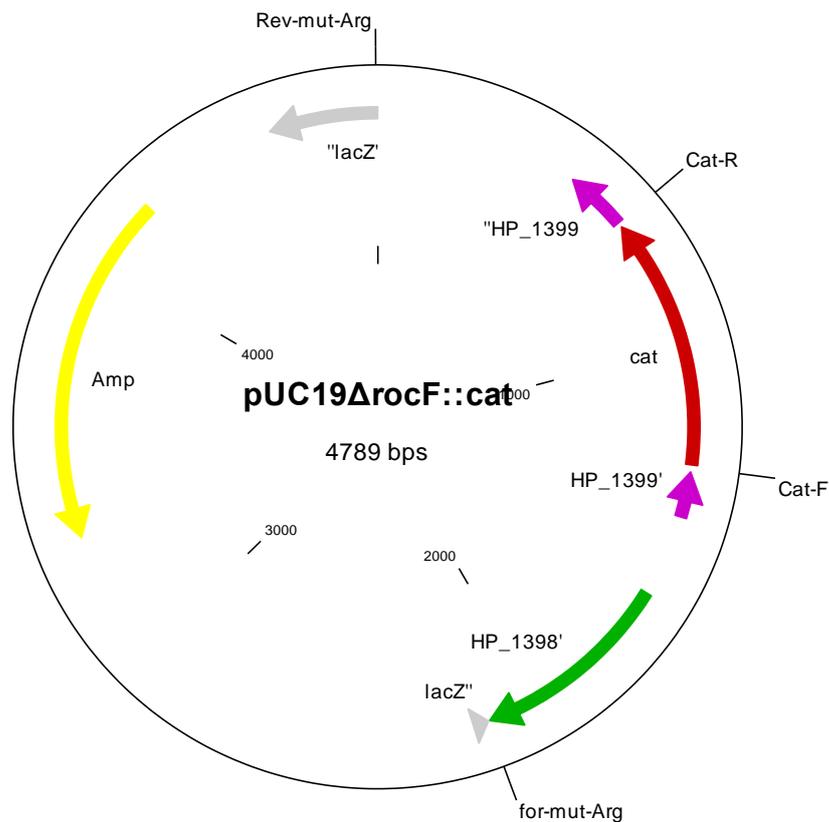


Figure 3.21 Plasmid map of pUC19 $\Delta$ *rocF*::*cat*, deletion-inactivation of *rocF* was carried out by inverse PCR strategy and inserting promoter-less chloramphenicol resistance cassette. This figure indicate primers that used for identification the insertion and orientation of the cassette within the plasmid namely *rocF* cloning primer pair for-mut-Arg (hp1399-F) and rev-mut-Arg (hp1399-R) in addition to *cat* primers cat-F (CAT-F) and cat-R (CAT-R).

### 3.3.3.2 Kanamycin insertion within pUC19-*rocF* inverse PCR product

As carried out with the insertion of *cat* cassette within pUC19 $\Delta$ *rocF*, a promoter-containing kanamycin resistance cassette ends with 5' *Bam*HI restriction sites was used to create another positive control *rocF* mutated construct. The inverse PCR product and kanamycin (880 bp fragment) were similarly digested with *Bam*HI (figure 3.22) and then constructed by ligation to form a 5014 bp plasmid. The plasmid (consists of kanamycin cassette inserted in pUC19 $\Delta$ *rocF*) was introduced into *E. coli* DH5 $\alpha$  by electroporation. Positive colonies (containing recombinant plasmids) were selected by plating on the surface of LA media supplemented with ampicillin (100  $\mu$ g/ml) and kanamycin (50  $\mu$ g/ml) after overnight growth at 37 °C. Colony PCR strategy was followed to identify the presence and the orientation of the insert ( $\Delta$ *rocF*::*kan*) by using *rocF* cloning primers (hp1399-F and hp-1399-R) and promoter kanamycin primers (pro.kan-F and kan-R). The cassette existence and orientation was identified by colony PCR using hp1399-F and kan-R primers. This resulted in an expected 1700 bp PCR product according to clone manager 9.0 calculations (figure 3.23).

The insertion of kanamycin within pUC19 $\Delta$ *rocF* was verified by sequencing using hp1399-F, hp1399-R, pro.kan-F and kan-R primers. Sequencing results showed that there are no sequence errors in the flanking regions of *rocF*. Furthermore, the existence of the kanamycin cassette within the mutated *rocF* allele was verified. The construct containing mutated *rocF* and promoter kanamycin cassette was designated as pUC19 $\Delta$ *rocF*::*kan* (figure 3.24).

Finally, pUC19-*pgk*::*cat* and pUC19 $\Delta$ *ppsA*::*kan* constructed plasmids (in addition to pUC19 $\Delta$ *rocF*::*cat* and pUC19 $\Delta$ *rocF*::*kan* control plasmids respectively) were prepared for the next step of this study, that is the recombination of the mutated *pgk* and *ppsA* alleles (in addition to the mutated *rocF* control alleles) within the genome of *H. pylori* 26695.

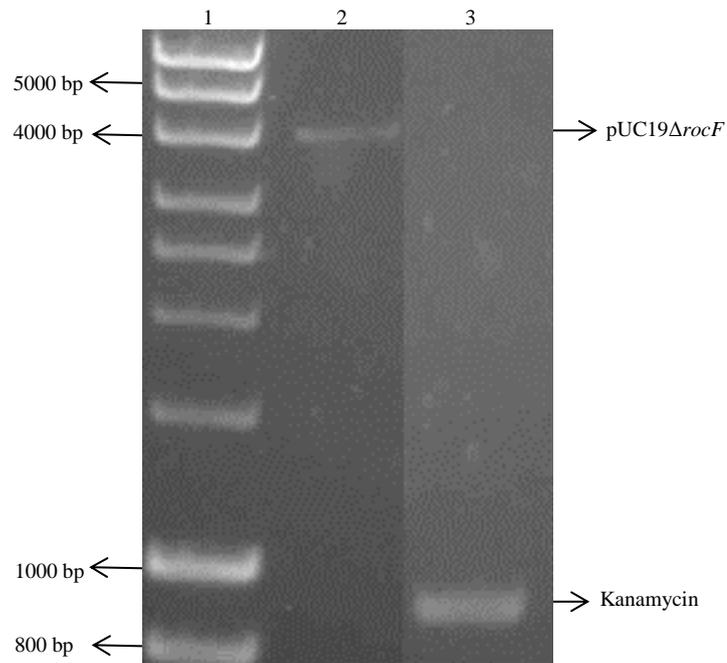


Figure 3.22 Inverse PCR mutagenesis (Deletion-inactivation strategy) of *rocF*. The figure shows *Bam*HI restriction digestion of inverse PCR product resulted from *rocF*-Puc19 plasmid and promoted kanamycin cassette. Lanes contain; lane 1. DNA hyper ladder marker (1 kb), lane 2. Inverse PCR product of *rocF*-Puc19 construct (4800 bp) (using Inv.hp0121-F and Inv.hp0121-R) restricted with *Bam*HI, lane 3. *Bam*HI restriction digestion of promoted kanamycin cassette (880 bp). As there were some unrequired bands, the photo was cut and the required bands were combined.

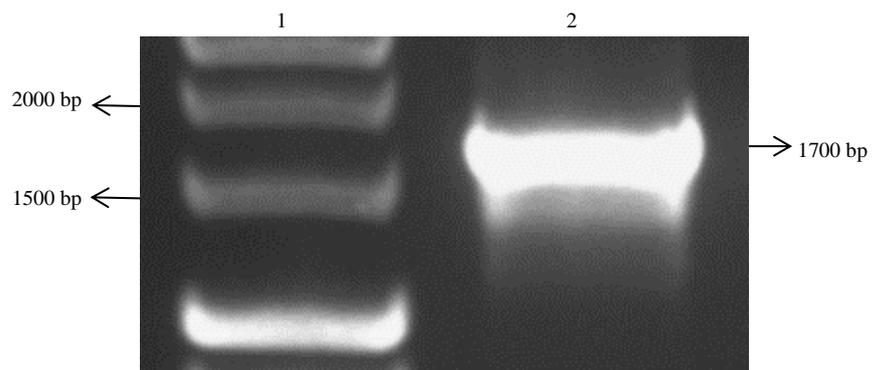


Figure 3.23 Mutagenesis of *rocF* by Deletion-inactivation strategy: identification of the insertion and orientation of promoted kanamycin in *pUC19ΔrocF* inverse PCR product was carried out by colony PCR. Lane 1. Hyper-ladder DNA marker (1kb), lane 2. PCR product amplified from *pUC19ΔrocF::kan* plasmid using hp1399-F and kan-R primer pair (predicted size 1700 bp).

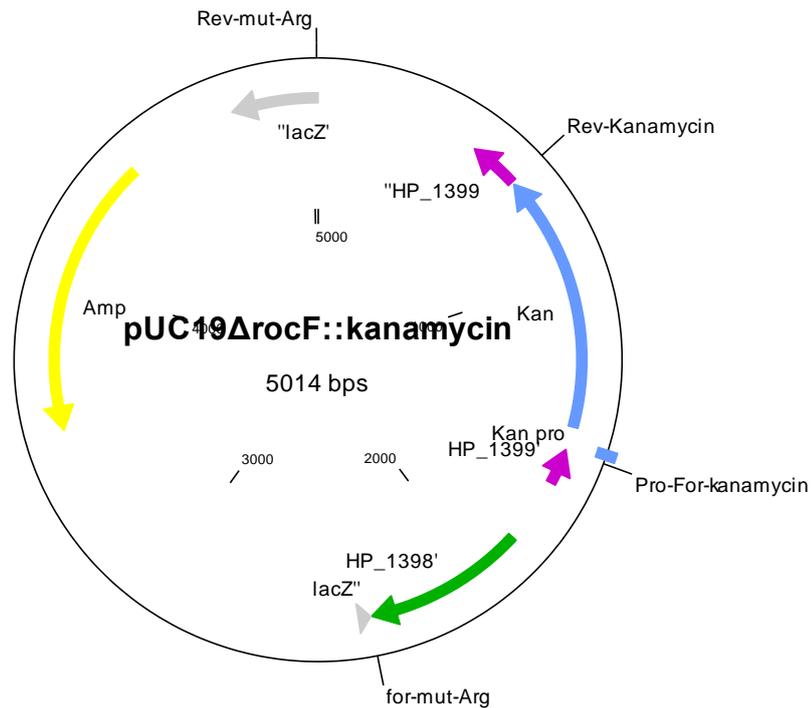


Figure 3.24 Plasmid map of pUC19 $\Delta$ *rocF*::*kan*, deletion-inactivation of *rocF* was conducted by inverse PCR mutagenesis with the insertion of promoter kanamycin cassette. This figure indicate *rocF* cloning primer pair for-mut-Arg (hp1399-F) and rev-mut-Arg (hp1399-R) in addition to kanamycin resistance cassette primers Pro-For-kanamycin (pro.kan-F) and Rev-kanamycin (kan-R) that used for identification the insertion and orientation of the cassette within the plasmid.

### 3.4 Natural transformation of mutated *pgk*, *ppsA* and *rocF* in *H. pylori* 26695

As a preliminary step for synthetic lethality investigations of *pgk* and *ppsA*, pUC19-*pgk*::*cat* and pUC19 $\Delta$ *ppsA*::*kan* constructed plasmids (in addition to pUC19 $\Delta$ *rocF*::*cat* and pUC19 $\Delta$ *rocF*::*kan* control plasmids respectively) were used for the transformation of *H. pylori* 26695, with the expectation that double crossover homologous recombination would take place between the flank regions of these mutated alleles and the wild type alleles in the chromosome of *H. pylori* 26695.

Natural transformation (described in section 2.5.6) was undertaken to introduce these constructed plasmids into *H. pylori* 26695. The plasmids were transformed into the bacteria and plated in the presence of antibiotic that correspond to the inserted

cassette resistance gene. To confirm the construction by double homologous recombination, chromosomal DNA was extracted (section 2.3) from *H. pylori* transformants and PCRs were done using different combinations of primers that localised upstream and downstream flank regions of the mutated alleles and primers of cassette resistance genes. After verifying the PCR products sizes, DNA sequencing was carried out to investigate any potential errors.

The following sections describe the last step of the first objective in this study namely the recombination of the mutated gene into the chromosome of *H. pylori* 26695.

### **3.4.1 Transformation of pUC19-*pgk*::*cat* and pUC19 $\Delta$ *rocF*::*cat* plasmids**

To replace the wild type alleles of *pgk* and *rocF* in *H. pylori* 26695 by *pgk*::*cat* and  $\Delta$ *rocF*::*cat* recombinant alleles respectively, pUC19-*pgk*::*cat* plasmid and pUC19 $\Delta$ *rocF*::*cat* control plasmid were separately introduced into *H. pylori* 26695 by natural transformation as described in section 2.5.6.

After three days of incubation on Columbia agar plates supplemented with 20  $\mu$ g/ml chloramphenicol, there 25-50 colonies grown on these antibiotic selective plates after allelic exchange. Colonies were selected and screened by colony PCR using a combination of *pgk* or *rocF* with *cat* primers to verify the existence and orientation of the chloramphenicol resistance cassette within the genomes of *pgk* and *rocF* mutated *H. pylori* strains. The following sections will deal with steps that verified *rocF* and *pgk* knockout after the introduction of pUC19-*pgk*::*cat* and pUC19 $\Delta$ *rocF*::*cat* into *H. pylori* 26695 by natural transformation.

#### **3.4.1.1 $\Delta$ *rocF*::*cat* verification in *H. pylori***

As controls increases the reliability of the results, pUC19 $\Delta$ *rocF*::*cat* control plasmid was introduced into *H. pylori* by natural transformation to compare the results with the transformation results of pUC19-*pgk*::*cat* into *H. pylori*. After positive colony selection, colony PCR screening was conducted using hp1399-F, hp1399-R, CAT-F and CAT-R

primers combinations. The PCR results verified the existence and the orientation of promoter-less chloramphenicol resistance cassette in the mutated *rocF* gene.

The genomic DNA of putative recombinant colonies were isolated and used as a template for PCR. This template and CAT-F and CAT-R primers pair were used for identifying the existence of the promoter-less *cat* cassette within  $\Delta rocF$  (figure 3. 25, panel A). Moreover, hp1399-F with CAT-R primers yielded the same expected PCR product that is approximately 1475 bp that verified the orientation of the cassette within the mutated gene (figure 3. 25, panel B).

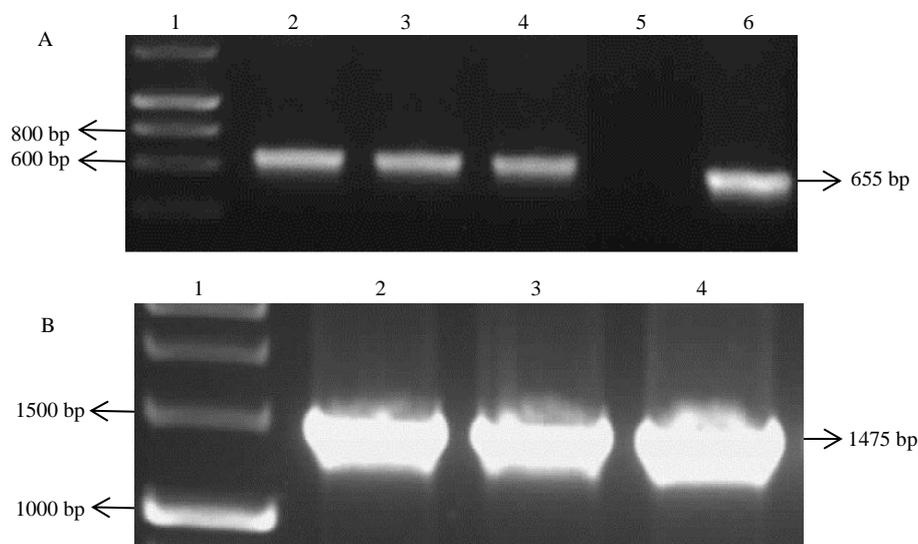


Figure 3.25 Identification of the presence and orientation of chloramphenicol resistance cassette within the mutated *rocF*. Panel (A) Verification of existence of the promoter-less cassette within  $\Delta rocF$  allele of *H. pylori* strain after natural transformation of pUC19 $\Delta rocF::cat$  plasmid. The cassette insertion verified by PCR using CAT-F and CAT-R primers and the genomic DNA of selected mutant strains to produce 655 bp expected size fragment. Lanes contain: (1) DNA Hyperladder marker (1kb), (2, 3 and 4) PCR product of random selected colonies of the  $\Delta rocF$  *H. pylori* strains, (5) -ve control (ddH<sub>2</sub>O) and (6) +ve control, the PCR product amplified from pUC19-*cat* plasmid using CAT-F and CAT-R primers. Panel (B) verification of the orientation of the cassette within the  $\Delta rocF$  allele by PCR amplification using hp1399-F and CAT-R primers and the genomic DNA of selected  $\Delta rocF$  *H. pylori* colonies. The expected molecular size of the fragment was approximately 1475 bp. Lanes contain: (1) Hyperladder molecular weight marker (1 kb), (2, 3 and 4) PCR product of the selected colonies.

After verifying the insertion of promoter-less *cat* cassette within  $\Delta rocF$  allele of *H. pylori*, the genomic DNA of one of the putative recombinant colonies was used in PCR

amplification using Extra-hp1399-F and Extra-hp1399-R primers (located 400 bp to the 5' and 3' ends of hp1399-F and hp1399-R flank regions primers respectively) (figure 3.26). This PCR produced approximately 3590 bp fragment which used as a template in sequencing using the same primer and the promoter-less *cat* primers pair for further verification of the existence of the chloramphenicol resistance cassette within the mutated allele. Sequencing results confirm the insertion of promoter-less *cat* within  $\Delta rocF$  in addition to sequence verification of the cassette and the flank region upstream and downstream the mutated gene.

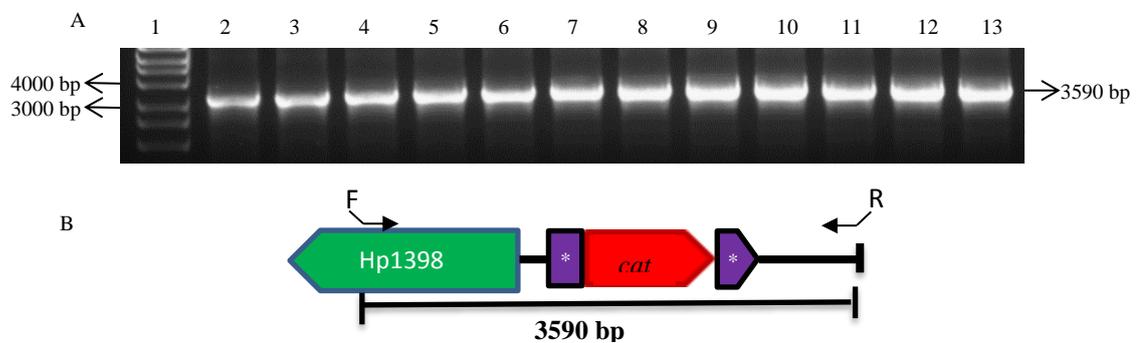


Figure 3.26 Gradient PCR of  $\Delta rocF::cat$  using Extra-hp1399-F and Extra-hp1399-R and chromosomal DNA of  $\Delta rocF::cat$  *H. pylori* strain as template. Panel A: Agarose gel electrophoresis of the gradient PCR amplification, lanes contain: lane 1. Hyper ladder DNA marker (1kb), lanes 2-13 PCR products of  $rocF::cat$  using 50-65 °C temperature gradient. The PCR products were purified and then used as template in sequencing using these primers in addition to promoter-less *cat* primers pair for further verification of chloramphenicol resistance cassette existence within  $\Delta rocF$  allele, and to investigate any potential errors in the  $rocF$  flank regions. White asterisks denote  $\Delta rocF$  allele. Panel B: schematic representation of the gradient PCR. Solid arrows show primer position for annealing, open arrows indicate coding sequence orientation.

### 3.4.1.2 *pgk::cat* verification in *H. pylori*

As with *pUC19ΔrocF::cat* plasmid control verification, Colonies of the *pgk* mutated *H. pylori* (produced from the introduction of *pUC19-pgk::cat* into *H. pylori* by natural transformation) were screened by colony PCR using the same set of primers which were used in *pgk* mutagenesis within *pUC19* (described in section 3.3.1). Similarly, hp1345-F and hp1345-R primers and *cat* specific primers (pro-CAT-F and CAT-R) were used to

verify the presence and the orientation of the cassette resistance gene within the mutated strain. The putative recombinant colonies were selected for chromosomal DNA extraction (described in section 2.3) to use the genomic DNA of each colony as a template in PCR. Using this template and the same set of primers used in colony PCR, the existence of *cat* within the genome of *pgk::cat H. pylori* strain was identified (figure 3.27, panel A). Concurrently, this PCR yielded an approximately 1835 bp product (figure 3.27, panel B) which identified the orientation of the cassette within *pgk*.

After successful identification of mutating the wild type *pgk* allele of *H. pylori* strain by the insertion of chloramphenicol resistance cassette, a PCR product was required for further verification by DNA sequencing. This can be achieved by the amplification of DNA fragment located between the extra flanking regions of *pgk*. After that, this fragment can be used as a template in the sequencing. The fragment was amplified from the genomic DNA of the *pgk::cat H. pylori* strain using Extra-hp1345-F and Extra-hp1345-R primers. The (Extra) primers were located approximately 200 base pairs upstream and downstream the 5' and 3' ends of *pgk::cat* flank regions respectively. These primers were designed to anneal outside the region cloned within pUC19 (pUC19-*pgk::cat*) to show the recombination has occurred at the right chromosomal locus.

The PCR produced two fragments, the size of the smaller fragment corresponds to the expected size of the wild type *pgk* allele in addition to the flank regions and the extra flank regions, and the other fragment corresponds to the same expected size of the smaller fragment in addition to the expected size of chloramphenicol resistance cassette. This observation was investigated and is covered in more detail later in this chapter (section 3.6).

As it was not possible to use the PCR product (located between the extra flanking regions of *pgk*) to verify the insertion of *cat* within *pgk* allele by sequencing. This can be explained that the sequencing was much more difficult because of the presence of two amplicons (mixed templates) in the PCR reaction. Accordingly, two PCR amplifications were carried out using genomic DNA of the *pgk::cat H. pylori* strain as template in addition to primer combinations of Extra-hp1345-F with CAT-R (figure 3.28) and Extra-hp1345-R with pro-CAT-F (figure 3.29) (table 2.2). The PCR products were purified and

sequenced using the same set of primers (Extra-hp1345-F with CAT-R and pro-CAT-F with Extra-hp1345-R). Sequencing verified the existence of the chloramphenicol resistance cassette within the *pgk* allele, and confirmed the absence of any sequence errors in the *pgk* flank regions during homologous recombination.

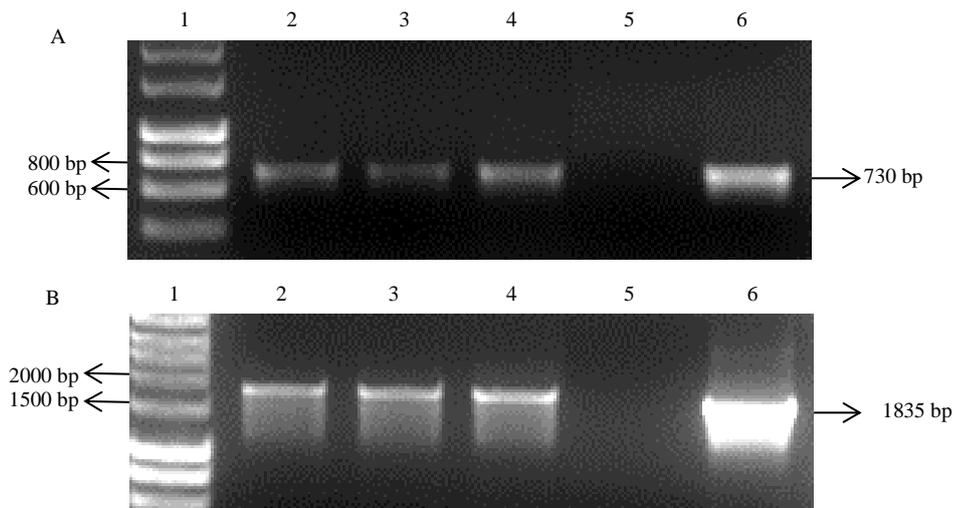


Figure 3.27 Conformation of existence and orientation of chloramphenicol resistance cassette within *pgk* wild type allele in the genome of *H. pylori* strain. Panel (A) Verification of existence of the cassette within the wild type allele of *pgk* mutated *H. pylori* strain after natural transformation of pUC19-*pgk::cat* plasmid. The existence verified by PCR using pro-CAT-F and CAT-R primers and the genomic DNA of random selected mutant strains to produce 730 bp expected size fragment. Lanes contain: (1) DNA Hyperladder marker (1kb), (2, 3 and 4) PCR product of random selected colonies of the *pgk* mutated strain, (5) -ve control (ddH<sub>2</sub>O) and (6) +ve control, the PCR product amplified from pUC19-*cat* plasmid using pro-CAT-F and CAT-R primers. Panel (B) verification of the orientation of the chloramphenicol resistance cassette within the *pgk* mutant allele by PCR amplification using hp1345-F and CAT-R primers and the genomic DNA of random selected mutant *H. pylori* strains. The expected molecular size of the fragment was approximately 1835 bp. Lanes contain: (1) Hyperladder molecular weight marker (1 kb), (2, 3 and 4) PCR product of random selected colonies of the *pgk* mutated strain, (5) Negative (-ve) control (ddH<sub>2</sub>O) and (6) positive (+ve) control, the PCR product amplified from pUC19-*pgk::cat* plasmid using hp1345-F and CAT-R primer pair.

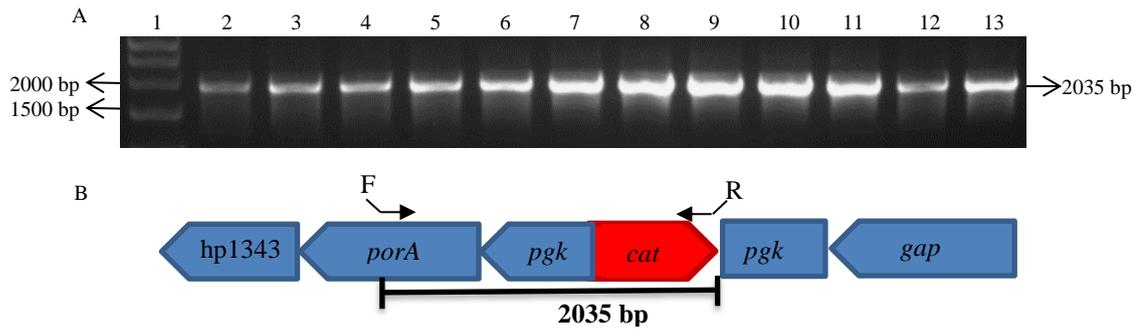


Figure 3.28 Gradient PCR of *pgk::cat* using Extra-hp1345-F and CAT-R. Panel A: Agarose gel electrophoresis of the gradient PCR amplification, lanes contain: lane 1. Hyper ladder DNA marker (1kb), lanes 2-13 PCR products of *pgk::cat* using 50-65 °C temperature gradient. The PCR products were purified and then used as template in sequencing using the same primers to investigate any potential errors in the 5' *pgk* flank region and to confirm the presence of *cat* resistance cassette. Panel B: schematic representation of the gradient PCR. Solid arrows show primer position for annealing, open arrows indicate coding sequence orientation.

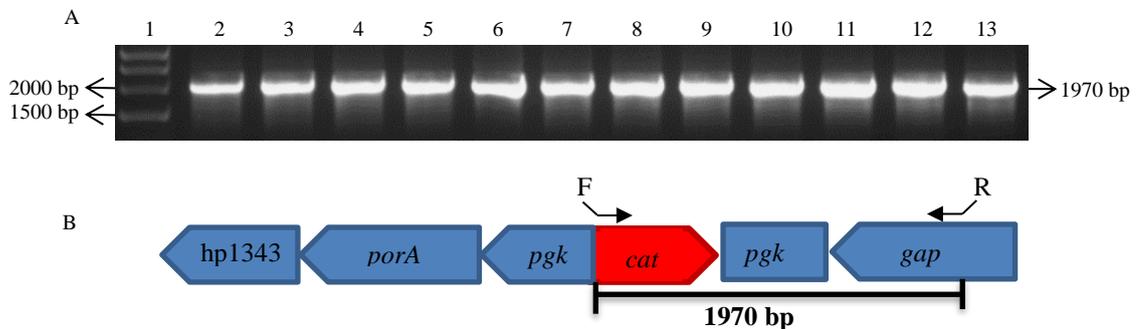


Figure 3.29 Gradient PCR of *pgk::cat* using pro-CAT-F and Extra-hp1345-R. Panel A: Agarose gel electrophoresis of the gradient PCR amplification, lanes contain: lane 1. Hyper ladder DNA marker (1kb), lanes 2-13 PCR products of *pgk::cat* using 50-65 °C temperature gradient. The PCR products were purified and then used as template in sequencing using the same primers to investigate any potential errors in the 3' *pgk* flank region and to confirm the presence of *cat* resistance cassette. Panel B: schematic representation of the gradient PCR. Solid arrows show primer position for annealing, open arrows indicate coding sequence orientation.

### 3.4.2 Transformation of pUC19 $\Delta$ *ppsA*::*kan* and pUC19 $\Delta$ *rocF*::*kan* plasmids

Another part of the recombination of the mutated genes within the chromosome of *H. pylori* (prior to synthetic lethality investigation) was conducted by creating *H. pylori* strains in which the target allele ( $\Delta$ *ppsA*) and the positive control allele ( $\Delta$ *rocF*) were introduced separately into *H. pylori* 26695. The mutagenesis strategy followed to create constructs containing mutated *ppsA* and *rocF* alleles were described previously (sections 3.3.2 and 3.3.3.2 respectively).

Initially, pUC19 $\Delta$ *ppsA*::*kan* plasmid and pUC19 $\Delta$ *rocF*::*kan* control plasmid were introduced into wild type *H. pylori* 26695 by natural transformation (described in section 2.5.6). Transformants of *ppsA* or *pgk* mutated allele were plated on Columbia agar plates supplemented with 50  $\mu$ g/ml kanamycin for the selection of recombinant cells. After three to seven days of incubation, no colonies of  $\Delta$ *ppsA*::*kan* *H. pylori* transformants strains were detected. The experiment was repeated three times using pUC19 $\Delta$ *rocF*::*kan* *H. pylori* transformants cells as a positive control. The concentration of kanamycin was decreased to 25  $\mu$ g/ml in some Columbia agar plates to increase recovery chances of mutated *ppsA* *H. pylori* cells. Furthermore, *H. pylori* J99 strain was used as recipient to transform pUC19 $\Delta$ *ppsA*::*kan* as *ppsA* and the flank regions (belonging to hp120, hp0122 and hp0123) in both strains are sharing 97-98 % sequence similarity. Similarly, no mutated *ppsA* *H. pylori* transformants were obtained. This preliminary result highlighted that *ppsA* allele might be essential alone in *H. pylori* 26695 and confirming that *ppsA* is potentially vital to the viability of this pathogen (table 3.1).

Regarding natural transformation of pUC19 $\Delta$ *rocF*::*kan* control plasmid into *H. pylori* 26695, less than 10 colonies of  $\Delta$ *rocF*::*kan* *H. pylori* strains were obtained after three days of incubation on kanamycin containing plates. These colonies were screened by colony PCR to confirm that the promoterized kanamycin resistance cassette was properly located within the chromosome of  $\Delta$ *rocF* *H. pylori* strain. Hp1399-F with kan-R and pro-kan-F with kan-R primers sets confirmed the orientation and the existence of the cassette. After that, putative recombinant colonies were selected for genomic DNA

extractions (described in section 2.3) which were used as templates in PCR amplifications. This PCR verified the expected size of approximately 2100 bp fragment using Extra-hp1399-F and kan-R (figure 3.30, panel A) and confirmed the presence of kanamycin resistance cassette within  $\Delta rocF$  *H. pylori* genome (figure 3.30, panel B). This result increases the reliability of the pUC19 $\Delta ppsA::kan$  transformation and recombination within the genome of *H. pylori* 26695. Furthermore, it decreases the possibility that the transformation failed due to technical error.

The genomic DNA of one of the putative recombinant colonies was used in PCR amplification after identifying the insertion of promoterized kanamycin cassette within  $\Delta rocF$  allele of *H. pylori* strain. Similarly to  $\Delta rocF::cat$ , this PCR was carried out using Extra-hp1399-F and Extra-hp1399-R primers (section 3.4.1). Approximately 3160 bp PCR product (figure 3.31) was sequenced using these primers and the pro-kan-F and kan-R primers pair for additional confirmation of kanamycin resistance cassette presence within the mutated allele. The results confirmed the insertion of promoterized kanamycin within  $\Delta rocF$  in addition to sequence verification of the cassette and the flank region upstream and downstream  $\Delta rocF::kan$ .

In conclusion, natural transformation results of pUC19 $\Delta ppsA::kan$  into *H. pylori* 26695 showed that this allele might potentially be essential in the enteric pathogen. However, pUC19-*pgk::cat* plasmid and pUC19 $\Delta rocF::cat$  control plasmid natural transformation results confirmed the recombination and allelic replacement of wild type allele with mutant allele within the genome of the bacteria. Hence, these transformant strains were used as recipients to introduce pUC19 $\Delta ppsA::kan$  by natural transformation for obtaining double mutants strain, which is the final step of synthetic lethality investigation of *ppsA* and *pgk* in *H. pylori* 26695.

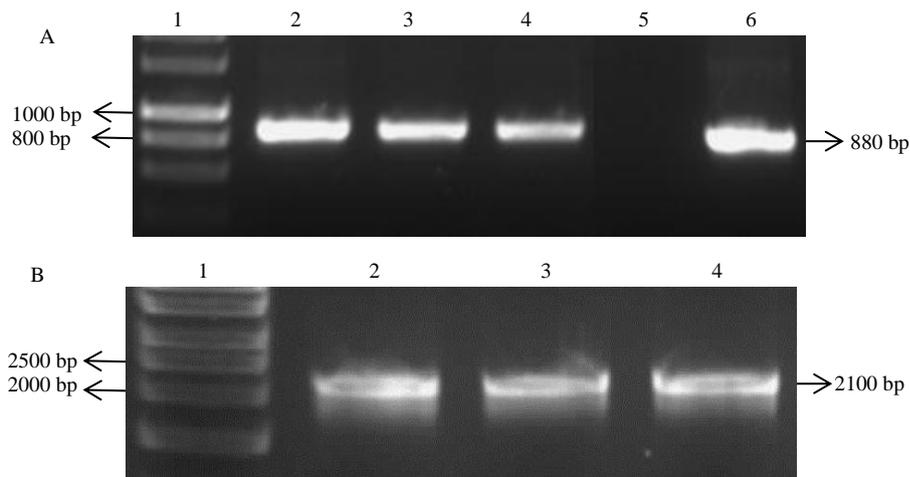


Figure 3.30 Location and orientation verification of  $\Delta rocF::kan$  in the genome of *H. pylori* strain. Panel (A) Identification the presence of promoter kanamycin resistance cassette within  $\Delta rocF$  allele of *H. pylori* strain after natural transformation of pUC19 $\Delta rocF::kan$  plasmid. This was carried out by PCR using pro-kan-F and kan-R primers and the genomic DNA of selected putative recombinant strains to produce 880 bp expected size fragment. Lanes contain: (1) DNA Hyperladder marker (1kb), (2, 3 and 4) PCR product of selected colonies of *H. pylori* strains with  $\Delta rocF::kan$  mutation, (5) -ve control (ddH<sub>2</sub>O) and (6) +ve control, the PCR product amplified from pUC19-kanamycin plasmid using pro-kan-F and kan-R primers. Panel (B) Promoter kanamycin cassette within *rocF* mutant allele orientation verification by PCR amplification using Extra-hp1399-F and kan-R primers and the genomic DNA of selected  $\Delta rocF::kan$  *H. pylori* strains. The expected molecular size of the fragment was approximately 2100 bp. Lanes contain: (1) Hyperladder molecular weight marker (1 kb), (2, 3 and 4) PCR products of selected colonies of  $\Delta rocF::kan$  strains.

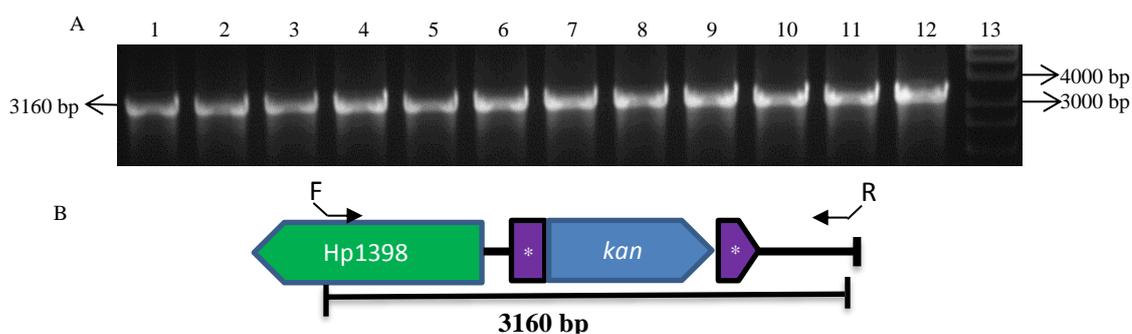


Figure 3.31 Gradient PCR of  $\Delta rocF::kan$  using Extra-hp1399-F and Extra-hp1399-R and chromosomal DNA of  $\Delta rocF::kan$  *H. pylori* strain as template. Panel A: Agarose gel electrophoresis of the gradient PCR amplification, lanes contain: lanes 1-12 PCR products of *rocF::kan* using 50-65 °C temperature gradient, lane 13. Hyper ladder DNA marker (1kb). The PCR products were purified and then used as template in sequencing using these primers in addition to promoter kanamycin primers pair for further verification of kanamycin resistance cassette existence within  $\Delta rocF$  allele, and to investigate any potential errors in the *rocF* flank regions. White asterisks denotes  $\Delta rocF$  allele. Panel B: schematic representation of the gradient PCR. Solid arrows show primer position for annealing, open arrows indicate coding sequence orientation.

### 3.5 Synthetic lethality of *ppsA* and *pgk*

As a final step in the synthetic lethality investigation of *ppsA* and *pgk* in *H. pylori* 26695, pUC19 $\Delta$ *ppsA*::*kan* construct was introduced by natural transformation into *pgk*::*cat* *H. pylori* strain and  $\Delta$ *rocF*::*cat* *H. pylori* control strain (as described in section 2.5.6). Transformant cells were grown on Columbia agar plates supplemented with 50  $\mu$ g/ml kanamycin. This was for allelic exchange selection of  $\Delta$ *ppsA*::*kan* fragment into the *ppsA* locus within the *pgk*::*cat* and  $\Delta$ *rocF*::*cat* *H. pylori* strains. After more than seven days of incubation under microaerobic conditions, no colonies of the transformed *pgk*::*cat* and  $\Delta$ *rocF*::*cat* *H. pylori* strains containing  $\Delta$ *ppsA*::*kan* mutated allele were detected on the selective medium. The concentration of kanamycin was decreased to 25  $\mu$ g/ml in some Columbia agar plates to increase chances of recovery. As another control, pUC19-*pgk*::*cat* plasmid was introduced to  $\Delta$ *rocF*::*kan* *H. pylori* strain by natural transformation. After three days of incubation, there were around 25-35 colonies on Columbia agar plates supplemented with 20  $\mu$ g/ml chloramphenicol.

Table 3.1 investigation of *ppsA* and *pgk* Synthetic lethality by natural transformation of *ppsA*, *pgk*, and *rocF* mutated constructs into the wild type, *pgk* mutated and *rocF* mutated *H. pylori* strains.

Construct	<i>H. Pylori</i> strain	Transformation result
pUC19 $\Delta$ <i>ppsA</i> :: <i>Kan</i>	26695	-
pUC19 $\Delta$ <i>ppsA</i> :: <i>Kan</i>	J99	-
pUC19 $\Delta$ <i>rocF</i> :: <i>Kan</i>	26695	+
pUC19 $\Delta$ <i>ppsA</i> :: <i>Kan</i>	<i>Pgk</i> :: <i>cat</i>	-
pUC19 $\Delta$ <i>ppsA</i> :: <i>Kan</i>	$\Delta$ <i>rocF</i> :: <i>cat</i>	-
pUC19- <i>pgk</i> :: <i>cat</i>	$\Delta$ <i>rocF</i> :: <i>kan</i>	+

As it was not possible to obtain *ppsA* with *pgk* double mutants to verify synthetic lethality, these results support the potential evidence of *ppsA* essentiality in *H. pylori* 26695 and J99 and highlight that *ppsA* alone might be vital to the viability of this pathogenic bacterium.

### 3.6 Wild type and inactivated *pgk* alleles variation in *pgk::cat H. pylori*

As described in section 3.4.1.1, the presence and orientation of the chloramphenicol resistance cassette was identified within the wild type *pgk* allele in the genome of random selected colonies of this mutated *H. pylori* strain. However, a PCR product was required (as a template) for further verification by DNA sequencing. This was achieved by the amplification of DNA fragment from the genomic DNA (of the random selected colonies) of *pgk::cat H. pylori* strain using Extra-hp1345-F and Extra-hp1345-R primers. This PCR produced two fragments namely 2540 bp fragment corresponds to the expected size of the wild type *pgk* allele (in addition to the flank regions and the extra flank regions) and 3270 bp fragment corresponds to the same expected size of this fragment in addition to the expected size of chloramphenicol resistance cassette. The PCR product were visualized by agarose gel electrophoresis for 2.5 hours using 1.5% agarose (figure 3.32).

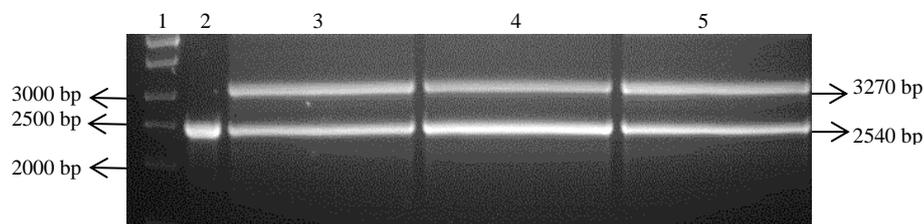


Figure 3.32 Agarose gel electrophoresis of wild and *cat* inactivated *pgk* PCR product from chromosomal DNA of *pgk::cat H. pylori* strain using Extra-hp1345-F and Extra-hp1345-R primes. Lanes contain: lane 1. Hyper ladder DNA marker (1kb), lane 2. Wild *pgk* allele PCR product from genomic DNA of *H. pylori* 26695 using Extra-hp1345-F and Extra-hp1345-R primes (positive control), lanes 3-5 PCR products of wild and *cat* inactivated *pgk* alleles from genomic DNA of three random selected colonies of *pgk::cat H. pylori* strain. The PCR products contain mixture of *pgk* allele (DNA band of 2540 bp) and *cat* inactivated *pgk* allele (DNA band of 3270 bp).

According to that variation, it was not possible to use this mixture of the wild type and the *cat* inactivated *pgk* alleles as a template in sequencing. However, this product was investigated to understand the reason behind this phenomenon. Each band was excised and purified from the gel and then used as templates in PCR amplification and

sequencing using Extra-hp1345-F with Extra-hp1345-R primers and promoterized *cat* primer pairs. Neither amplification nor sequencing positive results was obtained.

For further investigation, a unique *EcoRI* restriction site within *pgk* allele was identified with the aid of Clone Manager version 9. After that, *EcoRI* restriction digestion was carried out using the wild type with *cat* inactivated *pgk* alleles PCR product and a control *pgk* PCR product. These DNA fragments amplified from the genomic DNA of *pgk::cat* and the wild type *H. pylori* strains respectively using Extra-hp1345-F and Extra-hp1345-R primers. Digestion results verified the existence of wild type and *cat* inactivated *pgk* alleles as the same expected sizes calculated by clone manager was achieved experimentally. *EcoRI* digestion reported 959 bp and 1579 bp fragments resulted from 2538 bp wild type *pgk* allele fragment (figure 3.33 panel B). At the same time, another 2313 bp fragment resulted from the restriction of *pgk::cat* allele in addition to the same corresponding bands of the *EcoRI* restricted wild type *pgk* allele control. This means the presence of extra 734 bp which correspond to the same expected size of the *cat* gene with the cognate promoter that used to mutate *pgk* by insertion (section 3.3.1). The latter result verified the presence of *cat* within *pgk* allele (figure 3.33 panel A).

These results highlighted that the presence of wild type and *cat* inactivated *pgk* alleles in the PCR product of *pgk::cat H. pylori* might be due to single crossover or double crossover homologous recombination between one or both of *pgk::cat* flank regions and the flank region(s) of the wild *pgk* allele respectively. This means the existence of wild and inactivated *pgk* alleles in addition to the integration of pUC19 plasmid within the chromosome of *pgk::cat H. pylori* strain. As ampicillin gene version is a part of pUC19 plasmid (figure 3.13), this notion was experimentally checked by PCR amplification using combinations of ampicillin specific primers (amp-F and amp-R), *pgk* flank region primers (hp1345-F and hp1345-R) and Inverse *pgk* primers (Inv-hp1345-F and Inv-hp1345-R). No apparent PCR product was identified to verify the existence of pUC19 plasmid within the chromosome of *pgk::cat H. pylori* strain.

After that, gradient PCR amplifications at a range of 50-65 °C were conducted using genomic DNA of *pgk::cat H. pylori* strain and wild type *H. pylori* 26695 as templates in addition to different combinations of three *pgk* primers sets namely hp1345-F with

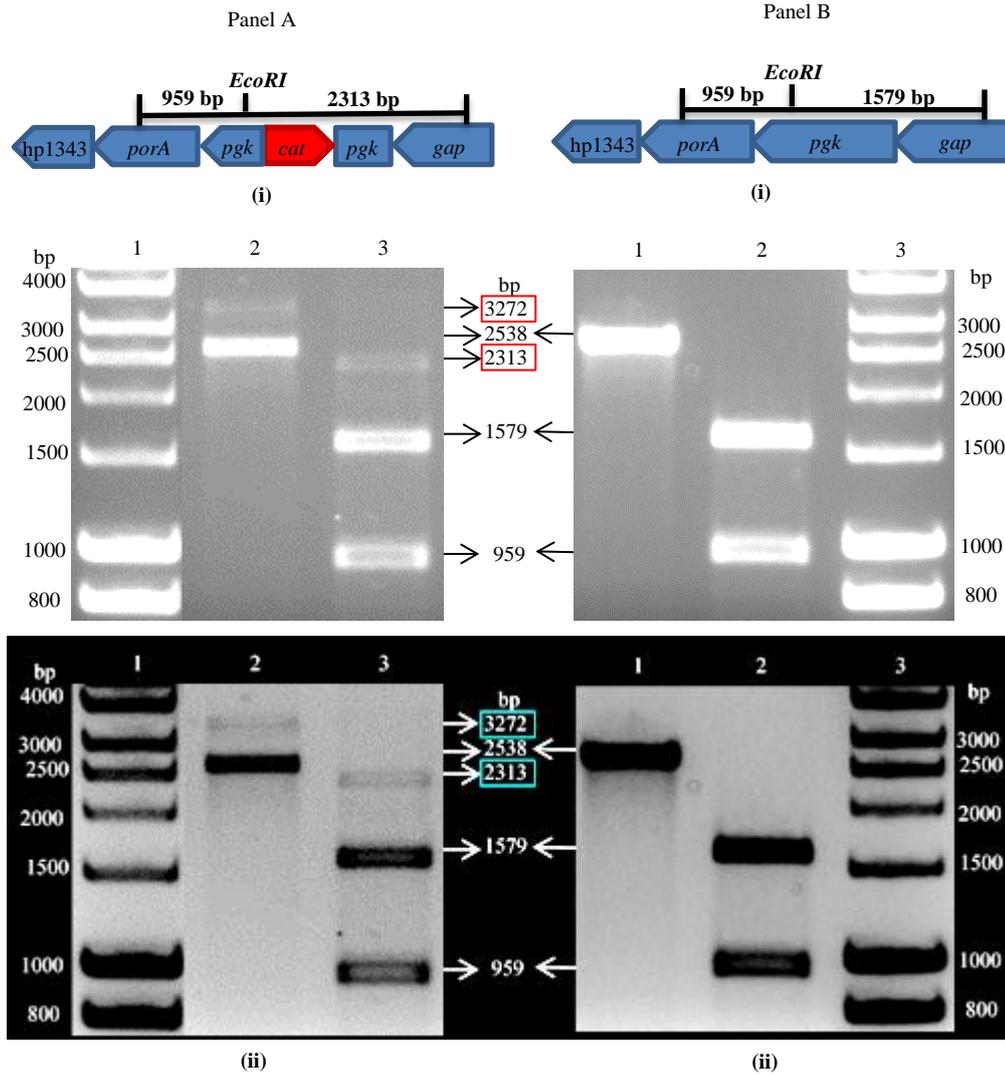


Figure 3.33 Original and inverse photo of *EcoRI* restriction digestion of wild and *cat* inactivated *pgk* alleles PCR product from chromosomal DNA of *pgk::cat H. pylori* strain using wild *pgk* allele (with flank and extra flank regions) PCR product from *H. pylori* 26695 as control. Panel A: (i) lanes contain: 1. DNA hyper ladder marker (1kb), 2. -ve control uncut wild and *cat* inactivated *pgk* alleles PCR product which consists of wild *pgk* allele fragment (2538 bp) and *cat* inactivated *pgk* allele fragment (3272 bp) (red box) and 3. *EcoRI* restriction of wild and *cat* inactivated *pgk* alleles PCR product. Wild *pgk* allele restriction resulted 959 bp with 1579 bp fragments while the digestion of *cat* inactivated *pgk* allele produced 959 bp with 2313 bp fragment (red box). The later fragment consists of *pgk* fragment (1579 bp) with promoter *cat* fragment (734 bp). (ii) Schematic representation of *EcoRI* restriction of *cat* inactivated *pgk* allele (with flank regions and extra flank regions). Panel B: (i) lanes contain: 1. -ve control Wild *pgk* allele (with flank and extra flank regions) PCR product (2538 bp), 2. +ve control *EcoRI* digestion of wild *pgk* allele PCR product showing 959 bp with 1579 bp fragments, and 3. DNA hyper ladder marker (1kb). (ii) Schematic representation of *EcoRI* restriction of wild *pgk* allele PCR product (with flank and extra flank regions).

hp1345-R, flank-hp1345-F with flank-hp1345-R and extraflank-hp1345-F with extraflank-hp1345-R. Nine PCR products with controls confirmed the presence of DNA fragments with molecular sizes correspond to that of the wild type and *cat* inactivated *pgk* alleles variation within all the PCR products. Hence, the presence of both wild type and mutant *pgk* alleles suggest two possibilities either the gene is essential to the viability of *H. pylori* or this presence is due to synergy phenomenon.

### 3.7 Summary

As *ppsA* and *pgk* were hypothesized as synthetic lethal metabolic genes by *in silico* double-deletion mutational studies (Thiele et al., 2005), *in vivo* gene knock out strategies were employed to investigate this hypothesis in the enteric pathogen *H. pylori* 26695. To achieve this goal, *pgk* and *ppsA* ORF's (in addition to *rocF* allele as a control) with flanking regions upstream and downstream of each allele (for double crossover homologous recombination) were successfully amplified. Plasmids carrying the amplified genes were cloned within pUC19 and then insertion and deletion inactivation mutagenesis strategies were employed by inserting antibiotics resistant cassettes inside the target alleles.

The pUC19-*pgk::cat*, pUC19 $\Delta$ *rocF::cat*, pUC19 $\Delta$ *ppsA::kan* and pUC19 $\Delta$ *rocF::kan* constructed plasmids were introduced by natural transformation into the genome of wild type *H. pylori* 26695 to mutate *pgk* and *ppsA* alleles using  $\Delta$ *rocF::cat* and  $\Delta$ *rocF::kan* alleles as controls respectively.

Natural transformation resulted *pgk::cat* and  $\Delta$ *rocF::cat* putative recombinant *H. pylori* strains. Conversely, *ppsA* allele could not be mutated as it might be essential to the viability of the enteric pathogen. For further verification, pUC19 $\Delta$ *ppsA::kan* plasmid was introduced to *pgk::cat* and  $\Delta$ *rocF::cat* *H. pylori* strains by natural transformation. Similarly, *ppsA* could not be mutated neither in *pgk::cat* strain nor in  $\Delta$ *rocF::cat* *H. pylori* control strain. These results were disagreed with the computational findings of *ppsA* and *pgk* synthetic lethality.

Another result was investigated in this chapter that is the variation of wild type and *cat* inactivated *pgk* alleles in *H. pylori*. The reason behind this potential variation was

checked via *EcoRI* restriction, single crossover homologues recombination possibility check and gradient PCR amplifications using different combinations of extra flank, flank and *pgk* primers sets. The results further verified the existence of the same variation. However, the variation results required more investigations such as DNA southern blotting to identify any relation between the variation and *pgk* essentiality to the viability of *H. pylori* 26695.

## Chapter 4. Structure and function of phosphoglycerate kinase and Bioinformatics of phosphoenolpyruvate synthase in *H. pylori* 26695

### 4.1 Introduction

*H. pylori* genome sequencing analysis and metabolic modelling studies revealed that the glycolytic pathway is incomplete in this enteric pathogen as genes for the key enzymes were not found (Tomb et al., 1997, Marais et al., 1999, Schilling et al., 2002). Similarly, C-13 NMR spectroscopy experimental studies failed to detect the activity of these enzymes (Chalk et al., 1994). Moreover, genome sequencing analysis and experimental studies suggested that glucose is the only carbohydrate utilised by the bacteria (Mendz and Hazell, 1993, Tomb et al., 1997, Psakis et al., 2009). However, this can be formed by the Entner-Doudoroff pathway which is constitutively active in *H. pylori* (Wanken et al., 2003). Nonetheless, glucose is utilised sparingly as the main carbon and energy sources in *H. pylori* are amino acids (Nagata et al., 2003).

The gluconeogenic pathway in the pathogenic bacteria is complete and so includes the key enzymes fructose-1,6-bisphosphatase (FBPase) and phosphoenolpyruvate synthase (PPSA) (Tomb et al., 1997). The latter is crucial to link the Krebs cycle with the gluconeogenic-glycolytic pathways in order to satisfy gluconeogenic requirements to provide precursors for anabolic processes (Doig et al., 1999). Accordingly, understanding of metabolism within *H. pylori* is important for the development of possible leads for structure based drug design. Within the carbohydrate metabolic pathways, the results of computational double deletion mutational studies were used to hypothesise that PPSA with phosphoglycerate kinase (PGK) are synthetic lethal genes, essential to the viability of *H. pylori* (Thiele et al., 2005). This identifies these metabolic enzymes as potential targets.

The previous chapter described how hp0121 (*ppsA*) and hp1345 (*pgk*) alleles were mutated and their role in synthetic lethality was investigated. In spite of the *ppsA* mutated allele alone showing a potential to be an essential gene to *H. pylori* viability, *pgk* can be targeted alone or with other enzymes of the Krebs cycle (such as fumarase)

as it deviates from the standard text book examples (section 1.5.3) (Kather et al., 2000, Chen et al., 2012). Hence it is worthwhile to determine the structure and characterise the functional role of *H. pylori* phosphoenolpyruvate synthase (hpPPSA) and phosphoglycerate kinase (hpPGK) in the context of the glycolytic-gluconeogenic pathways. In order to investigate the structural and kinetic assessment of hpPGK and hpPPSA, the enzymes ideally need to be prepared and purified in useful quantities. This was to be achieved by recombinant expression with affinity tags and purified by affinity chromatography. The aim of the work described in this chapter was the expression and purification and characterisation of the hpPGK and hpPPSA enzymes. For hpPGK, this was achieved by biochemical and X-Ray crystallographic analysis, whereas for hpPPSA it was necessary to employ structural prediction based on bioinformatics analysis.

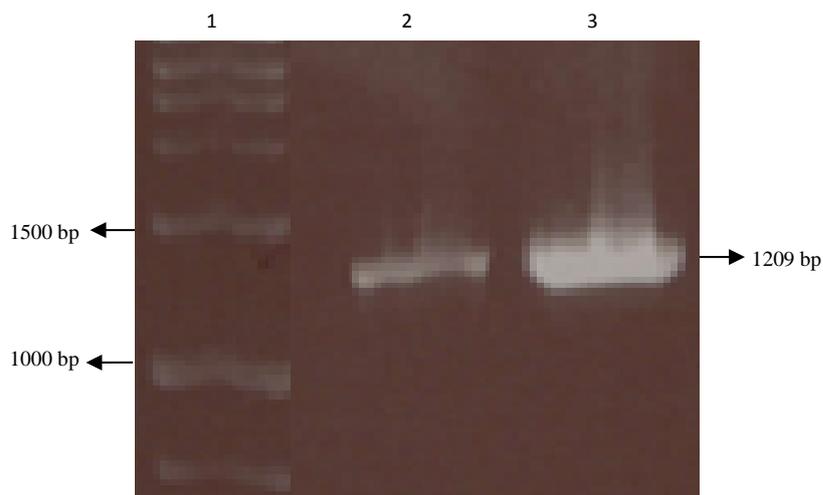
## 4.2 Cloning of *pgk* and *ppsA* in expression vectors

The aim of this part of the study is the amplification of hp1345 (*pgk*) and hp0121 (*ppsA*) open reading frames (ORFs) and manipulating the amplified genes in expression vectors. This step is essential to introduce the target genes constructed in high level expression vectors prior to overexpression optimisation and purification of the target enzymes produced by these genes.

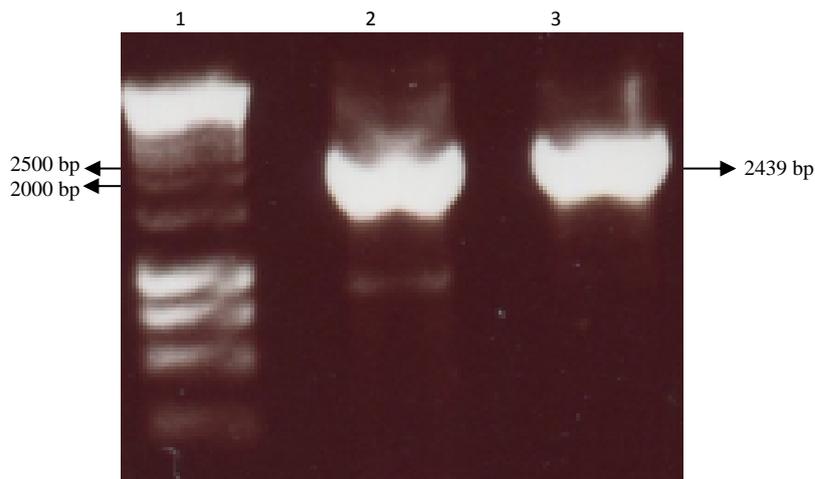
The *pgk* and *ppsA* genes were amplified from the genomic DNA of *H. pylori* 26695 using HPpgk-F with HPpgk-R and HPppsA-F and HP-ppsA-R (described in section 2.5.2.1) primer pairs respectively. These primers were designed according to the guidelines provided by the protein expression service, PROTEX lab, in the Molecular and Cellular Biology Department-University of Leicester. The successes of PCR amplification of *pgk* (1209 bp) and *ppsA* (2439 bp) were confirmed by agarose gel electrophoresis as shown in figures 4.1 and 4.2 respectively. The gel analysis shows obvious DNA bands of the amplified genes similar to the expected sizes reported in the genetic data bases.

The PCR products were submitted to PROTEX to insert the *pgk* and *ppsA* genes into the pLleic01 expression plasmid. The cloning was carried out by Dr. Xiowen Yang. This is to produce N-terminal hexa-histidine tag (N-His tag) and a Tobacco Etch Virus (TEV) protease cleavage site (at the 5' end of the coding sequence) proteins. However, only

*pgk* was cloned successfully in the Pleic01 expression plasmid. Further cloning trials were carried out to insert *ppsA* within other expression vectors of the PROTEX library. The gene was cloned into pLeic05 (C-terminal six histidine tag) and an untagged modified pLeic vector. Nonetheless, no apparent tagged hpPPSA protein were detected after expression optimisation. Finally, *ppsA* was successfully cloned into the pLEIC02 expression vector. The expressed product is an N-terminal glutathione tagged with a TEV cleavage site protein (N-GST- tag). Two pLeic01::*pgk* and pLeic02::*ppsA* constructs were obtained from PROTEX. These constructs were used as templates to identify the *pgk* and *ppsA* sequences by PNAFL using pLeic01 and pLeic02 forward with reverse primers pairs respectively. After sequence verification these constructs were used in the next stages of hpPGK and hpPPSA expression trials.



**Figure 4.1** PCR amplification of HP*pgk* gene tested at different annealing temperatures. Gel is 1% w/v agarose. Gels viewed under UV light with ethidium bromide (EtBr) used to visualise the products of PCR. Lane 1. Hyper ladder DNA (1kb) (Bioline), lane 2. HP*pgk* using 2 mM MgSO<sub>4</sub> with T<sub>m</sub>: 50 °C, lane 3. HP *pgk* using 2 mM MgSO<sub>4</sub> with T<sub>m</sub>: 55 °C



**Figure 4.2** PCR amplification of *HPppsa* gene at different annealing temperatures. Gel is 1% w/v agarose. Gels viewed under UV light with ethidium bromide (EtBr) used to visualise the products of PCR. Lane 1. Hyper ladder DNA (1kbp) (Bioline), lane 2. *HPppsa* using 1.5 mM MgSO<sub>4</sub> with T<sub>m</sub>: 50 °C, lane 3. *HPeno* using 1.5 mM MgSO<sub>4</sub> with T<sub>m</sub>: 55 °C.

## 4.3 Expression and purification of hpPGK and hpPPSA

### 4.3.1 Expression trials of hpPGK and hpPPSA

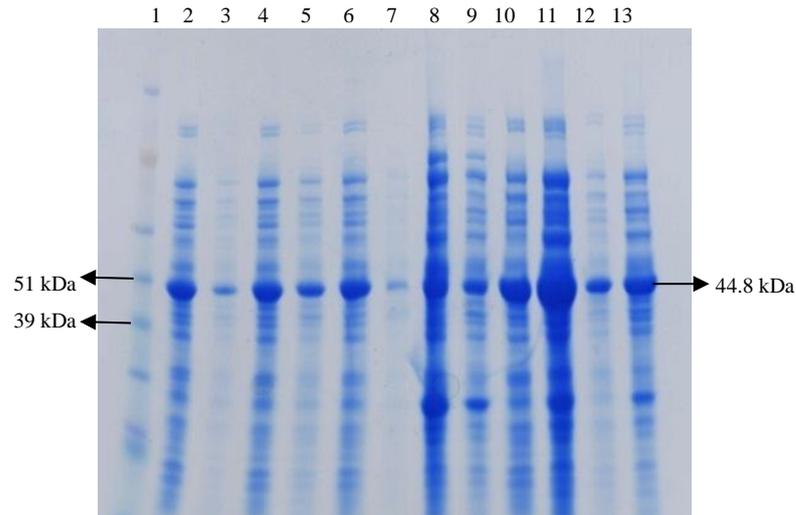
Prior to preparative expression and purification, the best expression conditions were investigated by inducing the transformed cells at different IPTG concentrations and temperatures as described in section 2.6.2. The best expression of soluble protein for hpPGK was obtained when the cells were expressed in *E. coli* Rosetta and induced at 20 °C by IPTG concentration of 100-150 μM (Figure 4.3). The molecular weight of the target protein bands seems to be similar to the theoretical molecular weight which is approximately 44,800 Dalton.

Several expression conditions and constructs were tested in order to investigate the best conditions for the various constructs, C-terminal His-tagged hpPPSA and untagged hpPPSA did not yield any over expressed proteins. The expression system was also tried in *E. coli* BL21 (DE3), *E. coli* ROSETTA and *E. coli* Origami B, none of which produced over expressed protein. However, Pleic2 (GST-tagged) yielded proteins in inclusion bodies

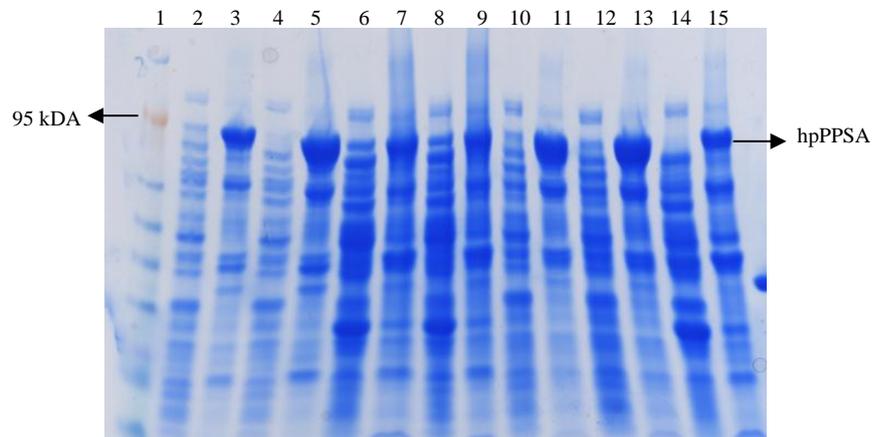
using 150-1000  $\mu$ M IPTG induction overnight at 20 °C (Figure 4.4). The target protein bands on SDS-PAGE gel seem smaller than the predicted molecular weight, this happens if the protein was synthesized as truncated unstable protein which then aggregated to form inclusion bodies of the immature and/or improperly folded proteins.

A potential reason may be because hpPPSA (approximately 91,000 Dalton (Da)) and a glutathione tag (GST) (approximately 26,000 Da) together exceed expected expression limits for proteins (100,000 Da) in *E. coli*. Furthermore, some foreign genes lose their spatio-temporal control of their expression when introduced into *E. coli*. This microenvironment can affect the newly synthesised recombinant protein in terms of changing the original source microenvironment such as pH, osmolarity, redox potential, cofactors and folding mechanisms.

Also, in overexpression conditions where efficient promoters are used, hydrophobic stretches in the recombinant protein are present at high concentrations and available for interaction with similar regions. These factors can lead to aggregation and instability of the recombinant protein. These protein aggregates known as inclusion bodies which form as a result of the imbalance between aggregation and solubilisation equilibrium (Rosano and Ceccarelli, 2014). Nonetheless, the identity of the target protein needs to be verified by peptide mass fingerprinting which may allow using the optimum overexpression condition in N-GST hpPPSA and a refolding purification.



**Figure 4.3** SDS-PAGE gel of the time course expression trials of the hpPGK in *E. coli* strain Rosetta (DE3) cells. Lane1. Protein markers (life technologies). Lanes 2, 4, 6 and 8. 400 $\mu$ M IPTG post-induction cell lysates taken after 2, 3, 4 hours and overnight at 37 °C respectively. Lanes 3, 5, 7 and 9. The same time points as lanes2, 4, 6 and 8 respectively except showing the insoluble fractions. Lanes 10 and 11. 150  $\mu$ M IPTG post-induction insoluble fraction and supernatant cell lysate respectively, for cells grown overnight at 20 °C. Lanes 12 and 13, without IPTG, whole cell lysate and soluble fraction respectively, for cells incubated overnight at 37 °C.

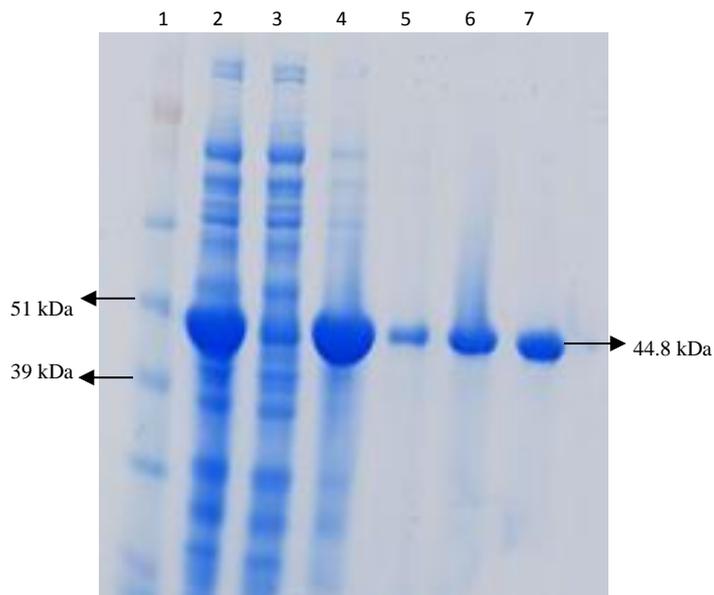


**Figure 4.4** SDS-PAGE gel of the time course of expression trials of hpPPSA in *E. coli* strain Rosetta (DE3) cells. Lane1. Protein markers (life technologies). Lanes 2, 4 and 6. 400 $\mu$ M IPTG post-induction soluble cell lysates taken after 3, 4 hours and overnight at 37 °C respectively. Lanes 3, 5, and 7. The insoluble fractions corresponding to lanes2, 4, and 6 respectively. Lanes8, 10 and 12. 50, 150 and 1000  $\mu$ M IPTG post-induction insoluble fraction and supernatant (soluble) cell lysate respectively, grown overnight at 20 °C. Lanes 9, 11 and 13 as with 8, 10 and 12 but for whole cells pellet fractions. Lanes 14 and 15, without IPTG, soluble fraction lysate and whole cells fraction respectively, grown overnight at 37 °C.

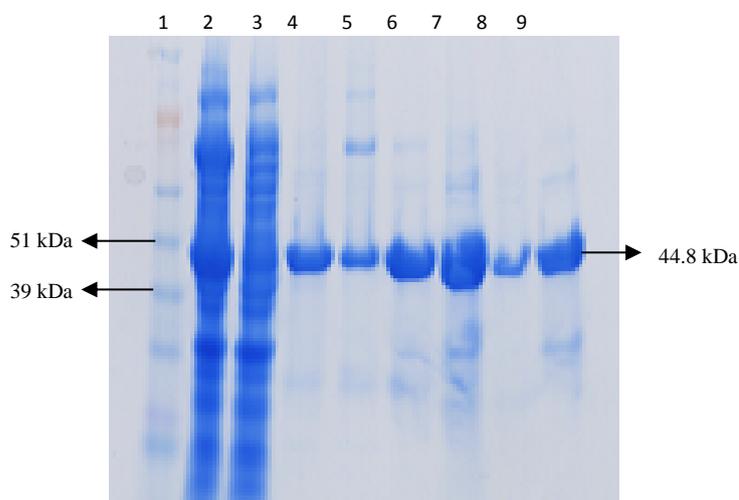
### 4.3.2 Purification of hpPGK

The soluble fractions of the *H. pylori* cell lysates (induced with 150 $\mu$ M IPTG) were subjected to small scale affinity chromatography (described in section 2.6.3) using nickel-charged nitriloacetic acid (Ni-NTA) resin. Figure 4.5 shows protein gel electrophoresis analysis of the samples taken from the purification steps and verifies the binding of His<sub>6</sub>PGK to the affinity resin. This result excludes the possibility of the over expression of untagged other native proteins.

These results show convincing evidence that an *E. coli*-based expression system can be used to produce hpPGK, and *E. coli* Rosetta (DE3) represents a suitable host for overexpression of hpPGK. Therefore, pLEICS-01 plasmid containing the ORF of hpPGK is used in this study for expression in *E. coli* Rosetta (DE3). The presence of N-terminal hexa histidine (N-His<sub>6</sub>) in hpPGK provides potential for single step protein purification by affinity chromatography using the Ni-NTA resin. However, Figure 4.6 shows the purification of hpPGK by requires two of steps Nickel affinity chromatography, the second binding step is important to enhance purity and to separate the cleaved tag hpPGK (flow through) from the uncleaved N-His<sub>6</sub>-hpPGK (elution). To verify the identity of the purified proteins, before kinetic and crystallographic investigation was undertaken, peptide mass fingerprinting was conducted (section 4.3.3).



**Figure 4.5** +/- IPTG SDS-PAGE gels of the small scale protein purification of the target protein by Ni-NTA affinity chromatography. Gel Samples taken from purification steps of hpPGK over expressed in *E. coli* strain Rosetta (DE3) with 150  $\mu$ M IPTG induction and. Lane 1. Protein markers (Life Technologies). Lane 2. Soluble cell lysate. Lane 3. Flow through fraction. Lane 4. hpPGK-Ni resin binding complex. Lane 5. hpPGK-Ni resin after washing. Lane6. hpPGK-Ni resin after elution. Lane 7. hpPGK elution fraction.



**Figure 4.6** SDS- PAGE gel of samples taken from protein purification of hpPGK using affinity chromatography by Ni-NTA resin. Lane 1. Protein markers (Life Technology). Lane 2. Soluble cells lysate. Lane 3. Flow through fraction. Lane4. hpPGK-Ni resin binding complex after washing. Lane 5. hpPGK-Ni resin after elution. Lane 6. hpPGK elution fraction before dialysis. Lane7. hpPGK Elution fraction after dialysis. Lane 8. hpPGK flow through after 2<sup>nd</sup> Ni resin binding (His<sub>6</sub> tag cleaved). Lane 9. hpPGK elution fraction from 2<sup>nd</sup> Ni resin binding after dialysis (His<sub>6</sub> tag not cleaved).

### 4.3.3 Peptide mass fingerprinting of hpPGK and hpPPSA

Trypsinogenised peptide mass fingerprinting from an SDS-PAGE gel containing samples of overexpressed hpPGK and purified tag-cleaved hpPGK and uncleaved N-His<sub>6</sub> hpPGK was sent to the PNAFL lab for identification by peptide mass fingerprinting. As seen in figures in the appendix regarding to these forms of hpPGK respectively, a high proportion of matched theoretical and practical peptides were obtained. These results unambiguously confirm the identity of the cleaved and uncleaved tag enzyme forms respectively, which verifies the enzyme identity prior to the characterization and crystallization work.

Another SDS-PAGE gel sample of overexpressed N-GST-hpPPSA was checked by peptide mass fingerprinting. Although the results demonstrated low rate of matching between the theoretical and practical peptide masses after trypsin digestion (figure is shown in the appendix). Nonetheless, this proportion was enough to verify the identity of the overexpressed hpN-GST-PPSA as the probability of obtaining these matches between sequences is quite low. This result allows using the optimum over expression conditions to investigate a refolding protein purification strategy of GST tagged hpPPSA.

A refolding purification trial was carried out using the rapid CST inclusion body solubilisation and renaturation kit (Cell Biolabs-USA) and affinity chromatography by Glutathione (GST) HiCap Matrix (Qiagen). The yield of purified refolded protein concentration was significantly low and/or with significant amount of contaminant proteins impurities. Accordingly, refolded purified hpPPSA was not suitable for kinetics and/or structural (by X-Ray crystallography) investigations.

Further methods need to be investigated, such as untagged conventional refolding protein purification strategies because SDS-PAGE analysis of hpPPSA after overexpression in *E. coli Rosetta* (DE3) revealed that the bacteria produced hpPPSA as potential tag-truncated and inclusion bodies aggregated protein. Another strategy might be to identify individual domains that fold properly and or truncate predicted disordered regions, requiring the cloning of new constructs into expression systems, with or without affinity tags for purification. Success would mean crystallisation trials

could be pursued for structure determination by X-Ray crystallography. As the time for the project was ran out, the following work on the structural investigation of hpPPSA focused on the prediction of the structure by bioinformatics techniques of sequence alignment and homology modelling.

#### 4.3.4 Size exclusion chromatography (SEC) of hpPGK

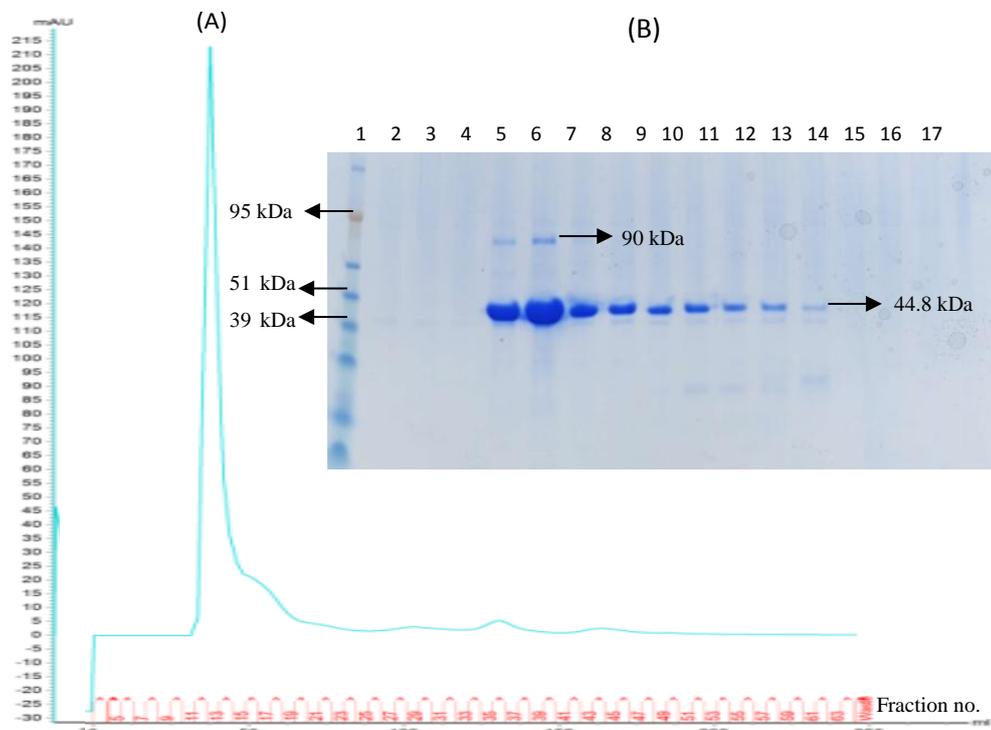
For oligomerisation state investigation, hpPGK was overexpressed and purified as described above. The enzyme contained N-terminal hexa Histidine tag and was purified using single binding step Ni-NTA affinity chromatography (Section 2.3.4). After dialysis the enzyme was concentrated and subjected to SEC (as described in Section 2.6.7).

A single peak was seen in size exclusion chromatography elution fractions, the peak has a shoulder (on the right) which may indicate impurities and/or degraded hpPGK (Figure 4.7 A). However, SDS-PAGE analysis of the fractions corresponded to this shoulder showed that these impurities were minor (Figure 4.7 B). The results verified the dimeric oligomeric state organization of hpPGK (with significant level of purity) as the peak corresponds to the area between IgG and BSA peaks in the standard protein calibration graph. These points correspond to the molecular masses of 160 kDa and 67 kDa respectively (Figure 2.1). The apparent molecular size of the hpPGK peak, is around 90 kDa which is a double of the molecular weight of a monomer of hpPGK. SDS-PAGE analysis of hpPGK samples taken from fractions of this peak revealed that the enzyme had been purified to near homogeneity with a molecular mass of approximately 45 kDa (for the monomer).

However, SDS-PAGE showed another protein band within the samples that represent the highest protein concentration fractions of the peak, thus unlike the monomeric protein bands of hpPGK on the gel these bands appeared to be as homodimer with similar expected size determined by SEC (90 kDa). The proper explanation behind that can be the incomplete denaturation of these hpPGK samples which resulted in a combination of monomeric and dimeric hpPGKs (Figure 4.7 B).

This supports the conclusion that hpPGK is a homodimer in solution unlike many bacterial PGKs (such as *E. coli*) which show a monomeric state. However, PGK of

*Corynebacterium glutamicum* was found to be active as a homodimer with a molecular weight of 104 kDa. Furthermore, dimeric PGKs reported from *Methanothermus fervidus* and *Pyrococcus woesei*, and tetrameric PGKs from *Sulfolobus solfataricus* and *Trypanosoma brucei* revealed multimerisation states (Reddy and Wendisch, 2014).



**Figure 4.7** Separation of hpPGK by Size Exclusion Chromatography (SEC) using Hi prep 16/60 Sephacryl S-200 HR column. The column was loaded with 4-5 ml of 10 mg/ml of hpPGK purified by Ni resin affinity chromatography. The experiment carried out using 50 mM HEPES pH 7.5, 1 mM DTT and 0.2 M NaCl, with a flow rate of 0.8 ml/min (24 cm/h). Panel (A) Chromatogram of hpPGK SEC. Protein concentration was detected by measuring the absorption at 280 nm. The observed peak has a shoulder (on the right) which may comprise impurities (significantly low according to SDS-PAGE analysis). The peak fractions containing pure enzyme were pooled, concentrated and flash frozen in liquid nitrogen and then stored at -80°C. Panel (B) SDS-PAGE gel of the peak fractions (11-19). Lanes contain: 1. Protein markers (life technologies), 5-13. The fractions of the active peak.

#### 4.4 Characterization of hpPGK

Enzymatic characterisation of hpPGK was conducted by identification of kinetic parameters ( $K_m$ ,  $V_{max}$  and  $k_{cat}$ ), and investigating the folding and stability of the protein prior to X-Ray structural study. The effect of hexa Histidine tag on hpPGK activity was also investigated by kinetic characterisation.

#### 4.4.1 Enzymatic analysis of hpPGK

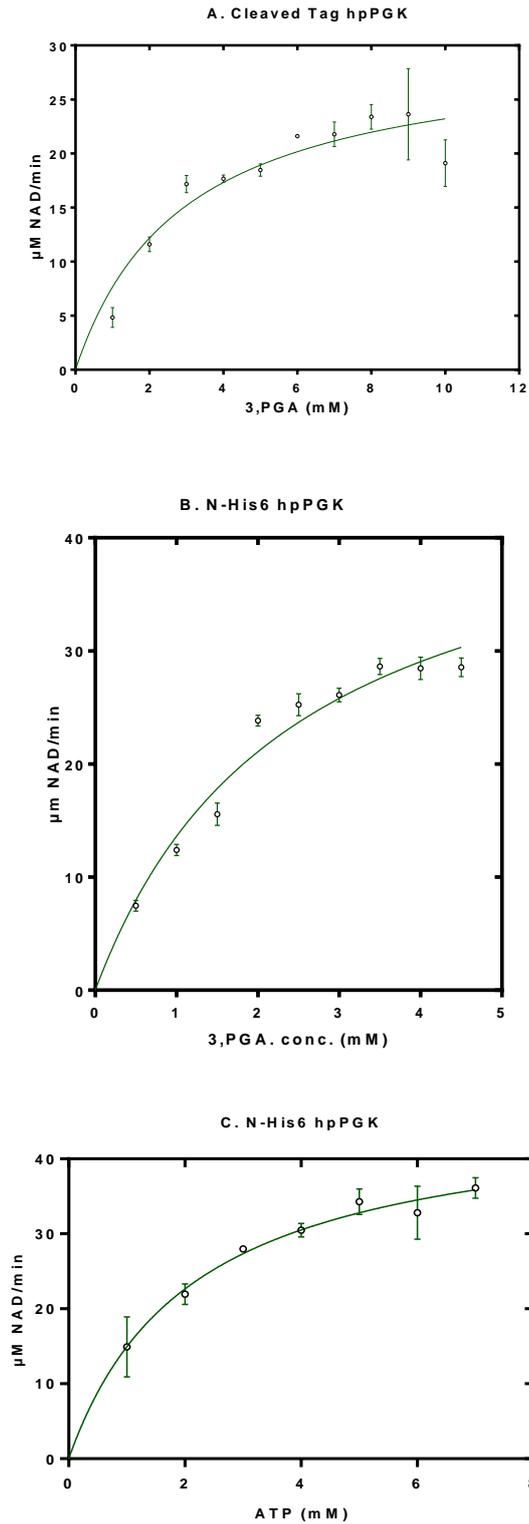
To measure hpPGK activity, an *in vitro* spectrophotometric coupled enzyme assay was used. The rate of the enzymatic reaction (the extent of product formation) was estimated as a function of time for a series of substrates concentrations (3,PGA and ATP).

Initially, the activity of the purified cleaved tag hpPGK and His<sub>6</sub>hpPGK were checked under the same assay conditions described in Section 2.7.1 except using various concentrations of 3, phosphoglycerate (3, PGA) as the rate limiting factor. After that the 3, PGA substrate saturation curve of the purified cleaved tag hpPGK and the other tagged hpPGK were plotted as shown in Figure 4.8 panels A and B respectively. The results showed that both of these enzyme forms follow Michaelis-Menten kinetics. This allowed an estimate of the kinetic parameters  $K_m$ ,  $V_{max}$  and  $k_{cat}$  as shown in Table 4.1. The  $K_m$  values in the Table are comparable between the cleaved tag hpPGK and His<sub>6</sub>hpPGK forms, i.e. the substrate affinity of these enzyme preparations are similar. Accordingly, it can be concluded that the hexa histidine tag does not affect the enzyme activity and histidine tag purified hpPGK is appropriate for crystallisation trials.

Finally, the same strategy was followed to determine  $K_m$ ,  $V_{max}$  and  $k_{cat}$  of N-His<sub>6</sub> hpPGK using different concentrations of ATP as shown in Figure 4.8 (panel C) and Table 4.1. The data emphasize that hpPGK follows Michaelis-Menten kinetics, and the measured kinetic parameters verify the activity of this enzyme.

**Table 4.1** Measured kinetics parameters of cleaved tag hpPGK and N-His<sub>6</sub>hpPGK with varying concentrations of 3,PGA and ATP respectively using methods described in Section 2.7.1.

Enzyme form	Rate limiting factor	$K_m$ (mM)	$V_{max}$ ( $\mu\text{M}/\text{min}$ )	$k_{cat}$ ( $\text{min}^{-1}$ )
Cleaved tag hpPGK	3,PGA	2.93±0.75	30.01±2.65	300.01±26.51
N-His <sub>6</sub> hpPGK	3,PGA	2.37±0.36	47.02±3.33	470.02±33.32
N-His <sub>6</sub> hpPGK	ATP	2.134±0.36	46.80±2.78	468.0±27.8



**Figure 4. 8** Enzymatic analysis of hpPGK. Panel (A) and (B) show the plot of initial rate for varying substrate concentration (3,PGA) at a fixed 2 mM concentration of ATP. Panel (C) shows the rate for varied concentrations of ATP at a fixed 2.5 mM concentration of 3,PGA. The rest conditions and components of the enzymatic assay were conducted as described in Section 2.7.1. The data in the figure carried out in 3 repeat experiments and the error bars indicate the standard deviation of the results.

## 4.4.2 Far ultraviolet (UV) Circular Dichroism (CD) spectrum of hpPGK

Circular Dichroism can be described as a technique used to estimate the secondary and tertiary structures of proteins by means of the wavelength dependent rotation of circularly polarized light by chiral protein molecules. The measurement units of CD signals are millidegrees (mdeg). CD can be calculated by subtracting absorbance differences between Right handed (R) (anti-clockwise) and Left handed (L) (clockwise) rotated circularly polarized light. Secondary and tertiary protein structure can be assessed by absorption determination in a wavelength range 160-250 and 260-320 nm for the analysis of secondary and tertiary protein structure respectively.

Prior to proceeding to crystallisation experiments the stability and folding of the enzyme was investigated. The measurements of protein secondary structure verify that N-His<sub>6</sub>hpPGK is a folded protein with  $\alpha$ -helical dominance. The CD spectrum acquired from the measurements of N-His<sub>6</sub>hpPGK indicates that there are two major peaks. The first one is positive at 190-200 nm, the other is negative peak at 208-222 nm (Figure 4.9). This is a main characteristic of  $\alpha$ -helical dominant protein structure.

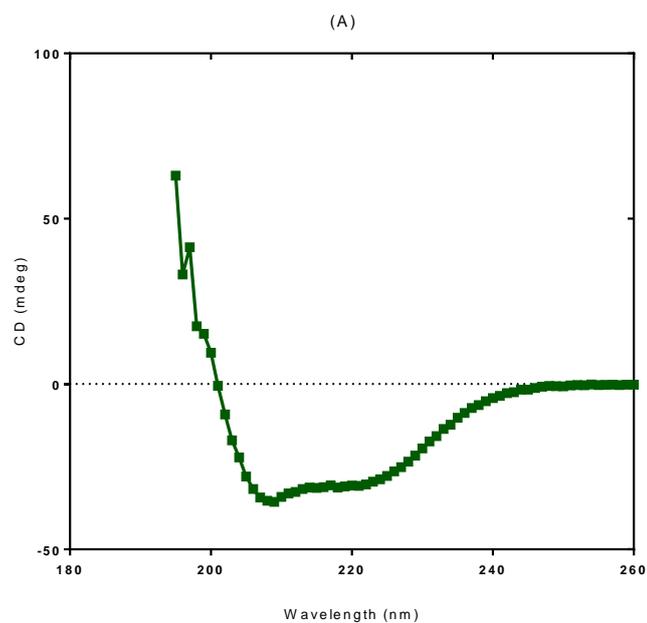
Following the thermal stability of N-His<sub>6</sub>hpPGK by measuring the CD at 220 nm showed that the enzyme is stable until about 45°C. Above this N-His<sub>6</sub>hpPGK starts melting to reach midpoint at 50°C, and the enzyme is completely melted at 60°C (Figure 4.10).

## 4.5 X-Ray crystallography of apo- N-His<sub>6</sub>- hpPGK

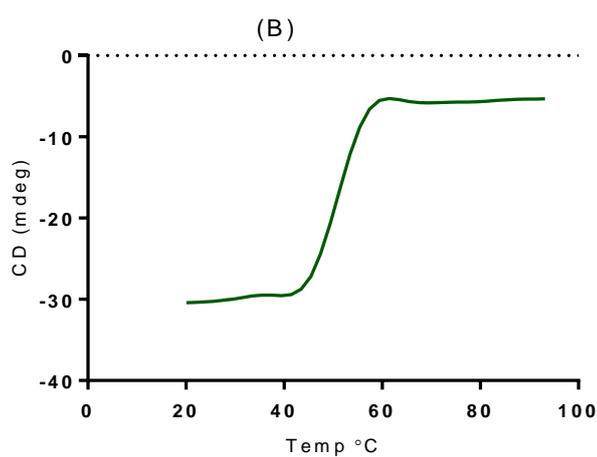
### 4.5.1 Crystallization of hpPGK

With the aim of determining the three dimensional structure of apo and liganded hpPGK, sitting drop crystallization trials were performed using the Mosquito crystallization robot (TTP Labtech), using several commercial available screens including JCSG+, MORPHEUS, MIDAS, NR-LBD, PACT premier, WIZARD and AmSO<sub>4</sub> Suite (Molecular Dimensions Ltd). Initially formed crystals did not produce good quality diffraction patterns. Therefore, manual screening using 24 well plates of conditions around the promising results was employed to improve these. This strategy includes

changing protein or precipitant concentration, crystallisation under a variety of pH values and different equilibration temperatures. In addition to these, 10% additive screens with micro-seeding and macro-seeding crystallisation techniques were followed.

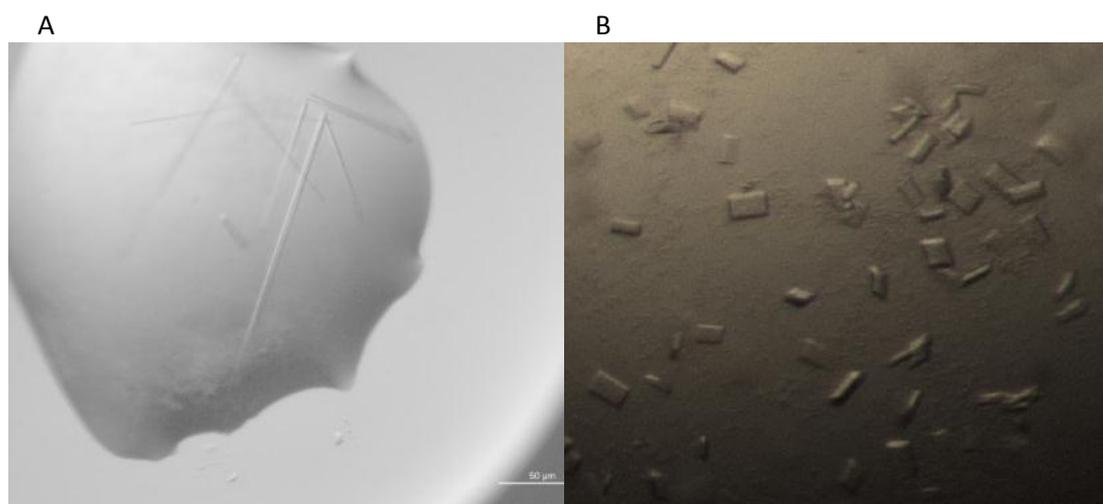


**Figure 4.9** Far-UV CD spectrum of apo N-His<sub>6</sub>hpPGK diluted to 0.3 mg/ml by 50 mM HEPES pH7.5, 200 mM NaF and 1 mM DTT.



**Figure 4.10** Far-UV CD spectrum of apo N-His<sub>6</sub>hpPGK at 222 nm with heat treatment using temperature range of 20-90°C.

A number of crystals were obtained under a variety of conditions and following different crystallisation techniques. Unfortunately, X-ray diffraction quality from any of these crystals was not sufficient to solve the structure of either apo or liganded hpPGK. However, 0.1M Bis-Tris pH6.5 and 28-33% of PEG MME 2000 (Cliff et al., 2010), resulted in needle crystals of apo-hpPGK (Figure 4.11 panel A) These gave low quality x-ray diffraction data when tested using IO4 beamline (Diamond Light Source). Subsequent improvement strategies were attempted including micro-seeding varying pH, precipitant or protein concentration and equilibration temperature and 10% additive screen. The best optimisation was achieved by adding 10% additives of JCSG+ screen which improved the crystals shape significantly (Figure 4.11 panel B). These 30-50  $\mu\text{m}$  length crystals were sent to the synchrotron beam (Diamond Light Source) for x-ray data collection.



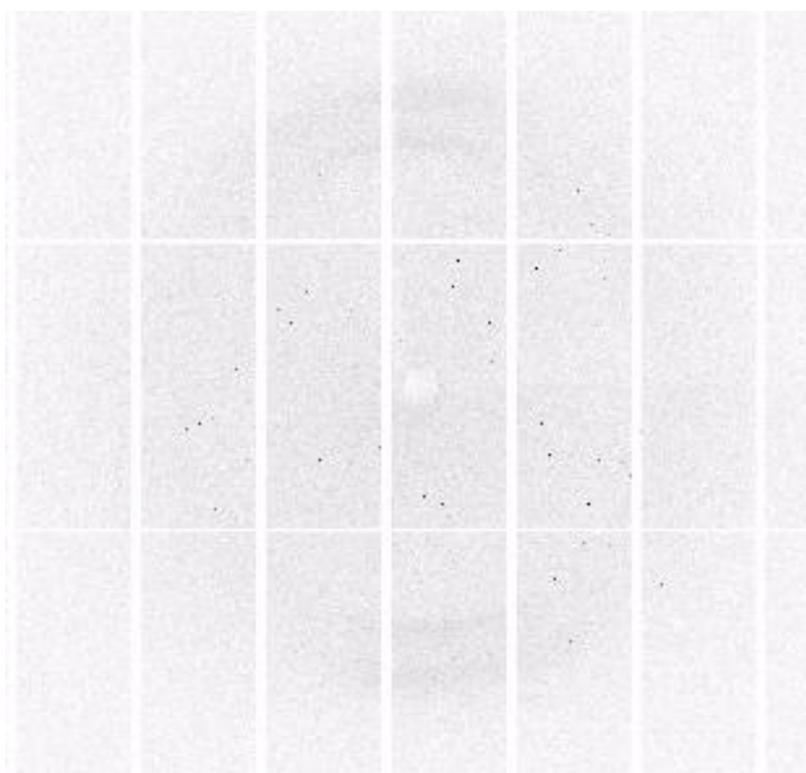
**Figure 4.11** Apo hpPGK crystals (13.5mg/ml) using the following crystallization conditions, (A) 0.1 M Bis-Tris pH6.5 and 30%PEGMME2000, the scale on the right corner is 50 $\mu\text{m}$  (B) The same conditions in A mixed with 10% of H2, in JCSG+ screen, which consists of 1 M Ammonium sulphate, 0.1 M Bis-Tris Ph5.5 and 1% PEG 3350. The crystals were grown for 1-2 days at 22°C and the results were long needles and plate shape crystals with around 30-50 $\mu\text{m}$  dimensions in A and B respectively.

#### 4.5.2 Data collection and processing of apo-hpPGK

As described in Section 2.5.2, crystals of apo-hpPGK are cryoprotected with 20-25% PEG 400 and stored in liquid nitrogen prior to sending to the Diamond Light Source synchrotron (DLS) (Oxford-UK) for data collection. Diffraction data were collected at the

microfocus beam line I24. A single cryo-cooled crystal (0.05x0.05 mm) was used to collect data using 0.97 Å wavelength and Pilatus3 6M detector with 300 mm crystal to detector distance. 1800 images were collected with 0.1° oscillation width with exposure time of 0.020 second. Figure 4.12 shows one of the diffraction images of apo-hpPGK.

The DLS auto-processing pipeline gave 2.66 Å diffraction data, which was sufficient for structure determination. The FastDP pipeline uses XDS, XSCALE, AIMLESS and POINTLESS respectively for indexing, merging and space group determination. The data collection and processing statistics are shown in Table 4.2.



**Figure 4.12** Diffraction image of apo-hpPGK with resolution of 2.66 Å. 1800 images were collected from a single cryo cooled crystal at 0.97 Å wavelength with 0.1 oscillation width. In the image it is possible to see the diffraction spots on the detector screen from the different reciprocal planes. For the maximum resolution limit, 2.0 Å were initially selected. However, the resolution limits of the outer shell are located before this namely 2.66 Å.

**Table 4.2** Data collection statistics of apo-hpPGK. Values for the highest resolution shell are shown in parentheses.

	Apo-hpPGK
Space group	I222
Unit cell dimensions(Å)	77.20 158.82 175.43
Resolution (Å)	29.63-2.66 (2.66-2.73)
Completeness (%)	98.8 (84.6)
†Rmerge	0.13 (0.78)
I/σ	11.5 (2.0)
Number of observation	206104 (11491)
Number of unique observation	31148 (1924)
Redundancy	6.6 (6.0)

†Rmerge =  $\sum_{hkl} [\sum_i |I_{hkl,i} - \langle I_{hkl} \rangle|] / \sum_{hkl,i} \langle I_{hkl} \rangle$ , where  $I_{hkl,i}$  is the intensity of an individual measurement of the reflection with Miller indices  $h$ ,  $k$ , and  $l$  and  $\langle I_{hkl} \rangle$  is the mean intensity of that reflection.

### 4.5.3 Structure determination of apo-hpPGK

The scaled and merged data output is the set of unique reflections and their amplitudes (F), stored in the mtz format, along with other information (e.g.  $\sigma(F)$ ,  $I$ ,  $\sigma(I)$  etc.). This was used for molecular replacement calculations. Molecular replacement requires, along with diffraction data, a search model and an estimate of the number of protein molecules to be found. The coordinates of the *Campylobacter jejuni* (PDB 3Q3V) apo PGK was used as a search model and the number of molecules to be found in each asymmetric unit (2) was determined by the Matthews program in the CCP4 6.5.0 suite. Molecular replacement was carried out using Phaser MR (in the Phenix suite). The

resultant phases from the derived coordinates with two monomers per asymmetric unit gave an interpretable map.

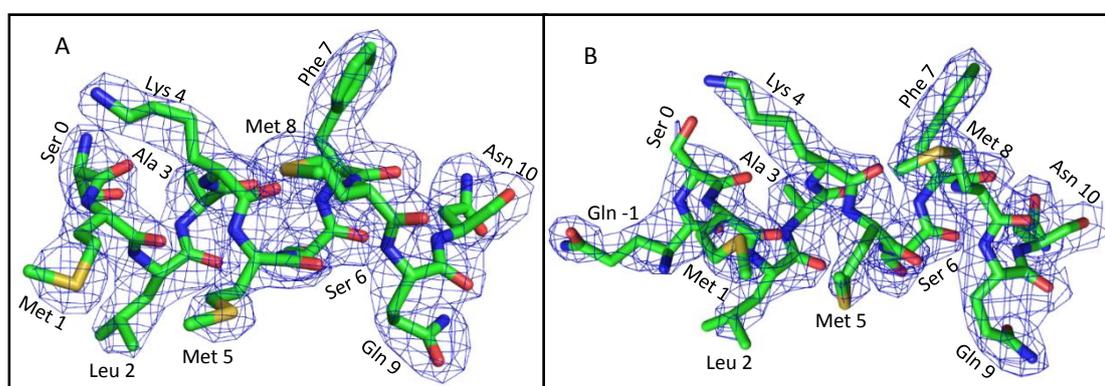
The preliminary model and phases were used as a basis for refinement, initially using the AutoBuild facility in Phenix (Adams et al., 2010) to obtain the best interpretation with the electron density map. The output map and coordinates were checked in WinCoot 0.8.1.1 (Emsley et al., 2010), and also compared to the output from Phaser. However, some regions in the structure of apo-N-His<sub>6</sub>-hpPGK showed poor correlation with the electron density maps. Accordingly, further refinement and manual rebuilding cycles were required to improve the model.

This process iteratively used Phenix refinement cycles, with the output map and coordinate files examined and adjusted with Coot. Eventually 395 residues out of 402 were modelled in each monomer of the dimer with Asp35, Glu36, Asn37, Leu 38 and Asn 39 unmodelled because of poor density, presumably because as they are part of a flexible loop. No electron density was identified for the C-terminal Arg401 and His402 in each molecule of hpPGK dimer.

Some interpretable electron density upstream of the N-terminal methionine residue was identified. The six histidine tag is linked to the N-terminal methionine by a peptide, resulting in the N-terminal addition of HHHHHHSSGVDLGTEENLYFQS. Accordingly, residues serine (Ser 0) and glutamine (Gln -1) were added to chain A and serine (Ser 0) was added to chain B using Coot to fit to the electron density. Subsequently, two cycles of restrained refinement was carried out with *REFMAC5* (in the CCP4 suite) (Murshudov et al., 2011). Final  $R_{work}$  of the model was 0.19 and  $R_{free}$  was 0.26 (shown in Table 4.3 which lists the refinement statistics). Figure 4.13 shows the refined N-terminal His tag linker residues in chain A and B of apo-hpPGK dimer.

**Table 4.3** Final refinement statistics of apo-hpPGK

Refinement	
Resolution (Å)	29.63-2.66
No. reflections	31558
Rwork/Rfree	0.19 /0.26
No. atoms	
Protein	6509
Solvent	52
Ligand (sulphates)	4
R.m.s. deviations	
Bond lengths (Å)	0.0121
Bond angles (°)	1.57
Overall B-factor (Å)	43.7



**Figure 4.13** hpPGK N-terminal residues (shown with 2Fo-Fc electron density map (blue) contoured at  $1\sigma$ ). Panels A and B show chains A and B respectively. The residues 1-10 are shown in addition to the addition residues from the linker i.e. Ser 0 in chain A (panel A) and Gln -1 with Ser 0 in chain B (panel B). The figure was created with PyMOL 1.2 (DeLano, 2002).

#### 4.5.4 Overall structure of apo hpPGK

The I222 crystal structure of hpPGK contains two molecules (chain A and chain B) per asymmetric unit (figure 4.14). The final model has 52 water molecules, four ligand sulphate molecules and two chloride ions (observed at non physiological locations), in addition to some unmodelled density which might be from the crystallisation medium or the cryoprotectant such as Bis-Tris, PEG MME 2000 or PEG 400.

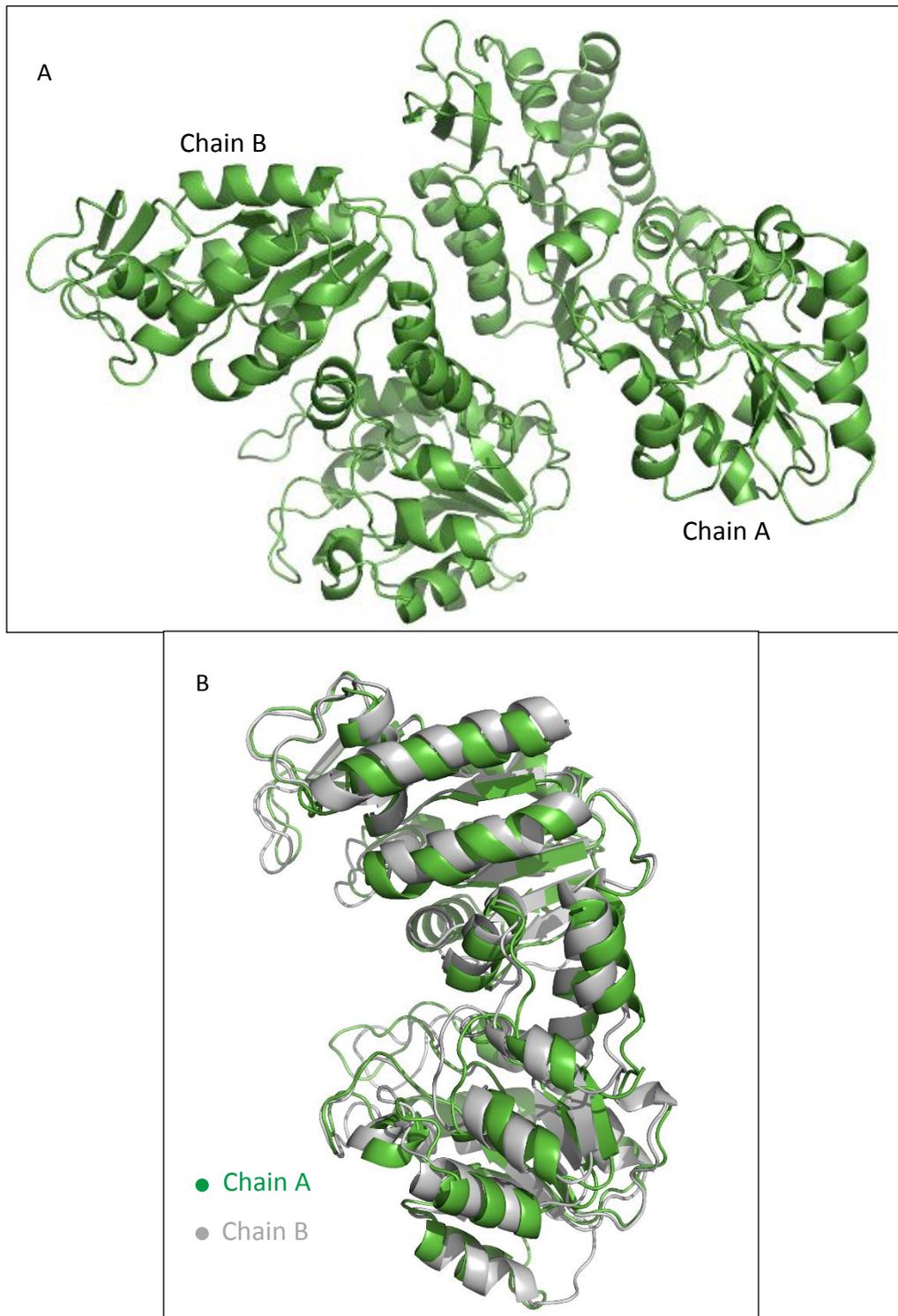
According to size exclusion chromatography analysis (section 4.3.4), the active biological unit of hpPGK is most likely found to be arranged as a dimer. This result was supported by superposition results of hpPGK monomers (chain A and chain B) as they showed very similar overall structure and homologues alignment (figure 4.14). Furthermore, interface analysis using PISA web application for hpPGK revealed interface with buried surface area of 690 and 678 Å<sup>2</sup> for chain A and chain B respectively. PISA results indicated that the interfacing residues are making twelve hydrogen bonds and two salt bridges.

An overview of the solved structure of a hpPGK monomer seen in its open conformation is shown in figure 4.15 A. The structure is comprised of two domains. A typical Rossmann fold super-secondary structure is found in each domain which composed of a central  $\beta$ -sheet sandwiched by two  $\alpha$ -helical layers, consistent with its role as nucleotide-binding enzyme (figure 4.15 B). The N-terminal domain (pale green in the figure) is formed of a six parallel stranded central  $\beta$ -sheet surrounded by two  $\alpha$ -helices on one side and three  $\alpha$ -helices on the other. This domain is responsible for the binding with 3-phosphoglycerate (3-PGA) or 1,3-bisphosphoglycerate (1,3-BPG). The C-terminal domain comprises a central  $\beta$ -sheet with five major parallel strands with three minor antiparallel strands, the  $\beta$ -sheet is surrounded by three  $\alpha$ -helices in the front and four  $\alpha$ -helices at the back (figure 4.15 B). The C-terminal domain controls ATP or ADP binding (dark green in the figure). These domains are linked to each other by a helix known as the hinge region (Gly175-Phe190). The C-terminal domain ends with a short helix followed by a C-terminal loop (Leu392- Lys400). This loop is considered as a second hinge region as it comes into proximity with the N-terminal domain.

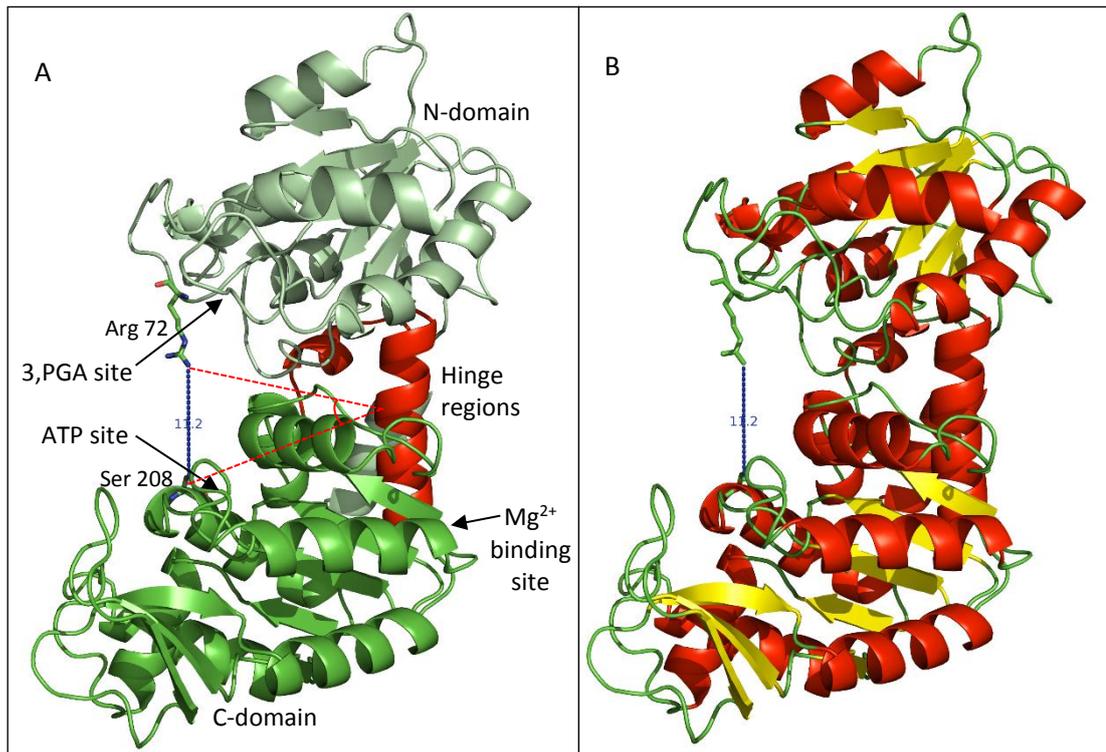
Generally, phosphoglycerate kinases adopt a fully “closed” conformation during catalysis when 3,PGA or 1,3BPG and nucleotide (ADP or ATP) are bound, whereas the resting “open” conformation is observed when either reactant is absent. A salt bridge was identified in the structure of the closed form of the human enzymes which is formed upon the transition from the resting state to the active state (completely closed conformation) (Auerbach et al., 1997). This bridge links a conserved Arg from the N-terminal domain (on top of the 3,PGA binding site) with an Asp from the C-terminal domain on top of the nucleotide binding site.

However, although the equivalent residue to the Asp residue is a conserved Asp in PGKs of other orthologues in *Campylobacter jejuni* it is Gly204, and in hpPGK it is Ser208 (figure 4.15). It is proposed the salt bridge creates a cover to fully close the cleft and form an active catalytic environment (Zheng et al., 2012). This could not be a salt bridge in the case in hpPGK, the link between Arg72 and Ser208 might be a hydrogen bond.

Nonetheless, the hpPGK adopts a typical “open” conformation with 11.2 Å (for chain A) and 15.28 Å (for chain B) distances between the side chains of the equivalent to “salt bridge” forming residues, that is NH<sub>2</sub> of Arg72 and OG of Ser208, which give angles of 31° and 41° for chains A (figure 4.15 A) and B respectively when triangulated to the middle of the hinge linker helix (C $\alpha$  of Ile183).



**Figure 4.14** Three dimensional structure of *H. pylori* 26695 apo-PGK shown as a cartoon representation. Panel A shows the dimer, (Panel B shows the superposition of hpPGK monomers (chain A in green, chain B in grey)).



**Figure 4.15** (Panel A) Crystal structure of *H. pylori* 26695 apo-hpPGK (chain A) showing the N-domain (pale green) (the binding site of 3-Phosphoglycerate or 1,3-Bisphosphoglycerate), hinge region (red) and C-domain (dark green) the nucleotide binding site (ATP or ADP) (indicated by arrows). The distance (11.2 Å) (shown in blue lines) between the side chains of Arg72 and Ser208 (shown in stick representation) that form a “salt bridge” upon closed conformation. (Panel B) Cartoon representation by secondary structure showing hpPGK structure with  $\alpha$ -helices coloured in red,  $\beta$ -sheets coloured in yellow and loops in green.

#### 4.5.5 ATP and 3,PGA binding sites of hpPGK using human PGK

The structure of open form apo-hpPGK, was compared to and superposed on the fully closed structure of human hpPGK (accession code 4AXX) in order to evaluate conformational changes and identify the residues responsible for the substrate and the nucleotide binding sites.

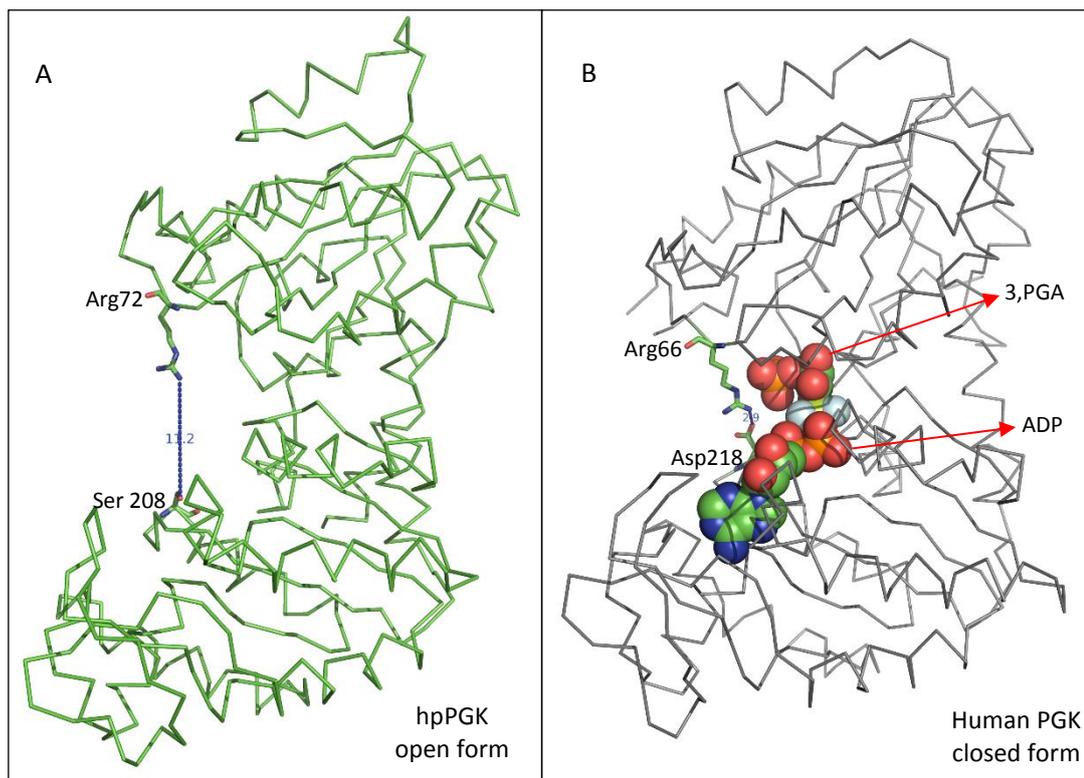
During the conformational change from fully opened to fully closed enzyme forms, the substrate 3,PGA binds to the N-terminal domain and is brought in close proximity, and finally contacts, the phosphate moiety from ATP embodying significant N-terminal domain rotational movement of around  $37^\circ$ . The degree of cleft closure can be assessed using the distance between the “salt bridge” forming residues (section 4.5.4). In typical closed conformation structures (such as human PGK) the distance between Arg65 and

Asp218 decreases to less than 5 to 6 Å. Upon domain closure, the real distance between these two residues will be nearer to 3 Å due to side chain movement accompany the salt bridge formation during catalysis (Zheng et al., 2012) (figure 4.16).

To identify and compare residues responsible for the substrate and nucleotide binding, 3,PGA (N-terminal) and ADP (C-terminal) domains of PGK were aligned separately to the corresponding domains of the fully closed human PGK (liganded with 3,PGA and ADP) (accession code 4AXX) The structure alignment achieved with root mean square value (RMSD) of 0.91 Å for 147 C-α atoms in the N-terminal domain and 1.1 Å for 165 C-α atom in the C-terminal domain.

Each domain of 3,PGA (or 1,3-BPG) and ADP (or ATP) binding sites of hpPGK and the superposed corresponding domains of human PGK were compared structurally. According to structure alignment analysis, the substrate binding residues in the cleft between N-terminal and C-terminal domains are highly conserved (figures 4.17 and 4.18). This conserved region corresponds well with the substrate binding site and the nucleotide binding sites in these domains respectively. The only variation in sequence was identified in the C-terminal domain where Ile300 of hpPGK replaces Gly313 in human PGK. However, the interaction of nitrogen number 6 (N-6) of adenosine in ADP is to the main chain of Gly313 rather than the side chain and thus would be unaffected.

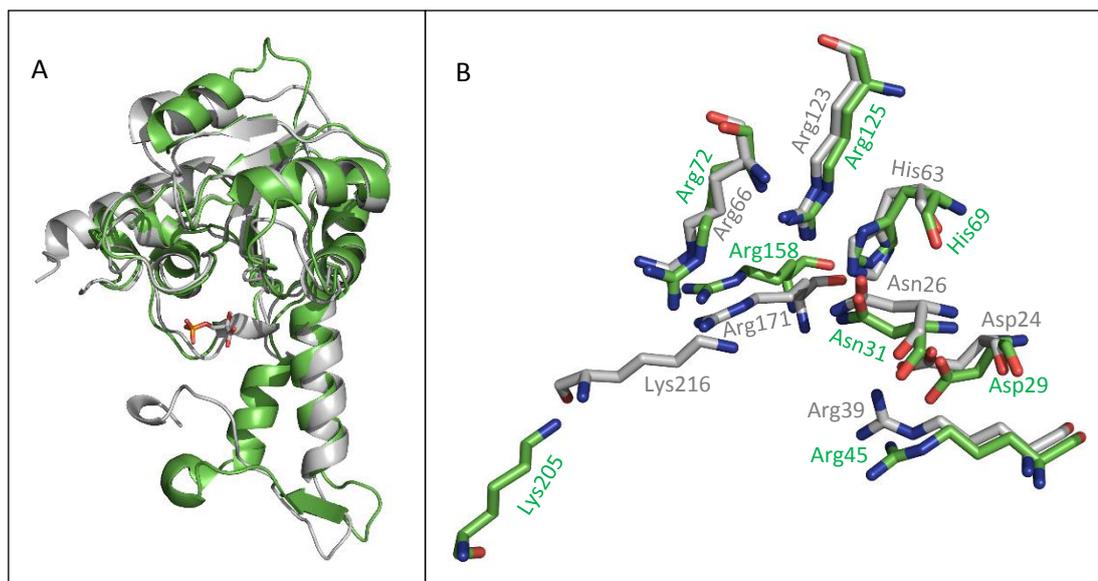
For additional confirmation, hpPGK residues identified as responsible for the substrate and nucleotide binding were checked by multiple sequence alignment approach using the MAFFT programme (Kato and Standley, 2013). The correspondence of these residues was analysed in order to compare them with the amino acid sequences of other PGK structures from human, *C. jejuni*, *Bacillus staerothermophilus* (*B. staerothermophilus*) and *Trypanosoma brucei* (*T. brucei*) deposited in the protein data bank (PDB) (figure 4.19). The sequence alignment analysis is consistent with the structural superposition of hpPGK with human PGK showing that the 3,PGA or 1,3-BPG and ATP or ADP binding residues are well conserved.



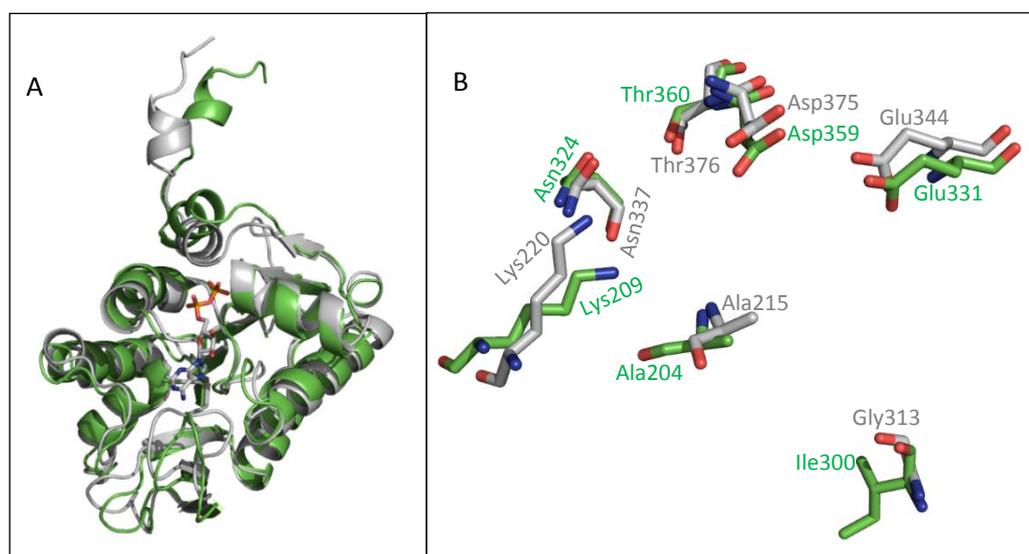
**Figure 4.16** Open and closed conformations of PGK (shown as  $C\alpha$  trace) from *H. pylori* and human (PDB accession 4AXX) respectively illustrating conformational changes of PGK during catalysis by mean of domain closure. (Panel A) The open form of hpPGK (green) shows the distance (11.2 Å) between the ‘‘salt bridge’’ forming residues (Arg72 and Ser208). (Panel B). The fully closed form of human liganded hpPGK (grey) illustrates the distance between the Arg65 and asp218 residues (2.9 Å). 3,PGA and ADP ligands are shown as CPK spheres.

#### 4.5.6 Sulphate ligands bound to hpPGK

The hpPGK structure solved in this study is similar to previously solved PGKs. The structure is an apo-form without any substrate. However, electron density was found consistent with sulphate molecules probably from the crystallisation condition (ammonium sulphate) (section 4.5.1). These molecules were modelled in both substrates binding sites, that is in the N-terminal and C-terminal domains (one molecule in chain A and three molecules in chain B) (figure 4.20). The results showed clear densities are seen for sulphate ions (Figure 4.21 A). Two sulphate molecules (number three and four) are bound similarly to each N-terminal domain of the hpPGK monomers.



**Figure 4.17** (Panel A) N-terminal domains superposition (on C $\alpha$ ) of apo-hpPGK (green) with 3,PGA liganded human PGK (grey). The superposition was carried out to compare the substrate binding residues with their homologues of hpPGK. (Panel B) shows the aligned positions of residues responsible for 3,PGA binding site of human PGK (grey) with their equivalent residues in hypothetical hpPGK (green).



**Figure 4.18** (Panel A) C-terminal domain alignment based on C $\alpha$  of open form of hpPGK (green) with closed form of human PGK liganded with ADP nucleotide (grey). The nucleotide binding residues in the human metabolic enzyme are compared with their equivalent residues in apo-hpPGK. (Panel B) shows detail of superposed residues in the ADP binding site of the mammalian PGK (grey) with their equivalent residues in hpPGK.

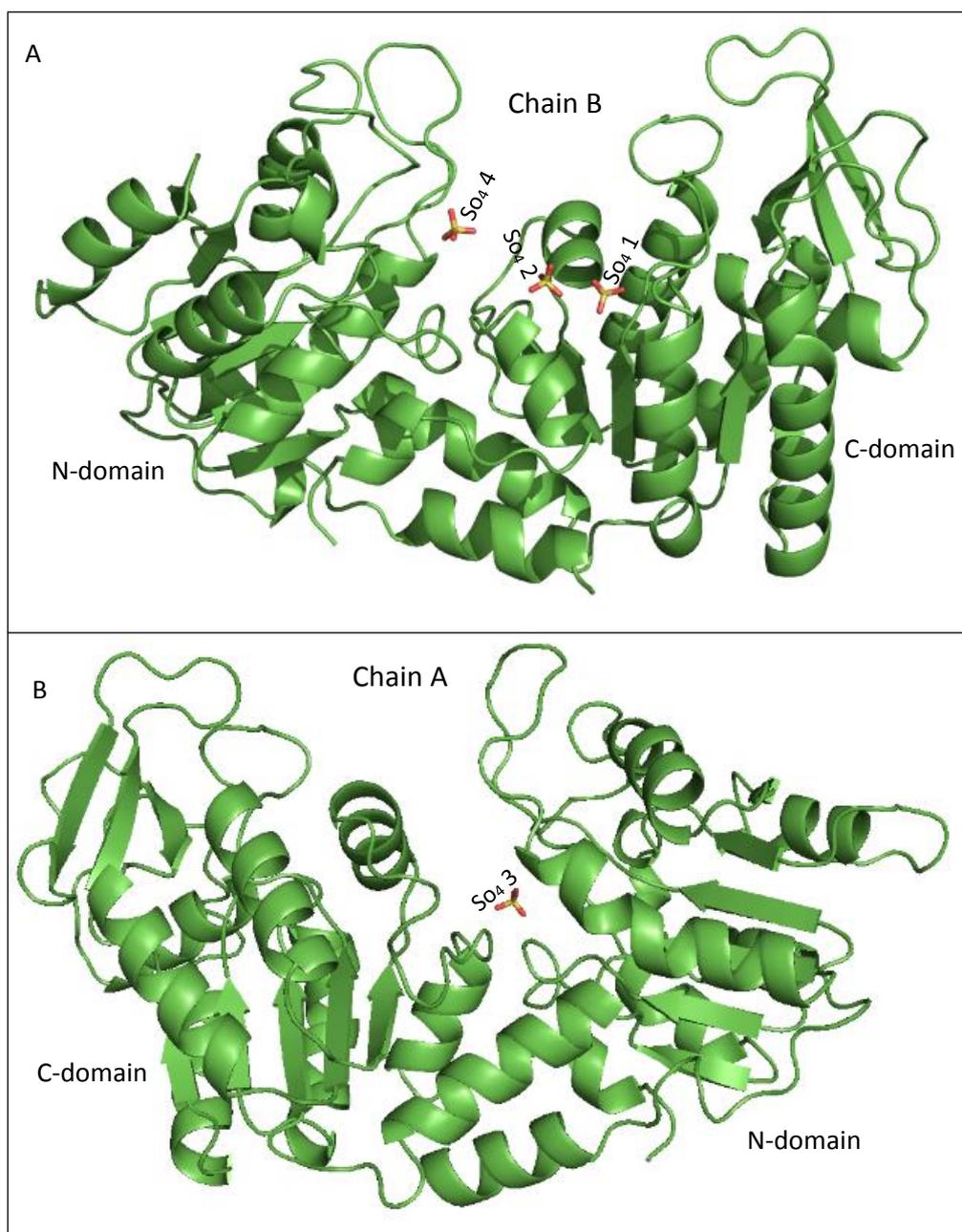
H. pylori	MLAKMSFMQNVKNIQEVVSHKRVLIRVDFNVPLDENLNITDDTRIRESLEPTIQYCIDNK
C. jejuni	SNA----XSDIISIKDIDLAKKKVFIRCDFNVQDDFLNITDDRRIRSAIPTIRYCLDNG
H. sapiens	MS-----LSNKLITLTKLVDKGRVVMRVDFNVPMKNN-QITNNQRIKAAVPSIKFCLDNG
T. maritima	-----EKMTIRDVDLKGKRVIMRVDFNVVPKDG-VVQDDTRIRAALPTIKYALEQG
B. staerothermop	M-----NKKTIRDVDVGRKRVFCRVDFNVPMEQG-AITDDTRIRAALPTIRYLLIEHG
T. brucei	-----EKKSINECDLKGKVLIRVDFNVVPKNG-KITNDYRIRSAIPLTKKVLTEG
	: : . : : * : * * * * * : : : : * : : * : : : : .
H. pylori	AKDIILVSHLGRPKGVVE-----KLSLKPFLKRLERLLNHEVVFQSQNI
C. jejuni	CS-VILASHLGRPKKEISS-----KYSLEPVAKRLARLLDKETVAKADV
H. sapiens	AKSVVIMSHLGRPDGVMPD-----KYSLEPVAVELKSLGKDVFLKDC
T. maritima	AK-VILLSHLGRPKGEPSP-----EFSLAPVAKRSELLGKQVFPVAV
B. staerothermop	AK-VILASHLGRPKGVVE-----ELRLDAVAKRGLGELLERPVAKTNEA
T. brucei	GS-CVIMSHLGRPKGIQMAQAGKIRSTGGVPGFQQKATLKPVAKRSELLLRPVTFAPDC
	. : * * * * * . : * . . * * * : :
H. pylori	V--QLKQALNENAPTRIFLLENIRFLRGEE-----ENDENLAKD-----LASLCD
C. jejuni	IGEDAKTKAXNLKAGEILLLENLRFKGET-----KNDENLAKE-----LASXVQ
H. sapiens	VGPEVEKACANPAAGSVILLENLRFHVEEKGKAKDASGNKVKAEPAKIEAFRASLSKLGD
T. maritima	VGDEVKKAWEELKEGEVLLLENLRFHPGET-----KNDPELAKF-----WASLAD
B. staerothermop	VGDEVKAAVDRLNEGDVLLLENLRFYFPGEE-----KNDPELAKA-----FAELAD
T. brucei	L--NAADVVSXKMSPGDVVLLLENLRFYKEEGSKK-----AKDREMAKI-----LASYGD
	: : . : . : . * * * * * * : : : : * * : : . : :
H. pylori	VFVNDAFGTSHRKHASTYGTAKFAPI--KVSGLLKKKEIDSFYQANHPLRPLLLIVGGA
C. jejuni	VYINDAFGVCHRAHSSVEAITKFFDEKHKGAGFLQKKEIDFASNLIKHPARPFVAVVGGG
H. sapiens	VYVNDAFGTAHRAHSSMVGVN--LPQ--KAGFLMKKEIDNYFAKALSPERPFLLAILGGA
T. maritima	IHVNDAFGTAHRAHASNVGIAQFIP---SVAGFLMEKEIKFLSKVLYNPEKPYVVVLGGA
B. staerothermop	LYVNDAFGAAHRAHASTEGIAHYLP---AVAGFLMEKELEVLGKALSNPDRPFTAIIGGA
T. brucei	VYISDAFGTAHRDSATMTGIPKILGN--GAGFLMEKEISYFAKVLGNPPRPLVAIVGGA
	: : . * * * * * . * : : : * * : : : * * : : * * : : * * : : * * :
H. pylori	KVSSILTLKLNILDLDKLIITAGAMSNTFLKAL-GYDVQDSSVEDALINDALELLQSAKE
C. jejuni	KVSGILQALTNLKPKVDKLIIGGGXAFTFLKAL-GYDIGNSLLEELLEEANKILTKGN
H. sapiens	KVADKIQLINMLDKVNEMIIGGMAFTFLKVLNMEIGTSLFDEEGAKIVKDLMSKAEK
T. maritima	KVSDIIGVITNLMEKADRILIGGAMMFTFLKAL-GKEVGSRRVEDKIDLAKELVEKAKE
B. staerothermop	KVKDKIGVIDNLLEKVDNLIIGGLAYTFVKAL-GHDVGSLLLEEDKIELAKSFMEKAKE
T. brucei	KVSDKIQLLDNMLQRIDYLLIGGAMAYTFLKAQ-GYSIGKSKCEESKLEFARSLKKAED
	** . * : : * : : : : * * . * * : : . : * : : . : : . : . : .
H. pylori	KKVKVYLPIDAVTTDDIILNPKHIKISPV-QDIEPKHKLADIGPASLKLFEVIESAPTIIL
C. jejuni	LGVKIYLPVDVVAAPACSQDVPXKFVPA-QEIPNGWKLADIGPASVRLFKVEISDAQTIW
H. sapiens	NGVKITLPVDVFTADKFDENAKTQATVAGSIPAGWMLDCGPSSSKYAEAVTRAKQIV
T. maritima	KGVEIVLPVDVAVIAQKIEPGVEKKVVRIDDGIPEGWMLDIDGPETIELFKQLSDAKTVV
B. staerothermop	KGVRFYMPVDVVADRFDANTKVVPI-DAIPADWSALDIDGPKTRELYRDVIRESKLVV
T. brucei	RKVQVILPIDHVCHTEF-KAVDSPLETEDQNIPEGHMLDIDGPKTIEKYVQTIGKCKSAI
	* . . : * * * * : : * : : . : * * * * : : * * : : . : : : .
H. pylori	WNGPLGVHAKQEFARGTIFLAHKIADT----YAFSLIGGGTIDAINRAGEKDNMSFIST
C. jejuni	WNGPXGVFIDKFSKSGSIKXSHYISEG----HATSVVGGGTDADVVARAGDADEXTFIST
H. sapiens	WNGPVGVEWEAFARGTKALMDEVVKATSR-GCITIIGGGTATCCAKWNTEDEKVSHVST
T. maritima	WNGPMGVFIDDFAEGTKQVALAIAALTEK-GAITVVGGGSAARAVNKFLEDKFSHVST
B. staerothermop	WNGPMGVFMDFAFHGTAKAIAEALAEAL---DTYSVIGGGDSAAAVEKFGIADKMDHIST
T. brucei	WNGPMGVFMVYPYSGTFAIAKAMGRGTHEHGIMSIIIGGGDSASAAELSGEAKRMSHVST
	* * * * * * * : : * : : : : : * * * * * : : * * * * * : : . : . : * * :
H. pylori	GGGASLELLEGGKILPCFEVLDNRH--
C. jejuni	GGGASLELIEGKEILPGVKALRSKENE
H. sapiens	GGGASLELLEGGKILPGVDALSNL---
T. maritima	GGGASLEFLEGKEILPGIASMRILKKA-
B. staerothermop	GGGASLEFMEGKQILPGVVALEDK---
T. brucei	GGGASLELLEGGKILPGVTVLDNRK---
	* * * * * * : : * * * * * :

**Figure 4.19** Multiple sequence alignment by MAFFT of human, *H. pylori*, *B. staerothermophilus* and *T. brucei* PGKs (with structures deposited in the PDB). The alignment shows 3,PGA binding residues (light green) in the N-terminal domain, ADP binding residues (dark green) in the C-terminal domain, hinge regions (helices) residues (red) and the “salt bridge” forming residues (blue). Variation identified in the latter C-terminal part of the “salt bridge” as Asp (D) in human, *T. brucei* and *B. staerothermophilus*, Gly (G) in *C. jejuni* and Ser (S) in *H. pylori*. The N-terminal part of the salt bridge that is Arg (R) is conserved.

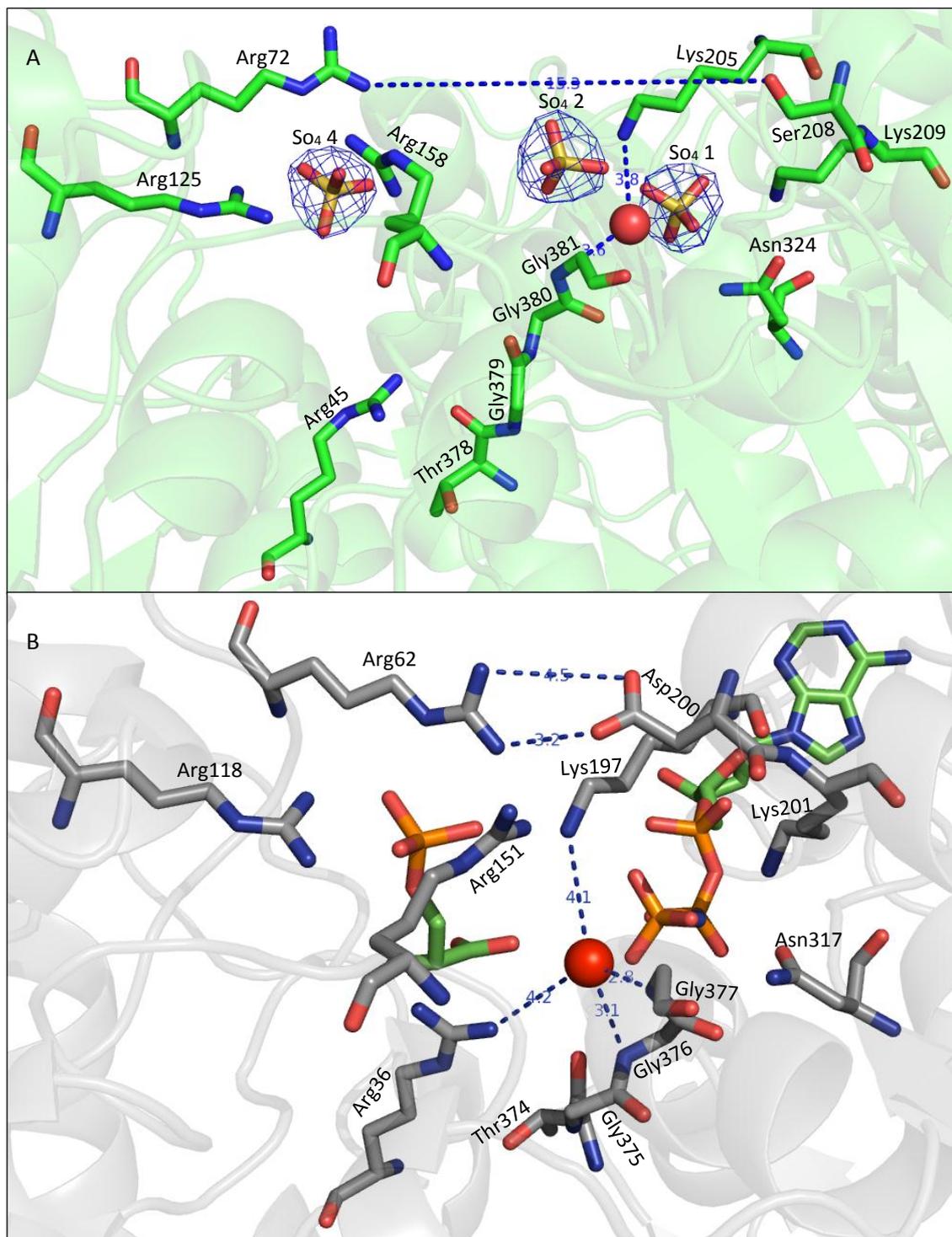
This sulphate is seen in the same site, and with equivalent bonding distances to that of *Plasmodium falciparum* PGK interacting with arginine residues. These arginines form part of the "Basic Patch" which consists of five arginine residues and three histidine residues (Smith et al., 2011). The majority of the patch residues are shown in figure 4.17 (section 4.5.5). The interactions of Arg72, Arg125 and Arg158 in the basic patch of hpPGK in the N-terminal domain of hpPGK are illustrated in figure 4.21 A. This result gives a direct evidence for anion binding in this site. Further kinetics investigation is required to check if sulphate is inhibitory (by competition) (Smith et al., 2011).

Comparing the active site binding residues with those of the closed structure of *Thermogota maritima* PGK (tmPGK) (accession code 1VPE) (Auerbach et al., 1997) (figure 4.21 A), it can be concluded that  $SO_4^{2-}$  ion (figure 4.20) corresponds to the same phosphate group bound to 3,PGA. Furthermore,  $SO_4^{2-}$  binds to a strictly conserved Lys205 residue in the C-terminal domain of chain B monomer of hpPGK. Moreover, a well-defined water molecule bound to the Lys205 and Gly381 (in the C-terminal domain) is positioned on the trajectory proposed for the phosphoryl transfer. This means that the molecule might represent the conformation of the putative fully liganded hpPGK immediately after the phosphoryl transfer reaction. Finally, comparison of  $SO_4^{2-}$  binding residues to their homologues in tmPGK (Lys 205 and Asn324) show that this ion molecule is positioned where the  $\gamma$ -phosphate of the nucleotide (AMP-PNP in tmPGK) is located in the C-terminal domain.

In conclusion, a sulphate molecule was identified in chain A bound to the basic patch in the N-terminal domain of hpPGK. This result supports the notion that this patch has regulatory effect rather than catalytic as there is no immediate structural change induced by the ion binding (Smith et al., 2011). Moreover, the crystal structure of hpPGK in the open conformation is similar to other solved structures PGKs of other organisms. However, three sulphate ions were observed in chain B for the first time. Comparison with the fully closed structure of tmPGK showed that these ions can explain phosphoryl transfer trajectory between 3,PGA and ATP during catalysis.



**Figure 4.20** Sulphate ligands bound to hpPGK structure. The figure shows three sulphate molecules bound to Chain B of the hpPGK dimer. Single sulphate is bound to chain A of the dimer.



**Figure 4.21** The active sites of hpPGK (panel A) and tmPGK (panel B). 2Fo-Fc map (prepared using ccp4 and Pymol) shows clear densities for three sulphate ions (shown as sticks) in chain B of hpPGK labelled SO<sub>4</sub> 1, SO<sub>4</sub> 2 and SO<sub>4</sub> 4. The sulphate binding residues are shown as sticks and central bound water molecule in each structure located in the putative phosphoryl transition state position (are shown as red spheres). Lys205 and Lys197 in hpPGK and tmPGK respectively are well oriented for gripping the phosphoryl group from the donor to the acceptor substrates 3,PGA and ATP or 1,3-BPG and ADP.

## 4.6 Homology modelling of hpPPSA

Experimental determination of protein structure may often be frustrated by cloning, expression or purification problems which may prevent the obtaining of sufficient amount of pure protein. Furthermore, subsequent crystallisation and/or data collection may also be problematic.

When the experimental techniques fail, homology modelling can be a useful computational technique to extrapolate the three dimensional structure of unknown proteins. This uses experimentally solved protein structures (from x-ray crystallography or NMR) to predict conformation of other proteins on the basis of sequence similarity likely reflecting structural similarity (Martí-Renom et al., 2000). Homology modelling can provide “low-resolution structure” with sufficient information about the spatial arrangement of critical residues in the protein which may lead to hypothesis generation, allowing the design of new experiments such as site-directed mutagenesis, to test these hypotheses. The following sections describe the fully automated steps of homology modelling of hpPPSA, namely primary sequence alignment, secondary structure prediction and the protein three dimensional structure prediction by the Phyre 2 web portal for protein modelling (Kelley et al., 2015).

### 4.6.1 Primary sequence alignment of hpPPSA

As of 2018, only one structure has been solved and deposited in the PDB for phosphoenolpyruvate synthase (PPSA) also known as pyruvate water dikinase. The paper describing this work has yet to be published. The crystal of the solved structure (PDB accession code 2OLS) was obtained from the enzyme from *Neisseria meningitides* PPSA protein (nmPPSA).

It is possible that the PDB might contain proteins with significant or useful sequence similarity to hpPPSA, but they might not be annotated as such, therefore the hpPPSA amino acid sequence was submitted in the Basic alignment Search Tool (BLAST) web-server (using the database of the National Centre of Biotechnology Information (NCBI)) for a primary sequence search against the Protein Data Bank (PDB) to identify available structurally related proteins (section 2.9.1). The search result verified that nmPPSA is

the only solved structure in the PDB for this enzyme, and shares 53% sequence identity with hpPPSA. The BLAST search also revealed that hpPPSA shared 35% sequence identity with the unusual rifampin resistance kinase, Rifampin phosphotransferase, from *Listeria monocytogenes* (RPH) (accession code 5HV1). RPH similarity with PPSA is restricted to the ATP-grasp domain and the swivel phosphohistidine domain that contain the conserved histidine residue essential for catalysis. However, the domain architecture of RPH differ than that of PPSA as it is found at the C-terminal region of the protein rather than at the centre. Moreover, the rest of RPH amino acid sequence exhibits relatively low similarity with PPSA (Stogios et al., 2016). Hence, the ATP-grasp domain that was solved in complex with Adenylyl-imidodiphosphate (AMP-PNP) was used to compare the ATP analogue binding sites of RPH to the hpPPSA.

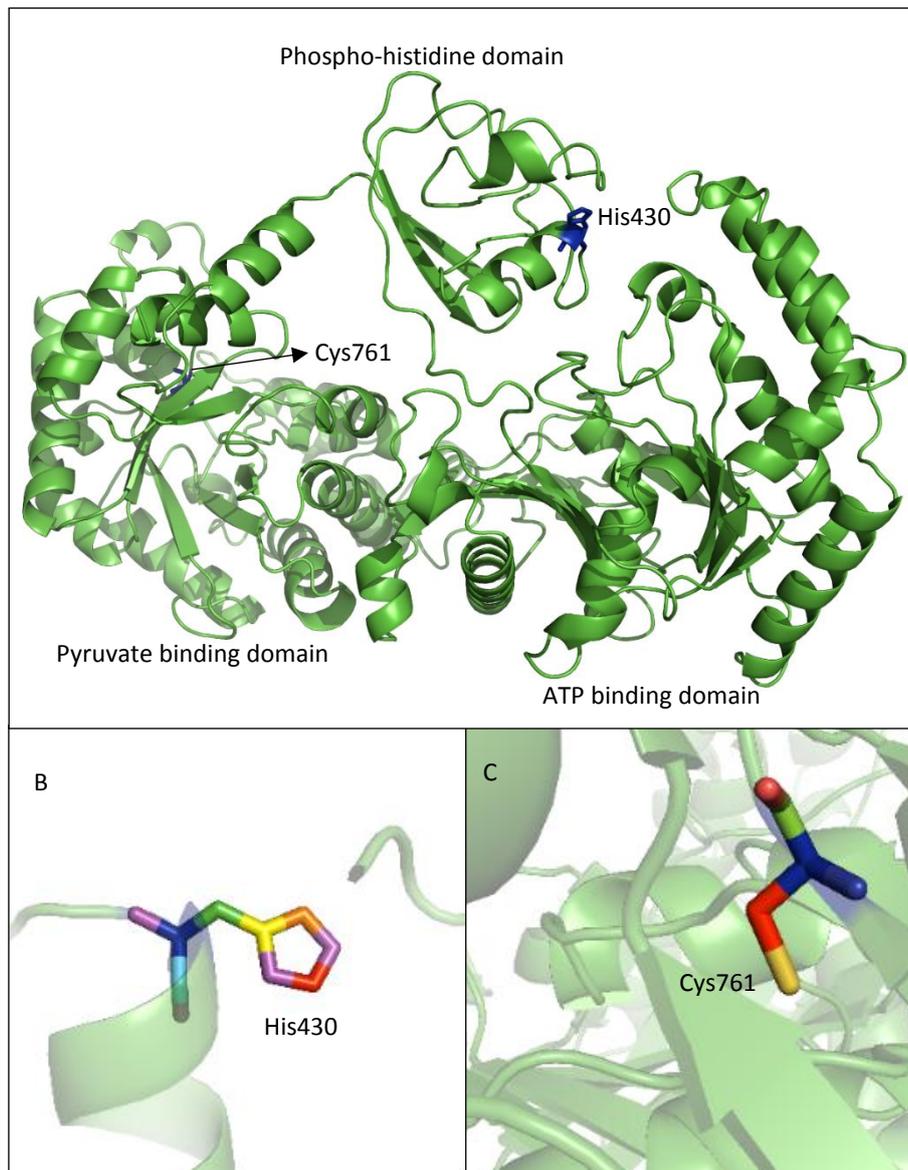
#### 4.6.2 Homology modelling of hpPPSA by Phyre2

After primary sequence alignment, the hpPPSA amino acid sequence was submitted to the Phyre2 web portal for homology modelling to predict the three dimensional structure. Phyre2 modelled 721 residues out of 812 (89% coverage) with 100% reported confidence (Figure 4.22 panel A).

Like other pyruvate water dikinases (section 1.3.6), The modelling results suggests that the hpPPSA structure is arranged in three domains, an ATP binding N-terminal domain (described in section 4.6.3), a central swivel phospho-histidine domain that harbours His-430, the essential catalytic conserved residue that hydrolyses the  $\beta$ -phosphate to generate AMP, Phosphate and a phospho-enzyme intermediate (figure 4.22 panel B) and a C-terminal domain that binds pyruvate and transfers the  $\beta$ -phosphate via swivelling domain mechanism to produce phosphoenolpyruvate (McCormick and Jakeman, 2015). The methylene group of the pyruvate moiety might be located close to Cys-761 (figure 4.22 panel C) which plays a catalytic role in providing the proton to the enolate form of pyruvate (Herzberg et al., 1996).

To help assess the validity of this homology modelling, the output from the web portal includes secondary structure prediction a measure of the confidence of disorder (figure 4.23). The results generally showed that the confidence in the model is high for nearly

all amino acid residues. However, the N-terminus region was poorly modelled (disordered) as are other short regions between residues 242 to 248 and residues 541 to 548. Other regions of low confidence (between residues 127 to 133 and residues 310-313) were identified as corresponding to homologues regions that were missing from the original experimental structure of the template (nmPPSA).



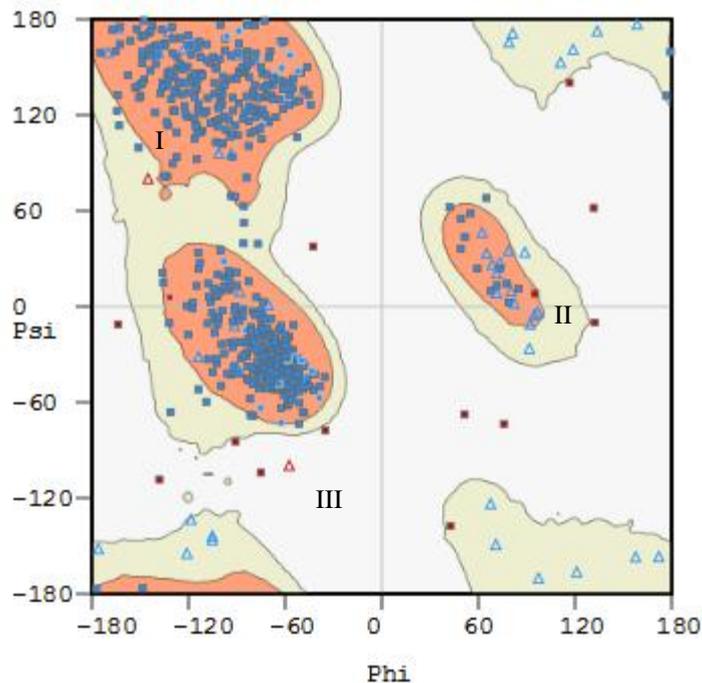
**Figure 4.22** Phyre2 Model of hpPPSA. Panel (A) shows the overall fold and domains. (B) Illustrates the essential conserved catalytic His-430 residue, in the central swivel domain. (C) Shows the catalytic residue Cys-761 in the C-terminal pyruvate binding domain which plays a key role in  $\beta$ -phosphate transfer.





**Figure 4.23** Secondary structure prediction and disorder confidence of hpPPSA calculated by Phyre2 (Kelley et al., 2015). The figure illustrates the index of confidence key, disordered regions and major secondary structure elements.

For further analysis of the quality of the model, Ramachandran plots of the Phyre2 hpPPSA model backbone dihedral angles were calculated. This showed that the majority of the amino acids residues of hpPPSA model in the most favoured or allowed regions (97.7%), with 2.3% of residues as outliers, falling in the disallowed regions (figure 4.24).



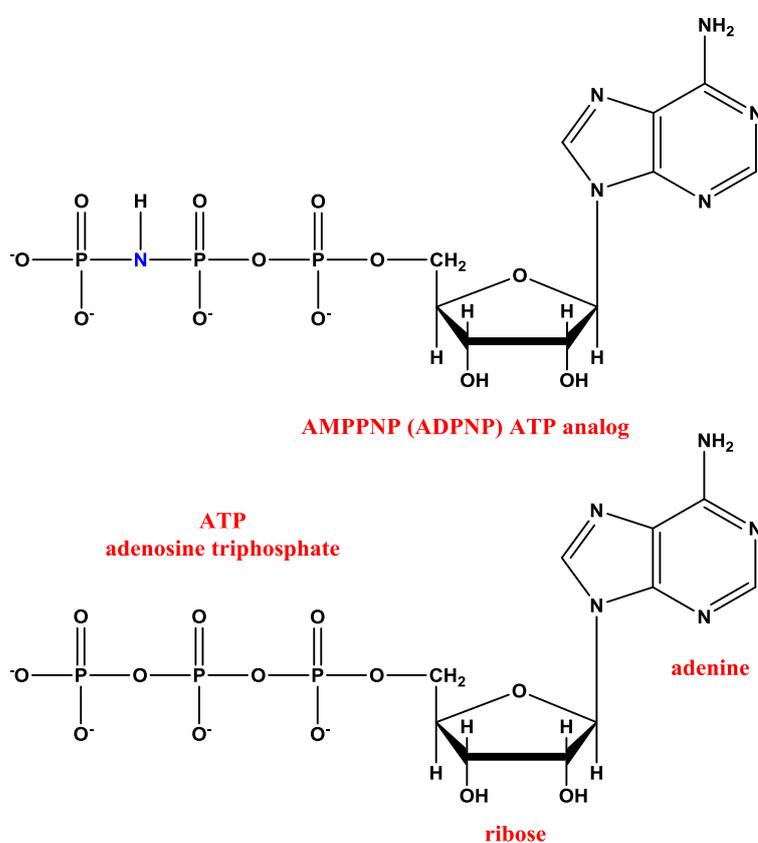
**Figure 4.24** Ramachandran plot of Phyre2 hpPPSA model. The plot was calculated within Win Coot 0.8.9. Amino acid residues in areas I and II represent the favoured and allowed regions, while residues in III area are the disallowed regions (highlighted in red colour).

Having a good degree of confidence and significantly high proportion of the residues in the allowed region, the ATP binding domain of hpPPSA model was then superposed onto the ATP-grasp domain of RPH of *Listeria monocytogenes* (PDB ID 5HV1 (Qi et al., 2016)) that had been solved in complex with the ATP analogue AMP-PNP.

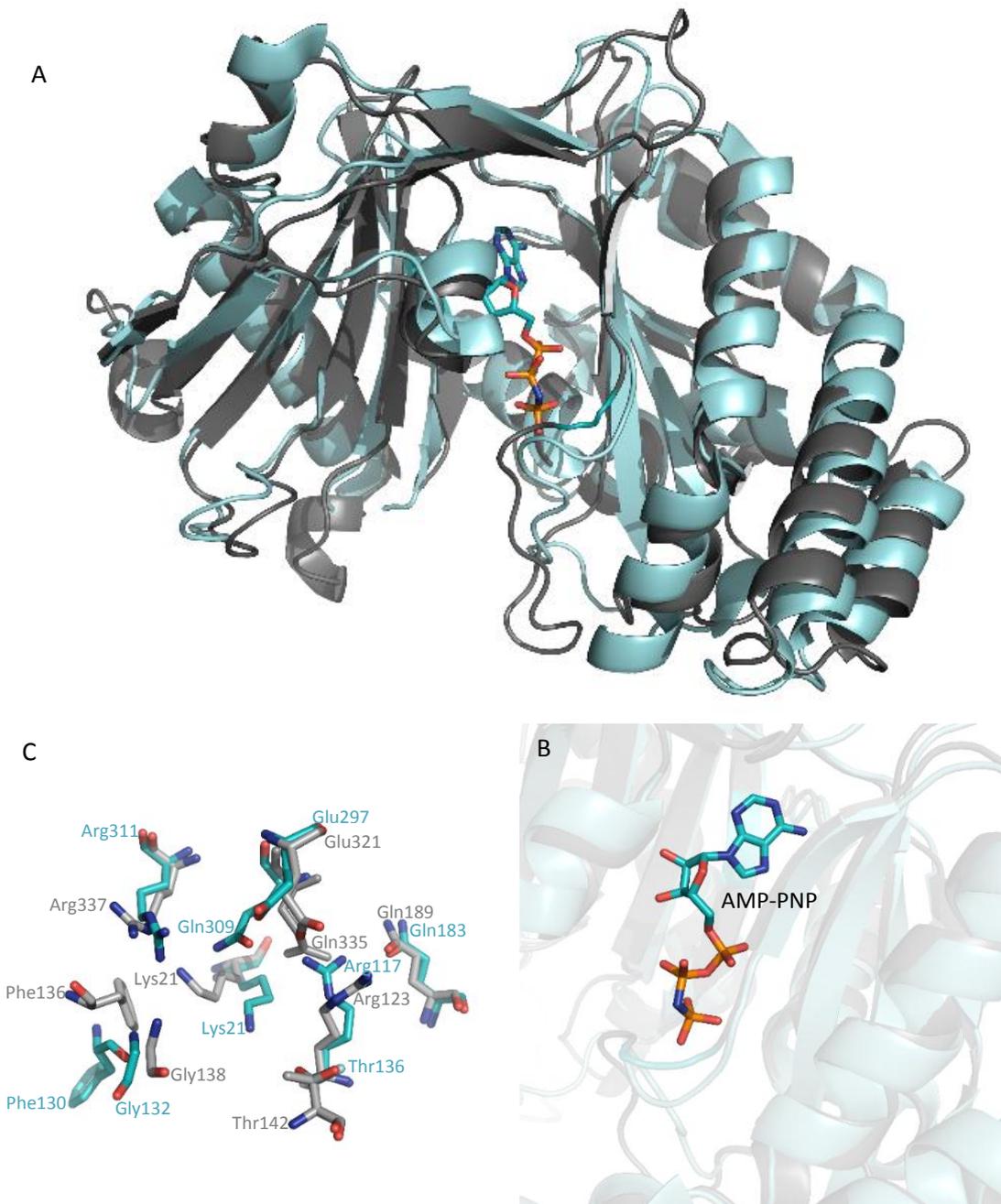
### 4.6.3 Ligand binding sites of hpPPSA ATP binding domain

As structural similarity of hpPPSA to RPH ATP-grasp domain is seen (as described in section 4.6.1), the ATP binding domain of hpPPSA model was overlaid to the RPH domain of *Listeria monocytogenes* (PDB ID 5HV1). The overlay result demonstrated that the overall fold of hpPPSA and RPH ATP-grasp domains are similar. The later structure had been solved in complex with Adenylyl-imidodiphosphate (AMP-PNP) a non-hydrolysable substrate analog of ATP (figure 4.25). The structure of the AMP-PNP binding site of RPH was compared to the ATP binding domain of hpPPSA model and the residues involved in ATP binding compared. All the residues identified to be involved in the nucleotide binding are very well conserved (figure 4.26).

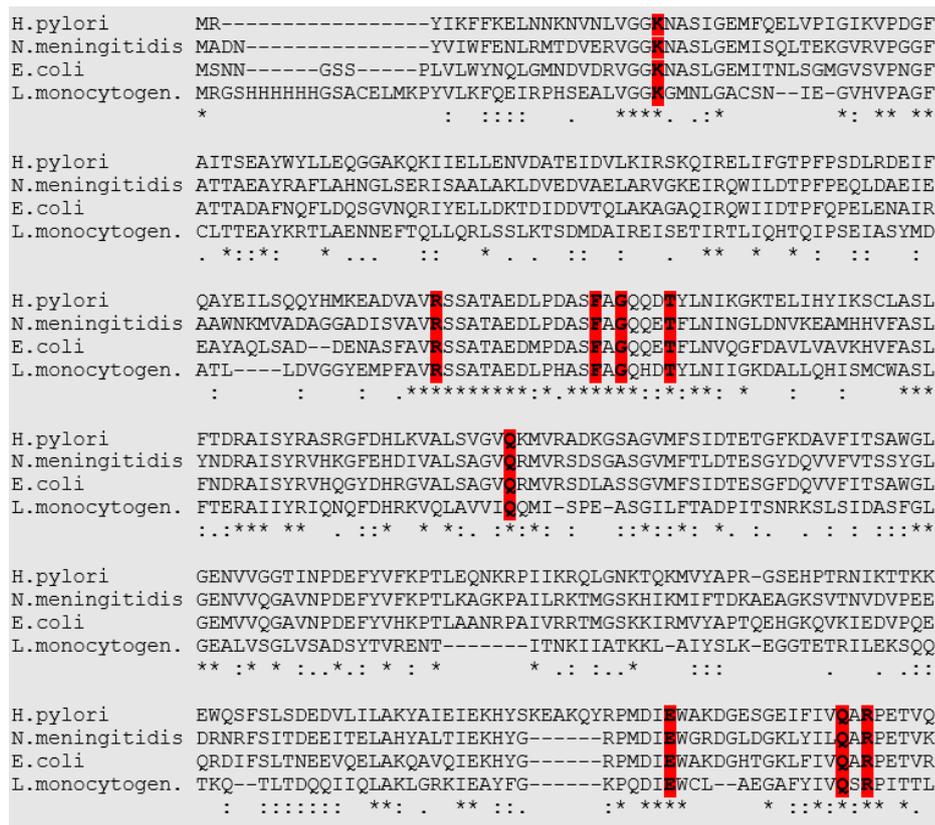
For further confirmation, the residues implicated from this in the interaction with ATP were further investigated by using the multiple sequence alignment approach in MAFFT (Kato and Standley, 2013). The conservation of the PPSA's ATP binding residues and ATP binding residues of RPH of *monocytogenes* was examined in the bacterial PPSAs of *H. pylori*, *E. coli* and *Neisseria meningitides* (figure 4.27 and table 4.4). The sequence alignments results showing that all the ATP binding residues in the nucleotide binding domain are well conserved.



**Figure 4.25** Schematic representation of the nucleotides analogue structures of adenosine triphosphate (ATP) and Adenylyl-imidodiphosphate (AMP-PNP).



**Figure 4.26** Comparison of ATP binding residues in ATP-grasp domain of *Listeria monocytogenes* RPH solved in complex with AMP-PNP with their equivalents in the ATP binding domain of *H. pylori* PPSA model. Panel (A) structural overlay of RPH ATP-grasp domain liganded with AMP-PNP structure (shown in cyan) with the ATP binding domain of hpPPSA model (shown in gray). Panel (B) Close view of AMP-PNP with transparent background of the overlaid RPH AT-grasp domain structure with ATP binding domain of hpPPSA homology model. Panel (C) shows the overlay of residues involved in AMP-PNP binding site of RPH ATP-grasp domain structure (shown in cyan) with equivalent residues involved in binding site of hpPPSA model (shown in gray).



**Figure 4.27** multiple sequence alignment of ATP binding domain in PPSA of *H. pylori*, *N. meningitidis* and *E. coli* in addition to the ATP-grasp domain in RPH of *L. monocytogenes*. All the amino acid sequences were obtained from KEGG genetic data bases (Kanehisa and Goto, 2000) except the RPH sequence from PDB 5HV1. The residues involved in ATP binding are highlighted in red.

**Table 4.4** Residues involved in ATP binding in *L. monocytogenes* RPH and their equivalents in *H. pylori*, *N. meningitidis* and *E. coli* PPSAs. The information is derived from figure 4.26 and figure 4.27.

Residue	<i>L. Monocytogenes</i>	<i>H. pylori</i>	<i>N. meningitidis</i>	<i>E. coli</i>
Lys	22	21	23	27
Arg	117	123	125	127
Phe	130	136	138	140
Gly	132	138	140	142
Thr	136	142	144	146
Gln	183	189	191	193
Glu	297	321	318	320
Gln	309	335	332	334
Arg	311	337	334	336

## 4.7 Summary

The work described in this chapter adds to the genetic studies in the previous chapter which investigated synthetic lethality of *ppsA* and *pgk* metabolic genes in the enteric pathogen *H. pylori*. The results of the genetic study highlighted that each one of these glycolytic-gluconeogenic genes might be essential alone to the viability of the pathogenic bacteria. Accordingly, this chapter has addressed the cloning, expression, purification and crystallisation of the *ppsA* and *pgk* genes from *H. pylori*.

The crystal structure of apo-hpPGK was solved and compared to the fully closed structure of human PGK. Structural superposition analysis showed that both 3-PGA or 1,3-BPG and ATP or ADP binding residues are well conserved. However, the enzyme can be targeted in terms of the contextual differences between *H. pylori* and the human's metabolic systems. Moreover, four sulphate ions were seen in the hpPGK dimer. The positions of these molecules as indicators of phosphate positions, allowed the path of phosphoryl transfer between the substrate and nucleotide domains during catalysis to be modelled, in addition to highlighting the regulatory effect of the basic patch in the former domain.

As hpPPSA purification was problematic, bioinformatics-based prediction of the enzyme structure was conducted using the only solved PPSA structure in the PDB as a search model namely *Neisseria meningitides* PPSA. Homology modelling analysis compared the essential catalytic conserved residues in the central and C-terminal domain that is His and Cys respectively. In addition, the ATP binding N-terminal domain was superposed to the corresponding domain of RPH of *Listeria monocytogenes* to compare the residues responsible for ATP binding. The results showed that these are conserved. As there is no corresponding orthologue of PPSA in human, this will keep the door open for further genetics, kinetics and structural studies toward possible leads for targeting this crucial preliminary gluconeogenic enzyme.

## Chapter 5. Discussion

### 5.1 Introduction

The majority of metabolic pathways have been extensively studied in mammals and in *E. coli* in particular. However, the facile assumption that characterising the pathways of these organisms allows general extrapolation to organisms whose metabolic context may be quite different, is unsafe. For example, only a few of the glycolytic-gluconeogenic enzymes of *H. pylori* have been studied until recently, and so our knowledge of the metabolism of this pathogen necessitates further study. To shed further light on the metabolism of the pathogen, the work described here focuses on some of the enzymes responsible for catalysing critical steps in glycolysis-gluconeogenesis. Computational mutational studies predict that in *Helicobacter pylori* the gene encoding phosphoenolpyruvate synthase hpPPSA is conditionally essential (synthetic lethal) with any of each one of the genes encoding glycolytic-gluconeogenic enzymes (PGK, PGM, ENO and GAPDH). The term synthetic lethal genes describes pairs of genes whose double deletion rendered biomass production impossible whereas neither of whose single deletion affects cell viability (Thiele et al., 2005). Understanding the genetics, structure and function of these enzymes in *H. pylori* metabolism may be a step forward toward new treatment strategies against the pathogen.

The following sections discuss the findings described earlier of the synthetic lethality investigations of *ppsA* and *pgk*, the kinetics and structural analysis of hpPGK and the bioinformatics-based structural analysis of hpPPSA.

### 5.2 Synthetic lethality investigation in *H. pylori*

Genomic studies identifying the genes essential to the viability of pathogenic bacteria is a fundamental step towards identifying potential targets for development of antibiotics against these organisms. Transposon mutagenesis and allelic replacement are the commonly used approaches for constructing recombinant strains to mapping essential genes (Amini and Tavazoie, 2011).

However, the gene encoding PPSA (*ppsA*) was not identified as a selective essential gene in *H. pylori* by genome prioritisation (a bioinformatics genome analysis to identify highly diverged *H. pylori* ORFs) and allelic replacement mutagenesis strategy (Chalker et al., 2001). Similarly, *ppsA* was designated as a non-essential gene in single enzyme-coding gene mutational studies using random mutagenesis and loop amplification (RMAL) (Jenks et al., 2001) and microarray tracking of global transposon mutagenesis (MATT) (Salama et al., 2004). In addition, the gene coding PGK in *H. pylori* (*pgk*) was assigned to be non-essential in the latter study.

Furthermore, *in silico* mutational deletion studies were carried out to identify essential and conditionally essential (synthetic lethal) genes in *H. pylori*. Single-deletion *in silico* studies were performed wherein reactions in central metabolism with their associated genes were deleted and their effects were computationally evaluated under a variety of medium conditions, resulting in a number of reactions that are considered essential. However, *in silico* single-deletion mutagenesis of *the ppsA* and *pgk* genes with their reactions were hypothesised as non-essential genes as they both failed to show essentiality under all of the conditions examined (Schilling et al., 2002, Thiele et al., 2005). On the other hand, *in silico* double deletion studies identified *ppsA* and *pgk* as synthetic lethal mutants, conditionally essential metabolic genes, as their double-knockout led to a lethal mutant whereby biomass production was rendered impossible in rich medium (Thiele et al., 2005). It should also be noted that the functionality of the individual gene products used in modelling are mostly annotated by extrapolation from similar genes in other organisms whose metabolic context may be quite different.

In order to test the synthetic lethality hypothesis, the experimental effect of single and double knockouts of *ppsA* and/or *pgk* on the viability of *H. pylori* was investigated with the appropriate controls. Mutagenesis of *ppsA* and *pgk* (with *rocF* control) was used to test their synthetic lethality in *H. pylori*. This is discussed in detail in the following sections where the results of the mutagenesis analysis are presented in the context of the literature.

### 5.2.1 Mutagenesis of *pgk*

Insertional-inactivation mutagenesis strategy was followed to mutate *pgk*. This was conducted by inserting a terminator-less chloramphenicol acetyl-transferase (*cat*) gene with cognate promoter into a *Bgl*III restriction site in the middle of *pgk*. Flanking regions from *gap* and *corA* genes (upstream and downstream the *pgk* gene) were used to ensure that the allelic exchange of the inactivated mutant allele with the wild type allele was achieved. Insertion of the antibiotic resistance cassette between the N- and C- domains of PGK means that they cannot be brought together for the phosphoglycerate kinase enzymatic function of phosphate transfer between these domains.

The *pgk* mutated allele was successfully introduced to *H. pylori*. However, a PCR product using *pgk::cat* *H. pylori* genomic DNA was required for further verification by sequencing (section 3.6). The PCR produced two variable fragments, one of them corresponds to the wild type *pgk* allele and the other represent *pgk::cat* mutated allele. This variation could reflect the following:

Non-homologous recombination under the selective pressure of chloramphenicol presence. However, the findings of *pgk* knockout in the current study do not support this hypothesis as the results (PCR verification and sequencing (section 3.4.1.2)) indicated that the allelic exchange occurred in the correct position on the chromosome as a result of homologous recombination between the mutated and wild type alleles. Nonetheless, a wild type allele may be present.

A single crossover event might have occurred at the chromosomal *pgk* locus to result in the integration of *pgk::cat* allele, the suicidal plasmid (pUC19) and the wild-type *pgk* allele (Reyrat et al., 1998, Heap et al., 2012). This crossover is in general more likely than a double crossover. This could be explained by *pgk* possibly being essential to the viability of *H. pylori* and therefore the bacteria retain a copy of the wild type allele in addition to the mutated allele. Nonetheless, *pgk* mutagenesis results that were achieved in this study require further investigation to identify whether this metabolic gene is essential to *H. pylori* viability or not. This can be pursued using strategies such as DNA Southern blotting by probe hybridisation of *pgk::cat* *H. pylori* genomic DNA with

labelled specific restricted fragments located within *pgk::cat* and wild-type *pgk* alleles. Additionally, Genome sequencing of the *pgk* mutated *H. pylori* should be carried out to identify any possible simultaneous non-homologous recombination of the mutated gene.

A further study with more focus on *pgk* essentiality is suggested by following a gene complementation strategy. This can be achieved by cloning a second copy of *pgk* and integrating it into a non-functional region within the *H. pylori* genome and placing it under the control of a constitutive promoter (activate gene expression independent of the influence of regulation), and then the *pgk* wild type allele could be knocked out with inserting a promoter-less and terminator-less antibiotic resistant cassette or the gene expression might be turned off by controlling transcription. Should the bacteria grow, then *pgk* essentiality will be verified.

This variation possibly resulted from a “synergy phenomenon” where putative recombinant cells (containing the mutated allele) can only exist if part of a colony with a wild type allele containing cells. This explanation agrees with the *pgk* essentiality reported in the literature.

The gene was identified as non-essential to the viability of *H. pylori* in a global transposon mutagenesis study. The result was verified as five transposon insertions (hits) were detected within *pgk* by amplification using the transposon specific primer in addition to the nonspecific primer ending in AGAG unique anchor nucleotides, which on average anneals every 260 bp based on the published *H. pylori* 26695 genome sequencing data (Tomb et al., 1997). The results were reported after MATT mutagenesis using the modified Tn7-based *in vitro* mutagenesis system. This result matched the number of expected hits per the *pgk* gene whose length is approximately 1200 bp (Salama et al., 2004). Moreover, *in silico* essentiality results for *pgk* were negative as the gene deletion in central metabolism of *H. pylori* showed ability to make all 47 key biomass constituents under three different substrate availability conditions (out of four) (Schilling et al., 2002).

In the light of these results, *pgk* was apparently reported as a non-essential gene by Transposon mutagenesis and *in silico* deletion mutagenesis. Nonetheless, potentially *pgk* was miss-interpreted as non-essential gene as the transposons insertion events occurred in non-essential regions of the gene which might led to the expression of the protein. Insertion-inactivation mutagenesis (using pUC19 suicidal plasmid, *pgk* and chloramphenicol resistance cassette) of the gene followed by allelic exchange strategy in the chromosome of *H. pylori* revealed that *pgk* knockout results was unexpected according to the variation of mutated and wild type alleles (section 3.6). These results are likely to be related to single crossover event with a chance of multiple copies of the mutated and wild type alleles. However, for unknown reasons some of the putative recombinant cells might revert back to wild type.

In future investigations, it might be possible to use different vectors other than pUC19 such as selection/counter-selection system for *pgk* mutagenesis prior to the introduction of the mutated gene into *H. pylori*. Upon introduction to the bacterial cell, the machinery of homologous recombination within the cell will identify the sites of homology between the vector and the chromosome and at some frequency will trigger the integration of the vector at the site of homology. This event is frequently selected for by the presence of an antibiotic resistance cassette within the vector. Following this selection, the vector should be removed to facilitate homologous recombination. A counter-selectable marker (such as SacB which is lethal when expressed in gram negative bacteria in the presence of sucrose) is included on the vector for this purpose. Only those cells that have excised the vector sequence through a second recombination event will survive upon induction of the counter-selectable condition. The counter-selection step is essential to find the desired mutants as the recombination event is quite rare (Solem et al., 2008).

Conclusively, in any case (either the enzyme coding gene is essential alone, conditionally essential with *ppsA* or non-essential to the viability of *H. pylori*), PGK will remain as a potential drug target either alone or with other metabolic gene products such as the gluconeogenic enzyme PPSA (Thiele et al., 2005), or possibly with Krebs cycle enzymes such as fumarase as this cycle in *H. pylori* deviates from the standard textbook examples

such as human (section 1.5.3) (Kather et al., 2000). Should *pgk* be essential to the viability of *H. pylori*, these findings will have important implications for the central metabolism of the pathogenic bacteria and developing therapies against this potential drug target.

### 5.2.2 Mutagenesis of *ppsA*

As described in section 5.2, *in silico* single-deletion mutagenesis and experimental random insertional transposon mutagenesis studies reported that *ppsA* is non-essential to the viability of *H. pylori*. In the last global transposon mutagenesis study, transposons were inserted randomly and well distributed along the genome of *H. pylori*, which led to the mapping of 5363 transposon mutants (including *ppsA*). Conversely, 344 genes predicted to be essential genes as they had no detectable transposon insertions (Salama et al., 2004).

However, a number of factors can cause systematic biases in random transposon mutagenesis experiments. For example, some genes (especially shorter ones) will inevitably be missed by chance, which will create false positive results as non-essential genes will be miss-interpreted as essential genes. Another drawback, especially with longer genes, the transposon insertion may occur at a non-essential part of the gene, such as the extreme ends, which may not disrupt the gene product function entirely. On the other hand, transposon insertion(s) in non-essential genes can cause polar effect and prevent the transcription of downstream gene(s) in the same operon which can lead to false positive results of the gene essentiality (Deng et al., 2013).

Transposon mutagenesis miss-interpretation potentially was the case with *ppsA* as the gene length is approximately 2400 bp and single transposon integration (hit) was identified (Salama et al., 2004). This result was reported despite using two different nonspecific primers with different anchor sequences. These primers were used to identify the target region(s) (gene(s)) flanking the inserted transposon(s) along the genome. The first primer ended in AGAG and on average anneals every 260 bp, whereas the other one ends in GATAT and on average this anneals every 1100 bp in the chromosome. Both sequences were based on the published *H. pylori* 26695 genome

sequencing data (Tomb et al., 1997). Accordingly, *ppsA* should have 2-8 transposon insertions (hits) in order to show reliable evidence of non-essentiality. Therefore, the results of *ppsA* transposon mutagenesis do not show an evidence for the disruption of the gene product. Two possible reasons behind this potential false positive result, transposon insertion may occur at a non-essential part of *ppsA* and/or the few number of transposon insertion(s) per the gene was insufficient to completely disrupt hpPPSA gene product function. Because of this, allelic replacement gene knock out strategy was followed here to clarify *ppsA* essentiality in *H. pylori* 26695.

The *ppsA* gene was mutated by the deletion-inactivation strategy using inverse PCR to enable deletion of significant part of the coding sequence (1800 bp out of 2400 bp) for the enzyme and subsequent insertion of a terminator-less kanamycin resistance cassette with promoter. Moreover, The same strategy was followed to mutate the non-essential *rocF* gene (coding for arginase) which was used as a positive control (Langford et al., 2006). The control construction was successful therefore in general terms the process for making the mutants was fit for purpose.

The knockout results of *ppsA* did not yield of viable recombinants. This provides preliminary evidence that *ppsA* is an essential metabolic gene, as *H. pylori* 26695 was unable to grow after gene inactivation. The potential explanation could be that the flank region downstream *ppsA* (section 3.2.2) that identified as essential genes namely hp0122 and hp0123 (predicted to be upstream of *ppsA* in terms of the direction of transcription) (Salama et al., 2004, Caspi et al., 2012) might be disrupted during the double crossover process or it could be *ppsA* is potentially essential to the metabolism and growth of *H. pylori* and therefore it cannot be mutated.

Mutagenesis result obtained by this method were potentially not consistent with global transposon mutagenesis conclusions and indicate that *ppsA* is likely to be an essential gene. However, it is recommended to clone a second copy of *ppsA* and integrating it into a non-functional region within *H. pylori* genome and place it under the control of constitutive gene expression promoter, and then *ppsA* wild type allele should be mutated by means of deletion inactivation strategy or turning off the gene by

controlling transcription. Should the bacteria grow after complementation, then *ppsA* allele essentiality will be confirmed.

In accordance with the latter result with respect to double deletion mutants, when the mutated *ppsA* and *pgk* introduced in *H. pylori* 26695 (using  $\Delta rocF::kan$  and  $\Delta rocF::cat$  as controls) in the final step of synthetic lethality investigation, the results revealed that all double mutants containing  $\Delta ppsA$  didn't grow after allelic exchange (section 3.5).

## 5.3 Structure and function of hpPGK

### 5.3.1 Kinetics and solution characterisation of hpPGK

This work shows hpPGK forms a homodimer whilst in solution. The PGK of *Corynebacterium glutamicum* (Reddy and Wendisch, 2014) and of some species of thermophilic archaea (Hess et al, 1995) show similar behaviour in solution unlike the majority of PGKs of eukaryotic and prokaryotic organisms which are seen to be monomers in solution (Fifis and Scopes, 1978, Mukherjee et al., 2002). Possibly, the dimerization may be important for providing stability to the enzyme as the contact area (interface between two subunits) reduces the surface area subjected to the aqueous environment whilst the protein is dimer (Kirino et al., 1994, Reddy and Wendisch, 2014). The thermal stability experiment carried out in this study revealed that the hpPGK protein is relatively thermostable, with a melting temperature of 45 °C, which is higher than PGKs of the mammalian *Mus musculus* (40 °C) and yeast *S. cerevisiae* (24 °C) (Ijeoma et al., 2008, Pegoraro and Lee, 1978) which are characterised as monomers in solution. These results agree with the above-mentioned explanation given by Kirino et al., 1994. In addition to thermal stability determinations secondary structure content of hpPGK was also determined by CD. The result suggested that the protein is well folded and alpha helical dominant (Figure 3.19). This result is in agreement with the literature where all structurally known PGKs are alpha helical dominant (Lee et al., 2006, Smith et al., 2011, Zheng et al., 2012).

Kinetics characterisation of hpPGK confirms the physiological role of PGK in *H. pylori*. The effect of the hexa-his tag on the activity was explored to find out whether addition of tag affects the activity and kinetic parameters of the enzyme or not. Comparing

enzymatic activity and kinetic parameters showed the tag had a negligible effect as expected. HpPGK shows Michaelis-Menten kinetics with  $K_m$  values for the substrates 3-PGA and ATP that are within the range of the other sources reported in the literature (Rao and Oesper, 1961, Zomer et al., 1998).

### 5.3.2 Crystal structure of hpPGK

X-ray diffraction data was obtained at relatively good resolution whereby secondary structure elements and many side chains can clearly be seen. This is enough to describe the overall structure and compare with the other structures in the literature. The crystal structure of hpPGK in the open conformation is similar to other solved structures PGKs from both prokaryotic and eukaryotic organisms (Zheng et al., 2012). The residues along the cleft formed between the N-terminal and C-terminal domains during catalysis are highly conserved. Only one residue variation was identified in the C-terminal domain where Ile300 of hpPGK replaces Gly313 in human PGK. However, this does not alter the nucleotide binding as the interaction is to the main chain rather than the side chain in human PGK (Cliff et al., 2010).

Interestingly, another variation is seen in one of the “salt bridge” forming residues of the closed form whereby Ser208 is in the site corresponding to Asp218 in human PGK. This suggest that the link between the domains during catalysis (closed conformation) at this point would have to be a hydrogen bond rather than salt bridge. Structural data from the fully closed structure of hpPGK is required to validate this hypothesis.

Sulphate molecules were identified bound in the basic patch to the N-terminal domain of hpPGK chain A (section 4.5.6). Smith and his co-workers suggested that the ion does not induce any structural changes and discussed its potential regulatory effect (Smith et al., 2011). It has been shown that sodium sulphate activates the enzymatic activity of yeast and *Plasmodium falciparum* PGKs at low concentrations whereas it shows an inhibitory effect at high concentrations (Scopes, 1978, Pal et al., 2004). Similarly the salt reduces the activity of *Trypanosoma brucei* PGK (Zomer et al., 1998). However, it is not clear whether sulphate salts activate or deactivate hpPGK activity. Hence, further kinetic studies are required to investigate this effect.

As the sulphate molecule (bound to chain A) simulates the binding of the phosphate molecule of 3-PGA in the N-terminal domain of other PGK binary complexes, the conformational change in hpPGK might start with phosphoglycerate-PGK binary complex formation, rather than with the formation of nucleotide-PGK binary complex (Auerbach et al., 1997).

Three sulphate ions are observed in chain B of hpPGK for the first time. Comparison of the interactions with the residues in the fully closed structure of *Thermogota maritima* PGK shows that the positions of these ions can describe the phosphoryl transfer trajectory between 3-PGA and ATP (shown by the nucleotide analogue, AMP-PNP) during catalysis (Auerbach et al., 1997). From this, the phospho-transfer mechanism can be extrapolated. The close proximity of the substrate and the nucleotide binding sites (figure 4.21 B) after induced conformational change triggers the nucleophilic attack by the oxygen atom of  $\beta$ -phosphate in ADP at the phosphate number one of 1,3-BPG. This reaction involves a configuration-inversion mechanism at the  $\gamma$ -phosphorus atom (Webb and Trentham, 1980) and follows a single displacement mechanism (Ho et al., 1988). This directly stimulates phosphoryl transfer through a charge transition state between the substrates (Knowles, 1980). The additional negatively-charged pentagonal transition state phosphate has to be stabilised by PGK (Auerbach et al., 1997).

This model sets the stage for analysis of conformation of hpPGK during phosphoryl transfer. Firstly, a large scale hinge-bending motion brings hpPGK domains in close proximity. A potential inter-domain hydrogen bond rather than the conserved salt bridge (in the case of hpPGK) between Arg72 (N-terminal domain) and Ser208 (C-terminal domain) forms a strap to tie the two domains in a closed state. Secondly, the conserved Lys205 (C-terminal domain) might also be implicated in the domains closure by anchoring both phosphoryl groups of the N-terminal bound substrate namely 1,3-BPG followed by the attraction of the C1-bound phosphate of 1,3-BPG (N-terminal domain). This hypothesizes that a critical group is brought into position by the conformational change. Moreover, this residue is well-positioned for escorting the phosphate group from the donor BPG to the acceptor ADP. The residue seems to be in charge of neutralising the negative charge of the phosphate during catalysis, which

would lead to decrease in the free energy of the transition-state intermediate (Auerbach et al., 1997).

### 5.3.3 Implications for drug design

Consistent with it being a conserved enzyme, the genetic findings obtained in the previous chapter highlighted that hpPGK might be essential alone to the viability of *H. pylori*, conditionally essential with other gluconeogenic enzyme namely hpPPSA or it can be targeted with Krebs cycle enzymes (such as fumarase) to disrupt the anaerobic metabolism in the enteric pathogen. The Krebs cycle in *H. pylori* deviates from standard text book examples such as that of human as there are three deviated enzymes namely  $\alpha$ -ketoglutarate:ferredoxin oxidoreductase, malate quinone oxidoreductase and succinyl-CoA:acetoacetyl-CoA transferase. According to that hpPGK can be targeted in terms of the contextual metabolic difference as compared to the human's metabolic context. Hence, hpPGK can be targeted by other PGK inhibitors such as tubercidin or suramin. Should the enzyme show conditional essentiality with hpPPSA or Krebs cycle enzymes, the latter inhibitors can be used with hpPPSA or Krebs cycle inhibitors such as bismuth. Interestingly, there is negligible absorption of bismuth into the blood stream from bismuth salicylate in standard oral doses (Nwokolo et al., 1990). Hence, this will keep the door open for future studies to identify sulphate or phosphate therapeutics (as hpPGK inhibitors) not absorbed through the intestinal tract which might be useful against *H. pylori* but not human's PGK.

Despite extensive trials, no crystals of substrate- or substrate analogue-bound hpPGK were obtained to date. The expectation is that hpPGK in the ligand-bound form will be in the closed form, as the enzyme undergoes conformational changes for catalysis, whereby the flexible hinge allows closure of the  $\sim 15$  Å distance between the 3,PGA and ATP for phosphate transfer. Evidence for the large conformational change required for catalysis had been obtained from yeast PGK by nuclear magnetic resonance (NMR) (Tanswell et al., 1976), small angle X-ray scattering (SAXS) (Pickover et al., 1979), sedimentation experiments (Roustan et al., 1980) and the X-ray crystallographic structure (Watson et al., 1982). Thus the closed form will have a different conformation, and the packing in the crystal would be expected to be significantly different to the open form. That hpPGK

has not been crystallised when complexed with substrate or substrate analogues can be considered as strong circumstantial evidence that it too undergoes this conformational change.

The long-held suspicion about PGK flexibility was verified by solving the ternary structure of *Trypanosoma brucei* PGK (Bernstein et al., 1997). Moreover, two different crystal forms at high resolution solved structures of the ternary complex were achieved using the bisubstrate analogue adenylyl 1,1,5,5-tetrafluoropentane-1,5-bisphosphonate. These conformations represent a starting point for the discovery of inhibitors (drugs) against PGK (Bernstein et al., 1998). More recently, metal fluoride ( $\text{MgF}_3^-$  and  $\text{AlF}_3^-$ ) transition state analogue complexes (TSA) provided the first experimental characterisation (using X-ray crystallography and  $^{19}\text{F}$ -based NMR methods) of the fully closed form of human PGK structure (Cliff et al., 2010). This strategy can be extrapolated to produce crystals to produce TSA complexes as the dominance of charge balance in TSA binding can guide the drug targeting of other orthologues such as hpPGK.

#### 5.4 Homology model of hpPPSA

Overall, the hpPPSA model is assessed as being of good quality with a high degree of confidence, and with 97.7% of residues falling in the allowed and most favoured regions of the Ramachandran plot. Moreover, sequence alignment analysis and superposition of the nucleotide binding domain of the model onto the ATP-grasp domain of the rifampin phosphotransferase (RPH) structure of *Listeria monocytogenes* in complex with the ATP analogue AMP-PNP showed that the catalytic residues His430 and Cys761 and ATP binding residues are conserved.

However, there is no corresponding orthologue to PPSA in humans, and the findings of the genetics work in the previous chapter suggest this enzyme may be essential to the viability of *H. pylori*. In this case, the enzyme is potential drug target and this model could be used in virtual screening studies to identify potential theoretical hpPPSA inhibitors. This may provide leads for therapies in the future. Moreover, the homology modelling of hpPPSA has provided a framework for future discussion of strategies for

obtaining crystal structures, such as deleting the predicted disordered residues to improve cloning, expression, purification and crystallisation. Site directed mutagenesis studies of the catalytic residues in the histidine and pyruvate binding domains, can be used to experimentally test the model and to investigate their effects on hpPPSA in terms of the functional and/or the structural levels. This can then lead to further kinetics and/or structural studies to identify inhibitors against this potentially unique drug target.

## 5.5 Implications for *H. pylori* central metabolism

The genetic, kinetic, structural and bioinformatics findings of the metabolic enzymes hpPPSA and hpPGK described in this study need to be interpreted in the context of the central metabolism of the enteric pathogen with appropriate caution as additional investigations are required in the future to verify conclusions.

These interpretations must be consistent with the observation that the Microaerophile *H. pylori* efficiently thrives and colonises the gastrointestinal mucosa of humans and migrate (using polar flagella) deeply into the crypts and near the inner epithelial cells. The central metabolism in *H. pylori* is highly conserved and well adapted for the low oxygen level microaerobic environment. As the primary carbon and energy sources are amino acids and organic acids, *H. pylori* utilises glucose sparingly. Furthermore, the pathogenic bacteria must rely on phosphoenolpyruvate synthase (PPSA) to satisfy gluconeogenic requirements as it lacks anaplerotic reactions that connect the Krebs cycle with gluconeogenesis, for example linking oxaloacetate with phosphoenolpyruvate. These requirements are essential intermediates for anabolic processes such as synthesis of nucleic acids and cell wall material (Tomb et al., 1997, Alm et al., 1999, Doig et al., 1999, Schilling et al., 2002, Nagata et al., 2003).

As PGK in a glycolytic context produces two ATP molecules per 6-carbon sugars, it is a required step in the degradation of glucose to pyruvate. The enzyme is therefore critical target for disrupting respiration (Zheng et al., 2012).

In addition to the work pursued in this thesis, a number of experimental genetic, structural and mechanistic studies were carried out to characterize some of the

enzymes in glycolysis-gluconeogenesis pathways in *H. pylori*. For example biophysical and kinetic investigation indicated apparent activity of the key role enzyme fructose-1,6-bisphosphatase (FBPase) that catalyses the second irreversible gluconeogenic reaction (Ayna, 2016). Furthermore, kinetic and structural study focused on glyceraldehyde-3-phosphate dehydrogenases (GAPDHs) emphasised that one of the two identified GAPDHs of *H. pylori* intensely prefers NADPH over NADH, similar to the enzymes in gluconeogenic photosynthetic organisms and archaea (Elliott, 2009). In addition to the other findings, it can be hypothesised that *H. pylori* might be a gluconeogenic rather than glycolytic organism.

## 5.6 Conclusions

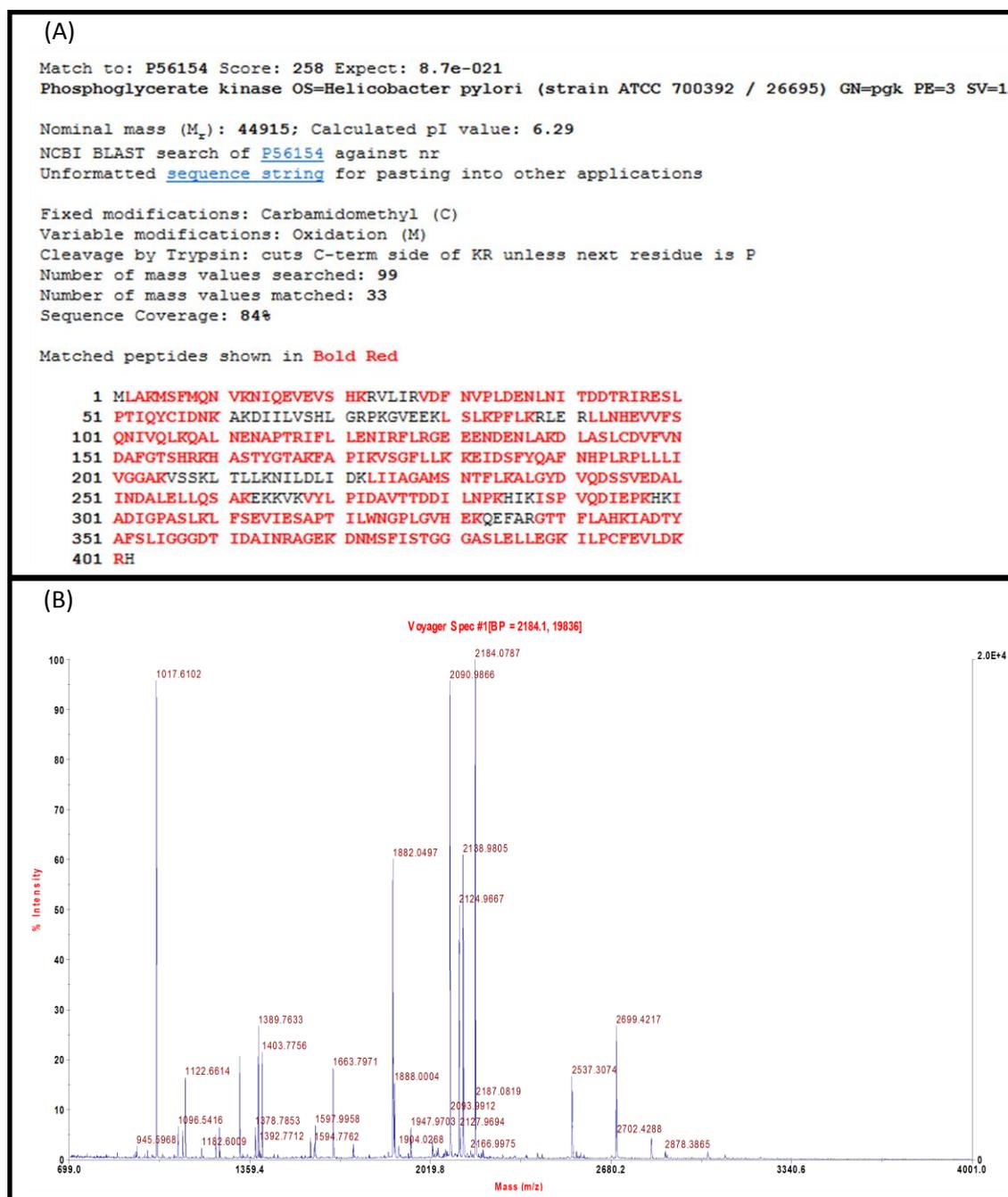
This study has emphasised the importance of testing synthetic lethality hypothesis of the gluconeogenic- glycolytic enzymes phosphoenolpyruvate synthase (hpPPSA) and phosphoglycerate kinase (hpPGK) based on computational findings using *in silico* double deletion mutagenesis to identify conditionally essential genes in *H. pylori*. The genetic findings reported here shed new light on *ppsA* potential essentiality to the viability of *H. pylori*. However, considerably more work will need to be done to verify *ppsA* essentiality by means of gene complementation analysis. Additionally, prior to this study *in silico* single deletion and transposon mutagenesis analysis studies reported that *pgk* is non-essential to the viability of the enteric pathogen. Nonetheless, it was difficult to verify these findings by allelic exchange mutagenesis due to wild type and mutated alleles variation results. Hence, this would be a fruitful area for further work to clarify *pgk* essentiality. In any case, *pgk* will remain as a valuable drug target due to contextual metabolic differences with humans. Should the essentiality of *pgk* alone be verified, this will lead to valuable insights on the central metabolism and drug targets in *H. pylori*.

Furthermore, the thesis has provided a deeper insight into bioinformatics, kinetics and structure of hpPPSA and hpPGK respectively. The homology modelling of hpPPSA has provided a framework for future discussion of genetic, mechanistic and structural strategies toward identifying therapies against this unique target due to the absence of corresponding orthologue in human. The crystal structure of the apo form of hpPGK was solved and compared to the fully closed form of human PGK, revealing that the

equivalent substrate and nucleotide binding residues along the cleft formed between the N-terminal and C-terminal domains during catalysis are highly conserved. Three sulphate ions bound along the cleft were identified for the first time, drawing the phosphate transfer path during catalysis. These findings may provide insights for future kinetic and/or structural studies to identify sulphate or phosphate based therapeutics. As *H. pylori* lives in the gastro-intestinal tract, it is exposed to and therefore vulnerable to drugs that are not absorbed through the human gut wall.

The studies described here demonstrated the wide diversity in bacterial metabolism, and why this variety has to be considered when a treatment for a specific pathogen is being developed.

## Appendix



**Figure A1** Peptide mass fingerprinting results of tag cleaved hpPGK. The output data is shown in: (A) Sequence panel shows the match between the theoretical and practical sequence. (B) Spectral information panel shows the mass of each peptide.

(A)

Match to: P00023 Score: 284 Expect: 1.3e-024  
His Tagged PGK - sequence provided by Aziz Al-Ethari 14/07/2014 - added to database by SYA

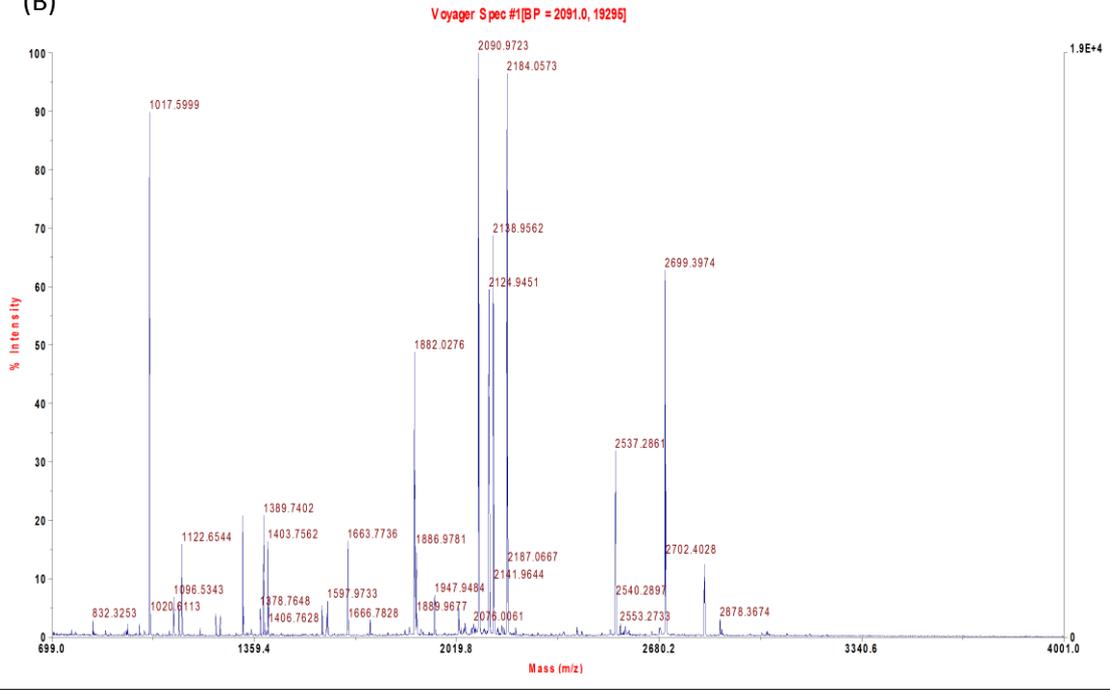
Nominal mass ( $M_r$ ): 47335; Calculated pI value: 6.30  
NCBI BLAST search of [P00023](#) against nr  
Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)  
Variable modifications: Oxidation (M)  
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P  
Number of mass values searched: 95  
Number of mass values matched: 36  
Sequence Coverage: 79%

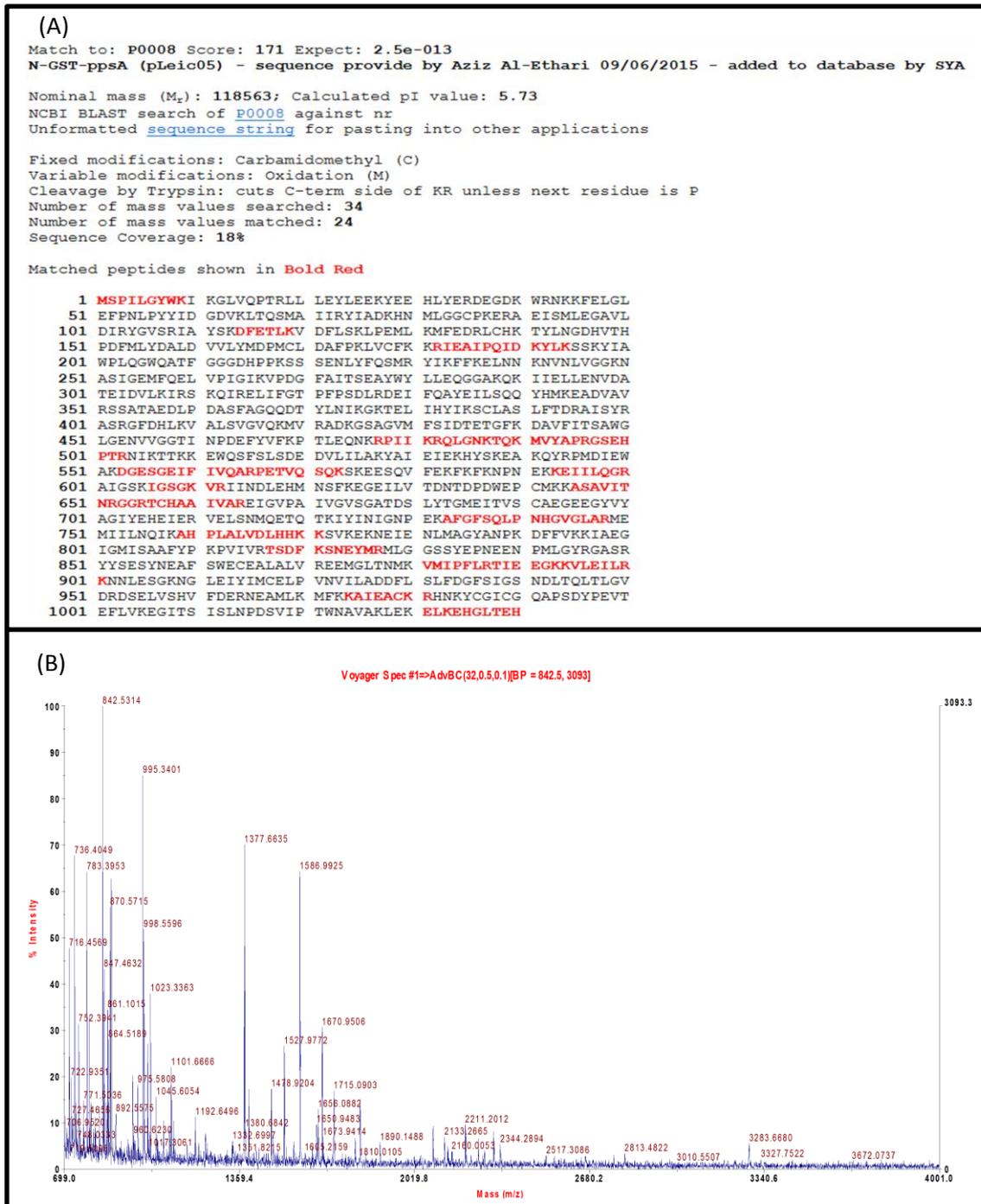
Matched peptides shown in **Red**

```
1 HHHHHHSSGV DLGTENLYFQ SMLAKMSFMQ NVKNIQEVEV SHKRVLIRVD
51 FNVPLDENLN ITDDTRIRES LPTIQYCIDN KAKDIILVSH LGRPQGVEEK
101 LSLKPFKRL ERLNHEVVF SQNIVQLRQA LNENAPTRIF LLENIRFLRG
151 EEENDENLAK DLASLCDVVF NDAFGTSHRK HASTYGTAKF APIKVSGLL
201 KKEIDSPYQA FNHPLRPLL IVGGAKVSSK LTLKRNILD IDKLIAGAM
251 SNTFLKALGY DVQDSSVEDA LINDALELQ SAKEKKVKVY LPIDAVTTDD
301 ILNPKHKIS PVQDIEPKHK IADIGPASLK LPSEVIESAP TILWNGPLGV
351 HEKQEFARGT TFLAKIADT YAFSLIGGD TIDAINRAGE KDNMSFISTG
401 GGASLELLEG KILPCFEVLD KRH
```

(B)



**Figure A2** Peptide mass fingerprinting results of N-His<sub>6</sub> hpPGK. The output data is shown in: (A) Sequence panel shows the match between the theoretical and practical sequence. (B) Spectral information panel shows the mass of each peptide.



**Figure A3** Peptide mass fingerprinting results of N- GST hpPPSA. The output data is shown in: (A) Sequence panel shows the match between the theoretical and practical sequence. (B) Spectral information panel shows the mass of each peptide.

## References

- ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, 66, 213-221.
- ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., ECHOLS, N., HEADD, J. J., HUNG, L.-W., JAIN, S., KAPRAL, G. J., GROSSE KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R. D., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2011. The Phenix software for automated determination of macromolecular structures. *Methods*, 55, 94-106.
- AFONINE, P. V., GROSSE-KUNSTLEVE, R. W., ECHOLS, N., HEADD, J. J., MORIARTY, N. W., MUSTYAKIMOV, M., TERWILLIGER, T. C., URZHUMTSEV, A., ZWART, P. H. & ADAMS, P. D. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, 68, 352-367.
- ALM, R. A., LING, L.-S. L., MOIR, D. T. & KING, B. L. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, 397, 176.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- AMINI, S. & TAVAZOIE, S. 2011. Antibiotics and the post-genome revolution. *Current Opinion in Microbiology*, 14, 513-518.
- ANSARI, S. & YAMAOKA, Y. 2017. Survival of *Helicobacter pylori* in gastric acidic territory. *Helicobacter*, 22.
- ASAKA, M. 2013. A new approach for elimination of gastric cancer deaths in Japan. *International journal of cancer*, 132, 1272-1276.
- ATHERTON, J. C. 2006. The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annu. Rev. Pathol. Mech. Dis.*, 1, 63-96.
- AUERBACH, G., HUBER, R., GRÄTTINGER, M., ZAISS, K., SCHURIG, H., JAENICKE, R. & JACOB, U. 1997. Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Structure*, 5, 1475-1483.
- AYNA, A. 2016. Glyceraldehyde-3-phosphate Dehydrogenase and Fructose-1, 6-bisphosphatase of the Enteric Pathogens *C. jejuni* and *H. pylori*. Department of Molecular and Cell Biology.

- BAAR, C., EPPINGER, M., RADDATZ, G., SIMON, J., LANZ, C., KLIMMEK, O., NANDAKUMAR, R., GROSS, R., ROSINUS, A. & KELLER, H. 2003. Complete genome sequence and analysis of *Wolinella succinogenes*. *Proceedings of the National Academy of Sciences*, 100, 11690-11695.
- BATTYE, T. G. G., KONTOGIANNIS, L., JOHNSON, O., POWELL, H. R. & LESLIE, A. G. 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, 67, 271-281.
- BAZZOLI, F., PORRO, G. B., MACONI, G., MOLTENI, M., POZZATO, P. & ZAGARI, R. 2002. Treatment of *Helicobacter pylori* infection. Indications and regimens: an update. *Digestive and Liver Disease*, 34, 70-83.
- BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.
- BERNSTEIN, B. E., MICHELS, P. & HOL, W. 1997. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature*, 385, 275-278.
- BERNSTEIN, B. E., WILLIAMS, D. M., BRESSI, J. C., KUHN, P., GELB, M. H., BLACKBURN, G. M. & HOL, W. G. 1998. A bisubstrate analog induces unexpected conformational changes in phosphoglycerate kinase from *Trypanosoma brucei*. *Journal of molecular biology*, 279, 1137-1148.
- BIZZOZERO, G. 1893. Ueber die schlauchförmigen Drüsen des Magendarmkanals und die Beziehungen ihres Epithels zu dem Oberflächenepithel der Schleimhaut Dritte Mittheilung. *Archiv für mikroskopische Anatomie*, 42, 82-152.
- BÖHM, G. 1997. CDNN: CD spectra deconvolution software version 2.1. University of Halle-Wittenberg, Halle, Germany.
- BRAGG, W. H. & BRAGG, W. L. 1913. The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 88, 428-438.
- BRECKAN, R. K., PAULSSEN, E. J., ASFELDT, A. M., KVAMME, J. M., STRAUME, B. & FLORHOLMEN, J. 2016. The All-Age Prevalence of *Helicobacter pylori* Infection and Potential Transmission Routes. A Population-Based Study. *Helicobacter*, 21, 586-595.
- BREUREC, S., GUILLARD, B., HEM, S., BRISSE, S., DIEYE, F. B., HUERRE, M., OUNG, C., RAYMOND, J., TAN, T. S. & THIBERGE, J.-M. 2011. Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. *PloS one*, 6, e22058.
- BROWN, L. M. 2000. *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiologic reviews*, 22, 283-297.

- BRÜNGER, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355, 472.
- BURY-MONÉ, S., KAAKOUSH, N. O., ASENCIO, C., MÉGRAUD, F., THIBONNIER, M., DE REUSE, H. & MENDZ, G. L. 2006. Is *Helicobacter pylori* a true microaerophile? *Helicobacter*, 11, 296-303.
- CAMPBELL, B. J., ENGEL, A. S., PORTER, M. L. & TAKAI, K. 2006. The versatile  $\epsilon$ -proteobacteria: key players in sulphidic habitats. *Nature Reviews Microbiology*, 4, 458-468.
- CASPI, R., ALTMAN, T., DREHER, K., FULCHER, C. A., SUBHRAVETI, P., KESELER, I. M., KOTHARI, A., KRUMMENACKER, M., LATENDRESSE, M., MUELLER, L. A., ONG, Q., PALEY, S., PUJAR, A., SHEARER, A. G., TRAVERS, M., WEERASINGHE, D., ZHANG, P. & KARP, P. D. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 40, D742-D753.
- CELLI, J. P., TURNER, B. S., AFDHAL, N. H., KEATES, S., GHIRAN, I., KELLY, C. P., EWOLDT, R. H., MCKINLEY, G. H., SO, P. & ERRAMILI, S. 2009. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proceedings of the National Academy of Sciences*, 106, 14321-14326.
- CHALK, P. A., ROBERTS, A. D. & BLOWS, W. M. 1994. Metabolism of pyruvate and glucose by intact cells of *Helicobacter pylori* studied by  $^{13}\text{C}$  NMR spectroscopy. *Microbiology*, 140, 2085-2092.
- CHALKER, A. F., MINEHART, H. W., HUGHES, N. J., KORETKE, K. K., LONETTO, M. A., BRINKMAN, K. K., WARREN, P. V., LUPAS, A., STANHOPE, M. J., BROWN, J. R. & HOFFMAN, P. S. 2001. Systematic identification of selective essential genes in *Helicobacter pylori* by genome prioritization and allelic replacement mutagenesis. *J Bacteriol*, 183, 1259-68.
- CHAO, Y., PATNAIK, R., ROOF, W., YOUNG, R. & LIAO, J. 1993. Control of gluconeogenic growth by pps and pck in *Escherichia coli*. *Journal of bacteriology*, 175, 6939-6944.
- CHEN, Z., ZHOU, Q. & GE, R. 2012. Inhibition of fumarase by bismuth(III): implications for the tricarboxylic acid cycle as a potential target of bismuth drugs in *Helicobacter pylori*. *Biometals*, 25, 95-102.
- CHU, C. H., LAI, Y. J., HUANG, H. & SUN, Y. J. 2008. Kinetic and structural properties of triosephosphate isomerase from *Helicobacter pylori*. *Proteins*, 71, 396-406.
- CLIFF, M. J., BOWLER, M. W., VARGA, A., MARSTON, J. P., SZABÓ, J., HOUNSLOW, A. M., BAXTER, N. J., BLACKBURN, G. M., VAS, M. & WALTHO, J. P. 2010. Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *Journal of the American Chemical Society*, 132, 6507-6516.

- COOPER, R. & KORNBERG, H. 1965. Net formation of phosphoenolpyruvate from pyruvate by *Escherichia coli*. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 104, 618-620.
- COOPER, R. & KORNBERG, H. 1967. The mechanism of the phosphoenolpyruvate synthase reaction. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 141, 211-213.
- CORTHÉSY-THEULAZ, I. E., BERGONZELLI, G. E., HENRY, H., BACHMANN, D., SCHORDERET, D. F., BLUM, A. L. & ORNSTON, L. N. 1997. Cloning and Characterization of *Helicobacter pylori* Succinyl CoA:Acetoacetate CoA-transferase, a Novel Prokaryotic Member of the CoA-transferase Family. *Journal of Biological Chemistry*, 272, 25659-25667.
- COSME, A., MONTES, M., IBARRA, B., TAMAYO, E., ALONSO, H., MENDARTE, U., LIZASOAN, J., HERREROS-VILLANUEVA, M. & BUJANDA, L. 2017. Antimicrobial susceptibility testing before first-line treatment for *Helicobacter pylori* infection in patients with dual or triple antibiotic resistance. *World J Gastroenterol*, 23, 3367-3373.
- DAVIS, I. W., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2004. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic acids research*, 32, W615-W619.
- DELANO, W. L. 2002. The PyMOL molecular graphics system. <http://pymol.org>.
- DEMIRAY-GÜRBÜZ, E., YILMAZ, Ö., OLIVARES, A. Z., GÖNEN, C., SARIOĞLU, S., SOYTÜRK, M., TÜMER, S., ALTUNGÖZ, O., ŞİMŞEK, İ. & PEREZ PEREZ, G. I. 2017. Rapid identification of *Helicobacter pylori* and assessment of clarithromycin susceptibility from clinical specimens using FISH. *The Journal of Pathology: Clinical Research*, 3, 29-37.
- DENG, J., SU, S., LIN, X., HASSETT, D. J. & LU, L. J. 2013. A Statistical Framework for Improving Genomic Annotations of Prokaryotic Essential Genes. *PLOS ONE*, 8, e58178.
- 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. & TRUST, T. J. 1999. *Helicobacter pylori* Physiology Predicted from Genomic Comparison of Two Strains. *Microbiology and Molecular Biology Reviews*, 63, 675-707.
- ELLIOTT, P. R. 2009. Structure and function of enteric pathogen glyceraldehyde-3-phosphate dehydrogenases. University of Leicester.
- EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D*, 66, 486-501.

- EPPINGER, M., BAAR, C., RADDATZ, G., HUSON, D. H. & SCHUSTER, S. C. 2004. Comparative analysis of four Campylobacterales. *Nature reviews. Microbiology*, 2, 872.
- ERNST, P. B. & GOLD, B. D. 2000. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annual Reviews in Microbiology*, 54, 615-640.
- ESHRAGHIAN, A. 2014. Epidemiology of *Helicobacter pylori* infection among the healthy population in Iran and countries of the Eastern Mediterranean Region: a systematic review of prevalence and risk factors. *World journal of gastroenterology: WJG*, 20, 17618.
- ESSA, A., KRAMER, J., GRAHAM, D. & TREIBER, G. 2009. Meta-analysis: four-drug, three-antibiotic, non-bismuth-containing “concomitant therapy” versus triple therapy for *Helicobacter pylori*, 109-18.
- EUSEBI, L. H., ZAGARI, R. M. & BAZZOLI, F. 2014. Epidemiology of *Helicobacter pylori* infection. *Helicobacter*, 19 Suppl 1, 1-5.
- EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62, 72-82.
- EVANS, P. & MCCOY, A. 2008. An introduction to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography*, 64, 1-10.
- EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69, 1204-1214.
- FALUSH, D., WIRTH, T., LINZ, B., PRITCHARD, J. K., STEPHENS, M., KIDD, M., BLASER, M. J., GRAHAM, D. Y., VACHER, S. & PEREZ-PEREZ, G. I. 2003. Traces of human migrations in *Helicobacter pylori* populations. *science*, 299, 1582-1585.
- FERNANDES, A. P., NELSON, K. & BEVERLEY, S. M. 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. *Proceedings of the National Academy of Sciences*, 90, 11608-11612.
- FIFIS, T. & SCOPES, R. K. 1978. Purification of 3-phosphoglycerate kinase from diverse sources by affinity elution chromatography. *Biochemical Journal*, 175, 311-319.
- FISCHER, W., PÜLS, J., BUHRDORF, R., GEBERT, B., ODENBREIT, S. & HAAS, R. 2001. Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Molecular Microbiology*, 42, 1337-1348.
- FOCK, K. M., KATELARI, P., SUGANO, K., ANG, T. L., HUNT, R., TALLEY, N. J., LAM, S. K., XIAO, S. D., TAN, H. J. & WU, C. Y. 2009. Second Asia–Pacific consensus guidelines for *helicobacter pylori* infection. *Journal of gastroenterology and hepatology*, 24, 1587-1600.

- FONVIELLE, M., COINÇON, M., DAHER, R., DESBENOIT, N., KOSIERADZKA, K., BARILONE, N., GICQUEL, B., SYGUSCH, J., JACKSON, M. & THERISOD, M. 2008. Synthesis and biochemical evaluation of selective inhibitors of class II fructose bisphosphate aldolases: towards new synthetic antibiotics. *Chemistry—A European Journal*, 14, 8521-8529.
- FOX, J. 2002. The non-H pylori helicobacters: their expanding role in gastrointestinal and systemic diseases. *Gut*, 50, 273-283.
- FRANK, J., DINGEMANSE, C., SCHMITZ, A. M., VOSSEN, R. H., VAN OMMEN, G.-J. B., DEN DUNNEN, J. T., ROBANUS-MAANDAG, E. C. & ANVAR, S. Y. 2015. The complete genome sequence of the murine pathobiont helicobacter typhlonius. *Frontiers in microbiology*, 6.
- FRENCK, R. W. & CLEMENS, J. 2003. Helicobacter in the developing world. *Microbes and infection*, 5, 705-713.
- GAO, W., CHENG, H., HU, F., LI, J., WANG, L., YANG, G., XU, L. & ZHENG, X. 2010. The evolution of *Helicobacter pylori* antibiotics resistance over 10 years in Beijing, China. *Helicobacter*, 15, 460-466.
- GARRITY, G. M., BELL, J. A. & LILBURN, T. 2003. Taxonomic Outline of the Prokaryotes. Bergey's Manual of Systematic Bacteriology. Release 4.0. Springer-Verlag. New York.
- GASKIN, D. J. H. & VAN VLIET, A. H. M. 2010. Random Mutagenesis Strategies for *Campylobacter* and *Helicobacter* Species. In: BRAMAN, J. (ed.) *In Vitro Mutagenesis Protocols: Third Edition*. Totowa, NJ: Humana Press.
- GE, Z., JIANG, Q., KALISIAK, M. S. & TAYLOR, D. E. 1997. Cloning and functional characterization of *Helicobacter pylori* fumarate reductase operon comprising three structural genes coding for subunits C, A and B. *Gene*, 204, 227-234.
- GEIS, G., LEYING, H., SUERBAUM, S., MAI, U. & OPFERKUCH, W. 1989. Ultrastructure and chemical analysis of *Campylobacter pylori* flagella. *Journal of clinical microbiology*, 27, 436-441.
- GHOTASLOU, R., LEYLABADLO, H. E. & ASL, Y. M. 2015. Prevalence of antibiotic resistance in *Helicobacter pylori*: A recent literature review. *World journal of methodology*, 5, 164.
- GOODWIN, A., KERSULYTE, D., SISSON, G., VELDHUYZEN VAN ZANTEN, S. J. O., BERG, D. E. & HOFFMAN, P. S. 1998. Metronidazole resistance in *Helicobacter pylori* is due to null mutations in a gene (rdxA) that encodes an oxygen-insensitive NADPH nitroreductase. *Molecular Microbiology*, 28, 383-393.
- GOODWIN, C. S., ARMSTRONG, J. A., CHILVERS, T., PETERS, M., COLLINS, M. D., SLY, L., MCCONNELL, W. & HARPER, W. E. S. 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb.

nov. and *Helicobacter mustelae* comb. nov., Respectively. *International Journal of Systematic and Evolutionary Microbiology*, 39, 397-405.

- GRAHAM, D. Y., LEE, Y. C. & WU, M. S. 2014. Rational *Helicobacter pylori* therapy: evidence-based medicine rather than medicine-based evidence. *Clinical Gastroenterology and Hepatology*, 12, 177-186. e3.
- GRAHAM, D. Y., LEW, G. M., KLEIN, P. D. & ET AL. 1992. Effect of treatment of *Helicobacter pylori* infection on the long-term recurrence of gastric or duodenal ulcer: A randomized, controlled study. *Annals of Internal Medicine*, 116, 705-708.
- GÜELL, O., SAGUÉS, F. & SERRANO, M. Á. 2014. Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol*, 10, e1003637.
- HEAP, J. T., EHSAAN, M., COOKSLEY, C. M., NG, Y.-K., CARTMAN, S. T., WINZER, K. & MINTON, N. P. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research*, 40, e59-e59.
- HERZBERG, O., CHEN, C., KAPADIA, G., MCGUIRE, M., CARROLL, L. J., NOH, S. J. & DUNAWAY-MARIANO, D. 1996. Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proceedings of the National Academy of Sciences*, 93, 2652-2657.
- HO, M., BRAMSON, H. N., HANSEN, D. E., KNOWLES, J. R. & KAISER, E. 1988. Stereochemical course of the phospho group transfer catalyzed by cAMP-dependent protein kinase. *Journal of the American Chemical Society*, 110, 2680-2681.
- HOFFMAN, P. S., GOODWIN, A., JOHNSEN, J., MAGEE, K. & VAN ZANTEN, S. V. 1996. 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*, 178, 4822-4829.
- HORIKI, N., OMATA, F., UEMURA, M., SUZUKI, S., ISHII, N., IIZUKA, Y., FUKUDA, K., FUJITA, Y., KATSURAHARA, M. & ITO, T. 2009. Annual change of primary resistance to clarithromycin among *Helicobacter pylori* isolates from 1996 through 2008 in Japan. *Helicobacter*, 14, 438-442.
- HSU, P.-I., LIN, P.-C. & GRAHAM, D. Y. 2015. Hybrid therapy for *Helicobacter pylori* infection: a systemic review and meta-analysis. *World journal of gastroenterology*, 21, 12954.
- HSU, P.-I. & WU, D.-C. 2016. 315 Reverse Hybrid Therapy Achieves a Similar Eradication Rate As Hybrid Therapy in the Treatment of *Helicobacter pylori* Infection. *Gastroenterology*, 150, S73.

- HUANG, C.-C., TSAI, K.-W., TSAI, T.-J. & HSU, P.-I. 2017. Update on the first-line treatment for *Helicobacter pylori* infection-a continuing challenge from an old enemy. *Biomarker research*, 5, 23.
- HUGHES, N. J., CLAYTON, C. L., CHALK, P. A. & KELLY, D. J. 1998. *Helicobacter pylori* porCDAB and oorDABC Genes Encode Distinct Pyruvate:Flavodoxin and 2-Oxoglutarate:Acceptor Oxidoreductases Which Mediate Electron Transport to NADP. *Journal of Bacteriology*, 180, 1119-1128.
- IJEOMA, O., HOLLOWELL, H. N., BODNAR, M. A. & BRITT, B. M. 2008. Thermodynamic analysis of the non-denaturational conformational change of baker's yeast phosphoglycerate kinase at 24° C. *Archives of biochemistry and biophysics*, 478, 206-211.
- IMANAKA, H., YAMATSU, A., FUKUI, T., ATOMI, H. & IMANAKA, T. 2006. Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol Microbiol*, 61, 898-909.
- JAWORSKI, W. 1889. Podrecznik Chorob zoladka. *Wydawnictwa Dziel Lekarskich Polskich*, 32.
- JENKS, P. J., CHEVALIER, C., ECOBICHON, C. & LABIGNE, A. 2001. Identification of nonessential *Helicobacter pylori* genes using random mutagenesis and loop amplification. *Res Microbiol*, 152, 725-34.
- JOOSTEN, R. P., LONG, F., MURSHUDOV, G. N. & PERRAKIS, A. 2014. The PDB\_REDO server for macromolecular structure model optimization. *IUCr*, 1, 213-220.
- KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28, 27-30.
- KATHER, B., STINGL, K., VAN DER REST, M. E., ALTENDORF, K. & MOLENAAR, D. 2000. Another Unusual Type of Citric Acid Cycle Enzyme in *Helicobacter pylori*: the Malate:Quinone Oxidoreductase. *Journal of Bacteriology*, 182, 3204-3209.
- KATOH, K. & STANDLEY, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30, 772-780.
- KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10, 845-858.
- KHALIFA, M. M., SHARAF, R. R. & AZIZ, R. K. 2010. *Helicobacter pylori*: a poor man's gut pathogen? *Gut pathogens*, 2, 2.
- KIBBE, W. A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35, W43-W46.

- KIRINO, H., AOKI, M., AOSHIMA, M., HAYASHI, Y., OHBA, M., YAMAGISHI, A., WAKAGI, T. & OSHIMA, T. 1994. Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus*. *European Journal of Biochemistry*, 220, 275-281.
- KLEYWEGT, G. J. & JONES, T. A. 1998. Databases in protein crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 54, 1119-1131.
- KNOWLES, J. R. 1980. Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry*, 49, 877-919.
- KONNO, M., FUJII, N., YOKOTA, S.-I., SATO, K., TAKAHASHI, M., SATO, K., MINO, E. & SUGIYAMA, T. 2005. Five-year follow-up study of mother-to-child transmission of *Helicobacter pylori* infection detected by a random amplified polymorphic DNA fingerprinting method. *Journal of clinical microbiology*, 43, 2246-2250.
- KOTSIKOROU, E., SAHOTA, G. & OLDFIELD, E. 2006. Bisphosphonate inhibition of phosphoglycerate kinase: Quantitative structure-activity relationship and pharmacophore modeling investigation. *Journal of medicinal chemistry*, 49, 6692-6703.
- KRIENITZ, W. 1906. Ueber das Auftreten von Spirochäten verschiedener Form im Mageninhalt bei Carcinoma ventriculi. *DMW-Deutsche Medizinische Wochenschrift*, 32, 872-872.
- KUIPERS, E. 1999. Exploring the link between *Helicobacter pylori* and gastric cancer. *Alimentary pharmacology & therapeutics*, 13, 3-11.
- KUIPERS, E., PENA, A., FESTEN, H., MEUWISSEN, S., UYTERLINDE, A., ROOSENDAAL, R., PALS, G. & NELIS, G. 1995a. Long-term sequelae of *Helicobacter pylori* gastritis. *The Lancet*, 345, 1525-1528.
- KUIPERS, E., THIJIS, J. & FESTEN, H. 1995b. The prevalence of *Helicobacter pylori* in peptic ulcer disease. *Alimentary pharmacology & therapeutics*, 9, 59-69.
- KUIPERS, E. J., LUNDELL, L., KLINKENBERG-KNOL, E. C., HAVU, N., FESTEN, H. P., LIEDMAN, B., LAMERS, C. B., JANSEN, J. B., DALENBÄCK, J. & SNEL, P. 1996. Atrophic gastritis and *Helicobacter pylori* infection in patients with reflux esophagitis treated with omeprazole or fundoplication. *New England Journal of Medicine*, 334, 1018-1022.
- KUSTERS, J., GERRITS, M., VAN STRIJP, J. & VANDENBROUCKE-GRAULS, C. 1997. Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. *Infection and immunity*, 65, 3672-3679.
- KUSTERS, J. G., VAN VLIET, A. H. & KUIPERS, E. J. 2006. Pathogenesis of *Helicobacter pylori* infection. *Clin Microbiol Rev*, 19, 449-90.

- LABIGNE, A. & JENKS, P. 2001. Mutagenesis. In: MOBLEY, H., MENDZ, G. & HAZEL, S. (eds.) *Helicobacter pylori: Physiology and Genetics*. Washington (DC): ASM Press.
- LANGFORD, M. L., ZABALETA, J., OCHOA, A. C., TESTERMAN, T. L. & MCGEE, D. J. 2006. In vitro and in vivo complementation of the *Helicobacter pylori* arginase mutant using an intergenic chromosomal site. *Helicobacter*, 11, 477-93.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26, 283-291.
- LEE, J. H., IM, Y. J., BAE, J., KIM, D., KIM, M. K., KANG, G. B., LEE, D. S. & EOM, S. H. 2006. Crystal structure of *Thermus caldophilus* phosphoglycerate kinase in the open conformation. *Biochem Biophys Res Commun*, 350, 1044-9.
- LEJA, M., AXON, A. & BRENNER, H. 2016. Epidemiology of *Helicobacter pylori* infection. *Helicobacter*, 21 Suppl 1, 3-7.
- LINZ, B., BALLOUX, F., MOODLEY, Y., MANICA, A., LIU, H., ROUMAGNAC, P., FALUSH, D., STAMER, C., PRUGNOLLE, F., VAN DER MERWE, S. W., YAMAOKA, Y., GRAHAM, D. Y., PEREZ-TRALLERO, E., WADSTROM, T., SUERBAUM, S. & ACHTMAN, M. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, 445, 915-918.
- MALATY, H. M., EL-KASABANY, A., GRAHAM, D. Y., MILLER, C. C., REDDY, S. G., SRINIVASAN, S. R., YAMAOKA, Y. & BERENSON, G. S. 2002. Age at acquisition of *Helicobacter pylori* infection: a follow-up study from infancy to adulthood. *The Lancet*, 359, 931-935.
- MALFERTHEINER, P., MÉGRAUD, F., O'MORAIN, C., HUNGIN, A., JONES, R., AXON, A., GRAHAM, D. & TYTGAT, G. 2002. Current concepts in the management of *Helicobacter pylori* infection—The Maastricht 2-2000 Consensus Report. *Alimentary pharmacology & therapeutics*, 16, 167-180.
- MANSOUR, K. B., BURUCOA, C., ZRIBI, M., MASMOUDI, A., KAROUI, S., KALLEL, L., CHOUAIB, S., MATRI, S., FEKIH, M. & ZARROUK, S. 2010. Primary resistance to clarithromycin, metronidazole and amoxicillin of *Helicobacter pylori* isolated from Tunisian patients with peptic ulcers and gastritis: a prospective multicentre study. *Annals of clinical microbiology and antimicrobials*, 9, 22.
- MARAIS, A., MENDZ, G. L., HAZELL, S. L. & MEGRAUD, F. 1999. Metabolism and genetics of *Helicobacter pylori*: the genome era. *Microbiol Mol Biol Rev*, 63, 642-74.
- MARSHALL, B. & GOODWIN, C. 1987. Notes: Revised Nomenclature of *Campylobacter pyloridis*. *International Journal of Systematic and Evolutionary Microbiology*, 37, 68-68.

- MARSHALL, B., ROYCE, H., ANNEAR, D., GOODWIN, C., PEARMAN, J., WARREN, J. & ARMSTRONG, J. 1984. Original isolation of *Campylobacter pyloridis* from human gastric mucosa. *Microbios Letters*, 25, 83-88.
- MARSHALL, B. & WARREN, J. R. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *The Lancet*, 323, 1311-1315.
- MARSHALL, B. J., ARMSTRONG, J. A., MCGECHIE, D. B. & GLANCY, R. J. 1985. Attempt to fulfil Koch's postulates for pyloric *Campylobacter*. *The Medical Journal of Australia*, 142, 436.
- MARTÍ-RENOM, M. A., STUART, A. C., FISER, A., SÁNCHEZ, R., AND, F. M. & ŠALI, A. 2000. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure*, 29, 291-325.
- MARTÍNEZ, L. E., HARDCASTLE, J. M., WANG, J., PINCUS, Z., TSANG, J., HOOVER, T. R., BANSIL, R. & SALAMA, N. R. 2016. *Helicobacter pylori* strains vary cell shape and flagellum number to maintain robust motility in viscous environments. *Molecular microbiology*, 99, 88-110.
- MATTHEWS, B. W. 1968. Solvent content of protein crystals. *Journal of molecular biology*, 33, 491-497.
- MCCORMICK, N. E. & JAKEMAN, D. L. 2015. On the mechanism of phosphoenolpyruvate synthetase (PEPs) and its inhibition by sodium fluoride: potential magnesium and aluminum fluoride complexes of phosphoryl transfer. *Biochemistry and Cell Biology*, 93, 236-240.
- MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40, 658-674.
- MEGRAUD, F. & LAMOULIATTE, H. 2003. the treatment of refractory *Helicobacter pylori* infection. *Alimentary pharmacology & therapeutics*, 17, 1333-1343.
- MENDZ, G. & HAZELL, S. 1993. Glucose phosphorylation in *Helicobacter pylori*. *Archives of biochemistry and biophysics*, 300, 522-525.
- MENDZ, G. L. & HAZELL, S. L. 1991. Evidence for a pentose phosphate pathway in *Helicobacter pylori*. *FEMS Microbiology Letters*, 84, 331-336.
- MENDZ, G. L. & HAZELL, S. L. 1996. The urea cycle of *Helicobacter pylori*. *Microbiology*, 142, 2959-2967.
- MENDZ, G. L., HAZELL, S. L. & BURNS, B. P. 1994a. The Entner-Doudoroff pathway in *Helicobacter pylori*. *Archives of biochemistry and biophysics*, 312, 349-356.
- MENDZ, G. L., HAZELL, S. L. & SRINIVASAN, S. 1995. Fumarate Reductase: A Target for Therapeutic Intervention Against *Helicobacter pylori*. *Archives of Biochemistry and Biophysics*, 321, 153-159.

- MENDZ, G. L., HAZELL, S. L. & VAN GORKOM, L. 1994b. Pyruvate metabolism in *Helicobacter pylori*. *Archives of Microbiology*, 162, 187-192.
- MENTIS, A., LEHOURS, P. & MEGRAUD, F. 2015. Epidemiology and Diagnosis of *Helicobacter pylori* infection. *Helicobacter*, 20 Suppl 1, 1-7.
- MITUI, M., PATEL, A., LEOS, N. K., DOERN, C. D. & PARK, J. Y. 2014. Novel *Helicobacter pylori* sequencing test identifies high rate of clarithromycin resistance. *Journal of pediatric gastroenterology and nutrition*, 59, 6-9.
- MOMTAZ, H., DABIRI, H., SOUOD, N. & GHOLAMI, M. 2014. Study of *Helicobacter pylori* genotype status in cows, sheep, goats and human beings. *BMC gastroenterology*, 14, 61.
- MONTANO, V., DIDELOT, X., FOLL, M., LINZ, B., REINHARDT, R., SUERBAUM, S., MOODLEY, Y. & JENSEN, J. D. 2015. Worldwide Population Structure, Long-Term Demography, and Local Adaptation of *Helicobacter pylori*. *Genetics*, 200, 947-963.
- MOODLEY, Y., LINZ, B., BOND, R. P., NIEUWOUDT, M., SOODYALL, H., SCHLEBUSCH, C. M., BERNHÖFT, S., HALE, J., SUERBAUM, S. & MUGISHA, L. 2012. Age of the association between *Helicobacter pylori* and man. *PLoS pathogens*, 8, e1002693.
- MORRIS, A. & NICHOLSON, G. 1987. Ingestion of *Campylobacter pyloridis* causes gastritis and raised fasting gastric pH. *American Journal of Gastroenterology*, 82.
- MORRIS, A. J., ALI, M. R., NICHOLSON, G. I., PEREZ-PEREZ, G. I. & BLASER, M. J. 1991. Long-term follow-up of voluntary ingestion of *Helicobacter pylori*. *Annals of internal medicine*, 114, 662-663.
- MUKHERJEE, K., GHOSH, S., RAY, M. & RAY, S. 2002. Purification and characterization of 3-phosphoglycerate kinase from Ehrlich ascites carcinoma cells. *Indian Journal of Biochemistry and Biophysics*, 39, 332-341.
- MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 67, 355-367.
- NACHAMKIN, I., ALLOS, B. M. & HO, T. 1998. *Campylobacter* species and Guillain-Barre syndrome. *Clinical microbiology reviews*, 11, 555-567.
- NAGATA, K., NAGATA, Y., SATO, T., FUJINO, M. A., NAKAJIMA, K. & TAMURA, T. 2003. L-Serine, d- and l-proline and alanine as respiratory substrates of *Helicobacter pylori*: correlation between in vitro and in vivo amino acid levels. *Microbiology*, 149, 2023-2030.
- NAKAMURA, H., YOSHIYAMA, H., TAKEUCHI, H., MIZOTE, T., OKITA, K. & NAKAZAWA, T. 1998a. Urease plays an important role in the chemotactic motility of

- Helicobacter pylori* in a viscous environment. *Infection and immunity*, 66, 4832-4837.
- NAKAMURA, S., AOYAGI, K., FURUSE, M., SUEKANE, H., MATSUMOTO, T., YAO, T., SAKAI, Y., FUCHIGAMI, T., YAMAMOTO, I. & TSUNEYOSHI, M. 1998b. B-cell monoclonality precedes the development of gastric MALT lymphoma in *Helicobacter pylori*-associated chronic gastritis. *The American journal of pathology*, 152, 1271.
- NARINDRASORASAK, S. & BRIDGER, W. 1977. Phosphoenolpyruvate synthetase of *Escherichia coli*: molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate. *Journal of Biological Chemistry*, 252, 3121-3127.
- NWOKOLO, C. U., MISTRY, P. & POUNDER, R. E. 1990. The absorption of bismuth and salicylate from oral doses of Pepto-Bismol (bismuth salicylate). *Alimentary Pharmacology & Therapeutics*, 4, 163-169.
- OPPERDOES, F. R. 1987. Compartmentation of carbohydrate metabolism in trypanosomes. *Annual Reviews in Microbiology*, 41, 127-151.
- OPPERDOES, F. R. & BORST, P. 1977. Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: The glycosome. *FEBS letters*, 80, 360-364.
- OWEN, R. 1998. *Helicobacter*-species classification and identification. *British medical bulletin*, 54, 17-30.
- PAL, B., PYBUS, B., MUCCIO, D. D. & CHATTOPADHYAY, D. 2004. Biochemical characterization and crystallization of recombinant 3-phosphoglycerate kinase of *Plasmodium falciparum*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1699, 277-280.
- PAUL, W., LANE, M. C. & PORWOLLIK, S. 2000. *Helicobacter pylori* motility. *Microbes and infection*, 2, 1207-1214.
- PEGORARO, B. & LEE, C.-Y. 1978. Purification and characterization of two isozymes of 3-phosphoglycerate kinase from the mouse. *Biochimica et Biophysica Acta (BBA)-Enzymology*, 522, 423-433.
- PELETEIRO, B., BASTOS, A., FERRO, A. & LUNET, N. 2014. Prevalence of *Helicobacter pylori* infection worldwide: a systematic review of studies with national coverage. *Digestive diseases and sciences*, 59, 1698-1709.
- PICKOVER, C. A., MCKAY, D. B., ENGELMAN, D. M. & STEITZ, T. A. 1979. Substrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. *Journal of Biological Chemistry*, 254, 11323-11329.
- PILKIS, S. J. & GRANNER, D. 1992. Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis. *Annual review of physiology*, 54, 885-909.

- PITSON, S. M., MENDZ, G. L., SRINIVASAN, S. & HAZELL, S. L. 1999. The tricarboxylic acid cycle of *Helicobacter pylori*. *European Journal of Biochemistry*, 260, 258-267.
- PLUMMER, M., FRANCESCHI, S., VIGNAT, J., FORMAN, D. & DE MARTEL, C. 2015. Global burden of gastric cancer attributable to *Helicobacter pylori*. *International journal of cancer*, 136, 487-490.
- PSAKIS, G., SAIDIJAM, M., SHIBAYAMA, K., POLACZEK, J., BETTANEY, K. E., BALDWIN, J. M., BALDWIN, S. A., HOPE, R., ESSEN, L. O. & ESSENBERG, R. C. 2009. The sodium-dependent d-glucose transport protein of *Helicobacter pylori*. *Molecular microbiology*, 71, 391-403.
- QI, X., LIN, W., MA, M., WANG, C., HE, Y., HE, N., GAO, J., ZHOU, H., XIAO, Y., WANG, Y. & ZHANG, P. 2016. Structural basis of rifampin inactivation by rifampin phosphotransferase. *Proceedings of the National Academy of Sciences*, 113, 3803-3808.
- RADER, B. A., WREDEN, C., HICKS, K. G., SWEENEY, E. G., OTTEMANN, K. M. & GUILLEMIN, K. 2011. *Helicobacter pylori* perceives the quorum-sensing molecule AI-2 as a chemorepellent via the chemoreceptor TlpB. *Microbiology*, 157, 2445-2455.
- RAO, D. R. & OESPER, P. 1961. Purification and properties of muscle phosphoglycerate kinase. *Biochemical Journal*, 81, 405-411.
- REDDY, G. K. & WENDISCH, V. F. 2014. Characterization of 3-phosphoglycerate kinase from *Corynebacterium glutamicum* and its impact on amino acid production. *BMC Microbiol*, 14, 54.
- RESENDE, T., CORREIA, D. M., ROCHA, M. & ROCHA, I. 2013. Re-annotation of the genome sequence of *Helicobacter pylori* 26695. *J Integr Bioinform*, 10, 233.
- REYRAT, J.-M., PELICIC, V., GICQUEL, B. & RAPPUOLI, R. 1998. Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis. *Infection and Immunity*, 66, 4011-4017.
- ROCHA, I., FÖRSTER, J. & NIELSEN, J. 2008. Design and application of genome-scale reconstructed metabolic models. *Microbial Gene Essentiality: Protocols and Bioinformatics*, 409-431.
- ROMÁN-ROMÁN, A., GIONO-CEREZO, S., CAMORLINGA-PONCE, M., MARTÍNEZ-CARRILLO, D. N., LOAIZA-LOEZA, S. & FERNÁNDEZ-TILAPA, G. 2013. vacA genotypes of *Helicobacter pylori* in the oral cavity and stomach of patients with chronic gastritis and gastric ulcer. *Enfermedades infecciosas y microbiología clinica*, 31, 130-135.
- ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172.

- ROSSMANN, M. G. & VAN BEEK, C. G. 1999. Data processing. *Acta Crystallographica Section D: Biological Crystallography*, 55, 1631-1640.
- ROUSTAN, C., FATTOUM, A., JEANNEAU, R. & PRADEL, L. A. 1980. Yeast 3-phosphoglycerate kinase: sulfate and substrate binding, their effect on the conformational state of the enzyme. *Biochemistry*, 19, 5168-5175.
- RUPP, B. 2009. *Biomolecular crystallography: principles, practice, and application to structural biology*, Garland Science.
- RUSSO KRAUSS, I., MERLINO, A., VERGARA, A. & SICA, F. 2013. An overview of biological macromolecule crystallization. *International journal of molecular sciences*, 14, 11643-11691.
- SALAMA, N. R., SHEPHERD, B. & FALKOW, S. 2004. Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *Journal of bacteriology*, 186, 7926-7935.
- SCHILLING, C. H., COVERT, M. W., FAMILI, I., CHURCH, G. M., EDWARDS, J. S. & PALSSON, B. O. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol*, 184, 4582-93.
- SCHWARZ, S., MORELLI, G., KUSECEK, B., MANICA, A., BALLOUX, F., OWEN, R. J., GRAHAM, D. Y., VAN DER MERWE, S., ACHTMAN, M. & SUERBAUM, S. 2008. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS pathogens*, 4, e1000180.
- SCOPES, R. K. 1973. Studies with a reconstituted muscle glycolytic system. The rate and extent of creatine phosphorylation by anaerobic glycolysis. *Biochemical Journal*, 134, 197-208.
- SCOPES, R. K. 1978. The Steady-State Kinetics of Yeast Phosphoglycerate Kinase. *European Journal of Biochemistry*, 85, 503-516.
- SECK, A., BURUCOA, C., DIA, D., MBENGUE, M., ONAMBELE, M., RAYMOND, J. & BREUREC, S. 2013. Primary antibiotic resistance and associated mechanisms in *Helicobacter pylori* isolates from Senegalese patients. *Annals of clinical microbiology and antimicrobials*, 12, 3.
- SHEN, Z., BATAK, F., MANNION, A., MILLER, M. A., BAKTHAVATCHALU, V., HO, C., MANNING, S., PASTER, B. J. & FOX, J. G. 2017. Novel urease-negative *Helicobacter sp.*'*H. enhydrae* sp. nov.'isolated from inflamed gastric tissue of southern sea otters. *Diseases of Aquatic Organisms*, 123, 1-11.
- SHIOTA, S., REDDY, R., ALSARRAJ, A., EL-SERAG, H. B. & GRAHAM, D. Y. 2015. Antibiotic resistance of *Helicobacter pylori* among male United States veterans. *Clinical Gastroenterology and Hepatology*, 13, 1616-1624.

- SLPPONEN, P., KEKKI, M., HAAPAKOSKI, J., IHAMÄKI, T. & SIURALA, M. 1985. Gastric cancer risk in chronic atrophic gastritis: Statistical calculations of cross-sectional data. *International journal of cancer*, 35, 173-177.
- SMITH, C. D., CHATTOPADHYAY, D. & PAL, B. 2011. Crystal structure of Plasmodium falciparum phosphoglycerate kinase: evidence for anion binding in the basic patch. *Biochem Biophys Res Commun*, 412, 203-6.
- SMYER, J. R. 1989. *Biochemical and genetic characterization of phosphoenolpyruvate synthase in salmonella typhimurium*. Texas Tech University.
- SOLEM, C., DEFOOR, E., JENSEN, P. R. & MARTINUSSEN, J. 2008. Plasmid pCS1966, a New Selection/Counterselection Tool for Lactic Acid Bacterium Strain Construction Based on the oroP Gene, Encoding an Orotate Transporter from *Lactococcus lactis*. *Applied and Environmental Microbiology*, 74, 4772-4775.
- SPEK, A. L. 2009. Structure validation in chemical crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 65, 148-155.
- ST. MAURICE, M., CREMADES, N., CROXEN, M. A., SISSON, G., SANCHO, J. & HOFFMAN, P. S. 2007. Flavodoxin:Quinone Reductase (FqrB): a Redox Partner of Pyruvate:Ferredoxin Oxidoreductase That Reversibly Couples Pyruvate Oxidation to NADPH Production in *Helicobacter pylori* and *Campylobacter jejuni*. *Journal of Bacteriology*, 189, 4764-4773.
- STACKEBRANDT, E., MURRAY, R. & TRÜPER, H. 1988. Proteobacteria classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives". *International Journal of Systematic and Evolutionary Microbiology*, 38, 321-325.
- STOGIOS, P. J., COX, G., SPANOGIANNOPOULOS, P., PILLON, M. C., WAGLECHNER, N., SKARINA, T., KOTEVA, K., GUARNÉ, A., SAVCHENKO, A. & WRIGHT, G. D. 2016. Rifampin phosphotransferase is an unusual antibiotic resistance kinase. *Nature Communications*, 7, 11343.
- SUGANO, K., TACK, J., KUIPERS, E. J., GRAHAM, D. Y., EL-OMAR, E. M., MIURA, S., HARUMA, K., ASAKA, M., UEMURA, N. & MALFERTHEINER, P. 2015. Kyoto global consensus report on *Helicobacter pylori* gastritis. *Gut*, 64, 1353-1367.
- SUTHERS, P. F., ZOMORRODI, A. & MARANAS, C. D. 2009. Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol Syst Biol*, 5, 301.
- SYCURO, L. K., PINCUS, Z., GUTIERREZ, K. D., BIBOY, J., STERN, C. A., VOLLMER, W. & SALAMA, N. R. 2010. Peptidoglycan crosslinking relaxation promotes *Helicobacter pylori*'s helical shape and stomach colonization. *Cell*, 141, 822-833.
- TAAL, B., BOOT, H., VAN HEERDE, P., DE JONG, D., HART, A. & BURGERS, J. 1996. Primary non-Hodgkin lymphoma of the stomach: endoscopic pattern and prognosis in low versus high grade malignancy in relation to the MALT concept. *Gut*, 39, 556-561.

- TADJROBEHKAR, O. & ABDOLLAHI, H. 2015. Survival and Chemotactic Behavior of *H. pylori* at Different Media pH. *Iranian Journal of Medical Sciences*, 29, 81-84.
- TANSWELL, P., WESTHEAD, E. W. & WILLIAMS, R. J. P. 1976. Nuclear-Magnetic-Resonance Study of the Active-Site Structure of Yeast Phosphoglycerate Kinase. *European Journal of Biochemistry*, 63, 249-262.
- THIELE, I., VO, T. D., PRICE, N. D. & PALSSON, B. O. 2005. Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol*, 187, 5818-30.
- THUNG, I., ARAMIN, H., VAVINSKAYA, V., GUPTA, S., PARK, J., CROWE, S. & VALASEK, M. 2016. The global emergence of *Helicobacter pylori* antibiotic resistance. *Alimentary pharmacology & therapeutics*, 43, 514-533.
- 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., FITZGERALD, 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. & VENTER, J. C. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 388, 539-47.
- TYTGAT, G. 1995. Endoscopic transmission of *Helicobacter pylori*. *Alimentary pharmacology & therapeutics*, 9, 105-110.
- UYTERLINDE, A. M., PENA, A., HAZENBERG, H., BLOEMENA, E., LINDEMAN, J., KLINKENBERG-KNOL, E. & MEUWISSEN, S. 1995. Increase of *Helicobacter pylori*-associated corpus gastritis during acid suppressive therapy: implications for long-term safety. *Am J Gastroenterol*, 90, 1401-1406.
- VAN ZANTEN, S. J. V., DIXON, M. F. & LEE, A. 1999. The gastric transitional zones: neglected links between gastroduodenal pathology and helicobacter ecology. *Gastroenterology*, 116, 1217-1229.
- VAS, M., VARGA, A. & GRACZER, E. 2010. Insight into the mechanism of domain movements and their role in enzyme function: example of 3-phosphoglycerate kinase. *Current Protein and Peptide Science*, 11, 118-147.
- VERLINDE, C. L., HANNAERT, V., BLONSKI, C., WILLSON, M., PERIE, J. J., FOTHERGILL-GILMORE, L. A., OPPERDOES, F. R., GELB, M. H., HOL, W. G. & MICHELS, P. A. 2001. Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resist Updat*, 4, 50-65.
- VERSALOVIC, J., SHORTRIDGE, D., KIBLER, K., GRIFFY, M. V., BEYER, J., FLAMM, R. K., TANAKA, S. K., GRAHAM, D. Y. & GO, M. F. 1996. Mutations in 23S rRNA are

- associated with clarithromycin resistance in *Helicobacter pylori*. *Antimicrobial Agents and Chemotherapy*, 40, 477-80.
- WANKEN, A. E., CONWAY, T. & EATON, K. A. 2003. The Entner-Doudoroff Pathway Has Little Effect on *Helicobacter pylori* Colonization of Mice. *Infection and Immunity*, 71, 2920-2923.
- WARREN, J. R. & MARSHALL, B. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *The lancet*, 321, 1273-1275.
- WATSON, H., WALKER, N., SHAW, P., BRYANT, T., WENDELL, P., FOTHERGILL, L., PERKINS, R., CONROY, S., DOBSON, M. & TUIITE, M. 1982. Sequence and structure of yeast phosphoglycerate kinase. *The EMBO journal*, 1, 1635.
- WEBB, M. R. & TRENTHAM, D. R. 1980. Analysis of chiral inorganic [16O, 17O, 18O]thiophosphate and the stereochemistry of the 3-phosphoglycerate kinase reaction. *Journal of Biological Chemistry*, 255, 1775-1778.
- WOESE, C. R. 1987. Bacterial evolution. *Microbiological reviews*, 51, 221.
- XIANG, Z. 2006. Advances in homology protein structure modeling. *Current Protein and Peptide Science*, 7, 217-227.
- YANG, J.-C., LIN, C.-J., WANG, H.-L., CHEN, J.-D., KAO, J. Y., SHUN, C.-T., LU, C.-W., LIN, B.-R., SHIEH, M.-J. & CHANG, M.-C. 2015. High-dose dual therapy is superior to standard first-line or rescue therapy for *Helicobacter pylori* infection. *Clinical Gastroenterology and Hepatology*, 13, 895-905. e5.
- ZHANG, Y.-X., ZHOU, L.-Y., SONG, Z.-Q., ZHANG, J.-Z., HE, L.-H. & DING, Y. 2015. Primary antibiotic resistance of *Helicobacter pylori* strains isolated from patients with dyspeptic symptoms in Beijing: a prospective serial study. *World journal of gastroenterology: WJG*, 21, 2786.
- ZHENG, H., FILIPPOVA, E. V., TKACZUK, K. L., DWORZYNSKI, P., CHRUSZCZ, M., POREBSKI, P. J., WAWRZAK, Z., ONOPRIYENKO, O., KUDRITSKA, M., GRIMSHAW, S., SAVCHENKO, A., ANDERSON, W. F. & MINOR, W. 2012. Crystal structures of putative phosphoglycerate kinases from *B. anthracis* and *C. jejuni*. *J Struct Funct Genomics*, 13, 15-26.
- ZOMER, A. W. M., ALLERT, S., CHEVALIER, N., CALLENS, M., OPPERDOES, F. R. & MICHELS, P. A. M. 1998. Purification and characterisation of the phosphoglycerate kinase isoenzymes of *Trypanosoma brucei* expressed in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1386, 179-188.
- ZUCCA, E., DREYLING, M. & GROUP, E. G. W. 2009. Gastric marginal zone lymphoma of MALT type: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Annals of Oncology*, 20, iv113-iv114.

ZULLO, A., DE FRANCESCO, V., HASSAN, C., MORINI, S. & VAIRA, D. 2007. The sequential therapy regimen for *Helicobacter pylori* eradication: a pooled-data analysis. *Gut*, 56, 1353-1357.