In silico analyzes of porins involved in niche adaptation:

Exploring the role of *Helicobacter pylori*
outer membrane phospholipase A in acid tolerance

Dissertation for the degree of Philosophiae Doctor (Ph.D.)

Hilde Synnøve Vollan

Division for Infection Control and Environmental Health,
Norwegian Institute of Public Health
and
Department of Clinical Molecular Biology (EpiGen), Division of Medicine,
Akershus University Hospital and University of Oslo

2017

Acknowledgements

I would first like to thank my principal supervisor prof. Geir Bukholm for providing motivation and encouragement throughout the project. Scientific hypotheses have been generated with great enthusiasm, based on his experience in the field of microbiology, medicine and global health. Throughout my battle with health issues, the tasks and workload have always been adjusted flexibly and insightful to fit my situation.

I have learned a great deal of bioinformatics through the courses at Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Nijmegen Medical Centre (Radboudumc), The Netherlands. I would especially thank my co-supervisor prof. Gert Vriend for taking the time to patiently teach me protein structure analyzes. The thesis is based on laboratory work by my co-supervisor Tone Tannæs who also has helped and supported me throughout these years. I would also like to thank my co-supervisor prof. Dominique A. Caugant for her insight and guidance with my thesis work. Finally, I would like to thank all my co-workers, especially Qin Ying Esbensen for the scientific help that improved this work.

The love of my life, Erik Gjerdrum, has always been by my side ensuring we got through every hurdle. I am forever thankful for my supportive family who has debated a field of science unknown to them at all hours of the day. I do believe they now find bioinformatics a word with meaning. Living with chronic pain is tough, but current research in the field of cluster headaches and migraines has given me hope, motivation and the pain free days necessary to complete my dissertation. My gratitude also goes to all the doctors and health practitioners who have helped ease painful days and enabled me to work on my thesis.

Sammendrag (Norwegian abstract)

Helicobacter pylori er en Gram-negativ bakterie som koloniserer ventrikkelslimhinnen hos mennesker. Dette er et utfordrende habitat hvor bakterien må både unngå vårt immunforsvar og tilpasse seg et surt miljø. Bakterien benytter et ustabilt genom med bl.a. fasevariable gener for å tilpasse seg endringer i miljøet. Et slikt fasevariabelt gen er *pldA* genet som koder for et aktivt eller trunkert fosfolipase A (OMPLA) i denne bakterien. OMPLA er et enzym man finner i yttermembranen hos Gram-negative bakterier, hvor den konverterer fosfolipider til lysofosfolipider. OMPLA blir vanligvis kun aktivert ved ekstrem stress ved nøytral pH. *H. pylori* med en aktiv OMPLA uttrykker i større grad virulensfaktorer sammenlignet med de som har et inaktivt protein, og er dermed assosiert med en økt risiko for sykdom. *H. pylori* må ha en intakt OMPLA for å kunne overleve surt miljø, uten at enzymet aktiveres.

Målet med denne oppgaven har vært å studere *pldA* genet med et fokus på nisje-adapsjon. Bioinformatiske verktøy ble benyttet til å studere hvordan aktiv OMPLA hjelper bakterien å overleve i surt miljø. Yttermembranproteiner hos Gram-negative bakterier har blitt analysert med fokus på nisjetilpasninger. Selv om bakterielle yttermembranporteiner innehar en rekke ulike funksjoner, er selve proteinstrukturen godt bevart. Den svært stabile β-barrel-konformasjonen er felles og nødvendig for at bakteriene skal overleve ekstreme miljøer. Kunnskapene vi fikk fra disse analysene ble benyttet til å utvikle OMPLA-modellen. Modelleringer av OMPLA-strukturen indikerer at OMPLA, i tillegg til å være et enzym, også kan fungere som en pore.

Studiene tyder på at OMPLA kan være involvert i urea metabolismen, som er en viktig overlevelses mekanisme for *H. pylori* i surt miljø. Cytoplasmisk urease katalyserer reaksjonen fra urea til karbondioksid og ammoniakk, men det finnes i dag ingen kjent yttermembran pore for hverken urea opptak eller ammonium utskillelse. Vår hypotese er at urea diffunderer inn gjennom OMPLA og ammoniakk skilles ut via OMPLA. OMPLA ser ut til å reguleres sammen med bl.a. to innermembran-transportere i et felles operon. Vi tror disse transportere også er med i urea metabolismen, og det er sannsynlig at de skiller ammonium (NH_3/NH_4^+) ut i periplasma. OMPLA og disse to innermembran proteinene ser ut til å ligge i samme operon hos alle gastriske *Helicobacter*. Analysene tyder på at *H. pylori* OMPLA er et nisje-tilpasset protein som er med på å nøytralisere det sure miljøet i magesekken.

Abbreviations

AbPirA	*Acinetobacter baumannii* PirA
Ail	Attachment invasion locus
AlpA	Adherence-associated lipoprotein A
AlpB	Adherence-associated lipoprotein B
AmCI	Ammonium Channels I
AmCII	Ammonium Channels II
AtpA	Adenosin-5'-trifosfat (ATP) synthase subunit α
ATPase	Adenosine TriphosPhatase
ATR	Acidic Tolerance Response
BabA	Blood group antigen-binding adhesin A
BabB	Blood group antigen-binding adhesin B
BLAST	Basic Local Alignment Search Tool
bp	base pairs
Cag	Cytotoxin-associated gene (*e.g.* CagA, CagPAI)
CAI	Codon Adaptation Index
CATH	Class, Architecture, Topology, Homology
CFB	Cytophaga-Flavibacterium-Bacteroides
Cir	Colicin I receptor
CMBI	Centre for Molecular and Biomolecular Informatics, Radboudumc
DNA	DeoxyriboNucleic Acid
dupA	duodenal ulcer promoter gene A
dPNAG	N-acetylated poly-β-1,6-N-acetyl-d-glucosamine
E-cadherin	Epithelial cadherin
eCAI	estimated Codon Adaptation Index
EcFadl	*E. coli* FadL
EcLptD	*E. coli* LptD
EcMaltoporin	*E. coli* Maltoporin
EcOmp	*E. coli* Outer membrane protein
EcOMPLA	*E. coli* Outer Membrane Phospholipase A
EHEC	Enterohemorrhagic *E. coli*
E_i	Shannon Entropy (used in EVA)
EM	Electron Microscopy

|	
EV	Entropy-Variability (*e.g.* EV plot)
EVA	Entropy-Variability Analyzes
FUR	Ferric Uptake Regulator
GDP	General Diffusion Porin
GGT	γ-GlutamylTransferase
GI	GastroIntestinal
GTR	Generalized Time-Reversible algorithm
HdBamA	*Haemophilus ducreyi* β-barrel assembly machinery A
HGT	Horizontal Gene Transfer
HK genes	HouseKeeping genes
HopQ	*Helicobacter* outer membrane protein Q
HtrA	High temperature requirement A
IM	Inner Membrane
IL-8	Interleukin 8
K80	Kimura's two parameter model from 1980
KEGG	Kyoto Encyclopedia of Genes and Genomes
KpLptD	*Klebsiella pneumoniae* LptD
KpOmp	*Klebsiella pneumoniae* Outer membrane protein
KpOmpA	*Klebsiella pneumoniae* OmpA
LG	Le and Gascuel substation matrix (2008)
LPS	LipoPolySaccharides
LRT	Likelihood Ratio Test
M2-M8	Phylogenetic analyzes of maximum likelihood Models 2 to 8
MALT	Mucosa-Associated Lymphoid Tissue
MD	Molecular Dynamics
MDR	Multi-Drug Resistance
ML	Maximum Likelihood
MLST	Multi Locus Sequence Typing
MSA	Multiple Sequence Analyzes
NanC	N-acetylneuraminic acid-inducible outer-membrane Channel
NCBI	National Center for Biotechnology Information
NgBamA	*Neisseria gonorrhoeae* β-barrel assembly machinery A
NMR	Nuclear Magnetic Resonance

NPH	Non-Pylori *Helicobacter*
nt	nucleotides
Occ	Outer membrane carboxylate channel (*e.g.* occK5 and occD)
OipA	Outer inflammatory protein A
OM	Outer Membrane
OMP	Outer Membrane Protein (*e.g.* OmpA, OmpC, OmpF, OmpG)
OMPdb	Outer Membrane Protein database
OMPLA	Outer Membrane Phospholipase A
OMV	Outer Membrane Vesicle
PaFadL	*Pseudomonas aeruginosa* FadL
PagL	PhoP/PhoQ-activated gene product L
PAI	PAthogenicity Island (e.g. CagPAI)
PaLptD	*Pseudomonas aeruginosa* LptD
PAML	Phylogenetic Analyzes of Maximum Likelihood
PaPirA	*Pseudomonas aeruginosa* PirA
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Pfam	Protein families' database
PIR	Protein Information Resource
PMF	Proton Motive Force
RbGDP	*Rhodobacter blastica* General Diffusion Porin
RcGDP	*Rhodobacter capsulatus* General Diffusion Porin
RFLP	Restriction Fragment Length Polymorphism
RNA	RiboNucleic Acid (*e.g.* messenger mRNA, transfer tRNA)
SabA	Sialic acid-binding adhesin A
Salmonella Typhimurium	*Salmonella enterica* serovar Typhimurium
SCOPe	Structural Classification Of Proteins extended
SeLptD	*Salmonella entrerica* LptD
SfLptD	*Shigella flexneri* LptD
Sp. or Spp.	Species in single or plural form
StMaltoporin	*Salmonella enterica* serovar Typhimurium maltoporin
StOMPLA	*Salmonella enterica* serovar Typhimurium OMPLA
T4SS	Type IV secretion system

TbpA	Transferrin binding protein A
TCDB	Transporter Classification DataBase
tfs3a	type IV secretion 3A
Ts/Tv	Transition /Transversion
VacA	Vacuolating cytotoxin autotransporter
V_i	Variability (used in EVA)
WHO	World Health Organization
XAR	eXtreme Acid Resistance
YpLptD	*Yersinia pestis* LptD

Preface

Evolution can be defined as the changes in a population over time, and evolutionary biology is a broad field of science that include studying how host-microbe interactions occur and affect population diversity [1, 2]. Bacteria and viruses are constantly evolving and adapting to their hosts, and tools that analyze these evolutionary processes can be powerful in combating microbial diseases.

Charles Darwin (1809-1882), depicted in Preface Figure 1, was a British naturalist and geologist. His work on evolution has affected many scientific fields [3, 4]. Darwin presented a controversial theory in his book entitled "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life" in the mid-19th century that has been debated ever since [5]. He is one of the founding fathers of modern biology and his theory explains the complexity of life on earth. His evolutionary theory used natural selection to explain both adaptation and speciation (where new species arises from already existing species). The evolutionary biologist and geneticist Theodosius Dobzhansky (1900–1975) continued Darwin's work and wrote an essay entitled "Nothing in biology makes sense except in the light of evolution", where he describes how evolution gives meaning to our world [1, 6].
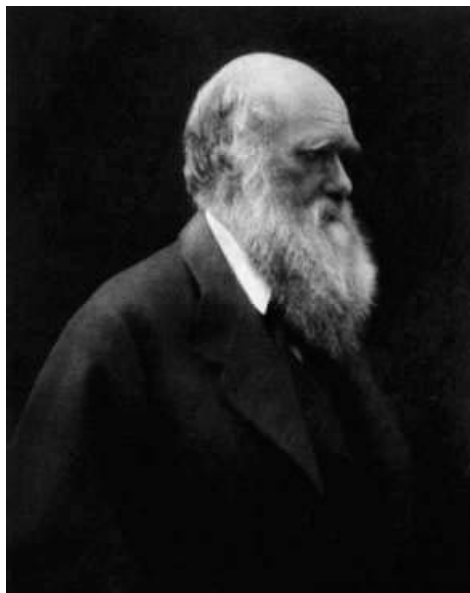
Preface Figure 1: Charles Darwin. *This is a photography of Charles Darwin (1809-1882) taken in 1868 by Julia Margaret Cameron.*

In this thesis, evolutionary algorithms and bioinformatic tools are used to study the evolution of bacterial proteins, see Preface Figure 2. These tools are used to shed light on how the ulcer-causing bacterium *Helicobacter pylori* can survive the harsh conditions found in the human stomach. *H. pylori* is a widely spread persistent bacterium that has co-evolved with humans for thousands of years [7]. This bacterium has evolved many mechanisms to evade human immune responses, and may colonize people for decades without doing any harm. However, *H. pylori* may suddenly attack the gastric epithelial cells, causing inflammation and eventually ulcer. In order to attack, *H. pylori* must survive the acidic bursts in the gut and swim through the gastric mucosa which is a protective gel-layer not suited for most microbial life.

Preface Figure 2: Bacterial protein. *This is as an example of a bacterial protein structure, the top-view of a trimeric osmoporin (PDB ID: 2J1N), isolated from the outer membrane of Escherichia coli.*

Evolutionary analyzes have been used to study a protein function and identify regions of high importance to a protein family. In this thesis, thousands of protein sequences were collected in order to study both sequence conservation and variability. Variable residues may indicate regions that is continuously changing and adapting to an environment, *e.g.* host evasion mechanism. These might be frequently changing surface-exposed residues that avoid being recognized by the host immune system.

Table of Contents

List of Figures

1. Introduction

1.1 *Helicobacter pylori*

1.1.1 The *Helicobacter* genus

The *Helicobacter* genus consists of a large group of microaerophilic, highly motile, helical rod-shaped Gram-negative bacteria. There are currently 52 *Helicobacter* species in the National Center for Biotechnology Information (NCBI) taxonomy database [8]. They colonize a wide variety of organs and host species [9-14]. The *Helicobacter* genus can be divided into two subgroups based on where the bacteria prefer to colonize. These two groups, gastric (stomach) and enterohepatic (liver or intestines), can be separated by morphology and through phylogenetic analyzes [12, 15-17].

1.1.2 Prevalence

The Australian scientists Barry Marshall and Robin Warren were the first to culture gastric *Helicobacter pylori* from human ventricle in 1982. They suggested that peptic ulcer was an infectious disease rather than caused by stress and/or other lifestyle factors. Peptic ulcer is an open sore, usually found in the stomach lining or duodenum. Marshall and Warren received the Nobel Prize in Physiology or Medicine in 2005 for their discovery of the bacterium *H. pylori* and its role in gastritis and peptic ulcer disease [18]. Today, more than half of the world's population is colonized with *H. pylori*. According to Hooi *et al.*, the African continent has the highest *H. pylori* prevalence, estimated ~70% on average. Country estimates ranged from ~20% in Switzerland to ~90% in Nigeria, compared with an *H. pylori* infection prevalence estimate of ~30% in Norway [19].

Although the life-cycle of *H. pylori* might not be fully traced, the bacterium is likely transmitted through an oral-oral or fecal-oral route [20, 21]. *H. pylori* has been detected in different environments, including oral cavity, feces, and contaminated water [21-23]. It may travel as passengers inside yeast or amoebas by forming vacuoles [24, 25]. Among possible sources of transmission, Kuipers *et al.* and Anand *et al.* reviewed the role of dental plaque, saliva, and periodontal disease in *H. pylori* infections [26, 27].

1.1.3 Colonization

The human gastric mucosa is a protective gel-like layer located in the stomach lining. Bacterial motility is difficult in this layer, and acidity makes the gastric mucosa a relatively sterile environment. It was long thought that the human stomach was sterile, but emerging research indicates that core microbiota can exist in this acidic environment [28]. Phyla detected in the human gastric mucosa includes *Actinobacteria*, *Proteobacteria*, *Firmicutes*, *Bacteroidetes* and *Fusobacteria* [29], but the ε-proteobacterium *H. pylori* is the most prevalent species [30, 31].

H. pylori has overcome the obstacles found in the human gastric mucosa mainly due to rotating flagella and its helical shape, aiding chemotaxis towards better places [32]. When the bacterium is closely associated with epithelial cells, it is exposed to a strong respiratory burst and can tolerate oxidative stress [33]. *H. pylori* usually prefers the gastric epithelial surface because it has a slightly alkaline or neutral pH compared to the acidic gastric lumen, as shown in Figure 1. The stomach lumen pH can fall below 2.5 [31].

The acidic resistance mechanisms found in *H. pylori* vary from the general survival bacterial mechanisms (Acidic Tolerance Response, ATR; and eXtreme Acid Resistance, XAR), reviewed by Lund *et al.* [34]. The two-component ArsRS proteins in *H. pylori* sense pH changes and can trigger the gene expression of more than 100 acid-response genes [35]. *H. pylori* senses host urea through chemotaxis [36], and urea is translocated into the cytoplasm to buffer the acidic environment through the urea pathway. This process has been described in several review articles, including those by McNulty *et al.*, Krulwich *et al.*, and Sachs *et al.* [37-39]. The *H. pylori* inner membrane urea channel, UreI, is involved in the transport of urea from periplasm to cytoplasm where it is closely associated with urease. This cytoplasmic urease catalyzes the urea-reaction by producing carbon dioxide and ammonia, which will maintain the cytoplasmic pH neutral. Carbon dioxide diffuses to the periplasm where α-carbonic anhydrase converts CO_2 to bicarbonate and participates in the periplasmic pH buffering at around ~6.1. Ammonia will mostly be protonated and will (most likely mainly in its protonated form) diffuse outside. As *H. pylori* acid tolerance mechanism is unraveling, multifunctional proteins crucial for this process have been discovered, including urease and UreI [40-42]. Urease may be multifunctional and is required for not only neutralizing cytoplasmic pH, but also implicated in colonization. However, urease's role in host colonization has been debated since *H. pylori* strains lacking urease may colonize the stomach of gerbils [43].
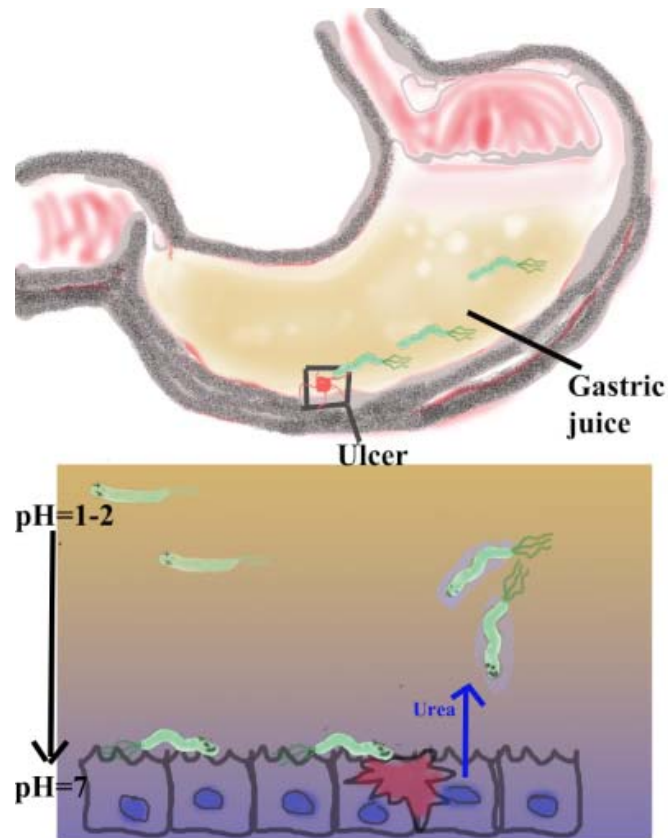
Figure 1: H. pylori colonization. *The gastric mucosa is a gel-like protective barrier between the stomach and epithelial cells. H. pylori surpass this harsh environment with regard to both the viscosity and acidity. Ammonium is a product from the urease reaction that buffers the acidic environment, resulting in a liquid-state mucosa that enables swimming (the light blue region surrounding the bacteria). It targets the neutral site found at the epithelial cells through chemotaxis (the blue arrow). H. pylori is capable of chemotaxis towards the host's urea excretion site. This is just an example using an ulcer wound (red) as an illustration.*

1.1.4 Virulence factors and pathogenicity

H. pylori colonization in the gastric mucosa requires factors associated with or secreted from its outer membrane (OM) [44]. Pathogenesis depends upon several factors, and in this thesis, virulence factors are defined as molecules associated with disease [31, 41, 45, 46],

Adhesins are crucial for the initial colonization and infection since they are involved in binding to the gastric epithelial cell surface [41]. Among the OM protein (OMP) adhesins implicated in colonization and inflammation are sialic acid-binding adhesin (SabA), blood-group-antigen-binding adhesin (BabA), adherence-associated lipoprotein A and B (AlpA and AlpB), outer inflammatory protein A (OipA), and *Helicobacter* OM protein Q (HopQ) [47]. The protein expression profiles of OipA, BabA, and SabA are higher in gastric cancer patients, and has been suggested to be us ed as biomarkers for gastric cancer [48].

The cytotoxin-associated gene A product (CagA), vacuolating cytotoxin autotransporter (VacA), duodenal ulcer promoter gene A (dupA), and OipA are among the virulence factors of most importance to clinical outcome [45, 47, 49-51].

- CagA is a toxin injected into host cells via the Type IV secretion system (T4SS) to induce an inflammatory response [41]. OMP adhesins previously mentioned enhance CagA translocation [47]. The Cag pathogenicity island *(cag*PAI) operon encodes both the virulence protein CagA and the T4SS secretion system that injects CagA into the host cell. This island is a 40-kB segment that consists of nearly 30 genes subdivided into ~11 operons [52, 53].

- VacA is a vacuole-inducing toxin [41]. Internalized VacA leads to swelling, forming vacuole-like organelles in the host cell. This multifunctional toxin may initiate apoptosis, activate inflammatory response in its host, inhibit T-cells or obstruct cell proliferation [41, 54]. VacA is aided by several proteins, and recent research implicates γ-glutamyltransferase (GGT) function in enhanced VacA vacuolization [55]. GGT is a conserved virulence factor that improve colonization and is associated with higher risk for developing peptic ulcer disease [56, 57].

- The *dupA* gene has been implicated in higher risk of duodenal ulcer. This gene may induce interleukin 8 (IL-8) and secrete urease, but inhibit gastric cancer [58]. The biological role of dupA has been debated, but it is believed that dupA is part of a novel T4SS transporter complex named Type IV secretion 3A (tfs3a). It is homologue to virB4, a T4SS adenosine triphosphatase (ATPase) [45], integrated in the genome with other virulence genes, and experiments show that it is a membrane-bound ATPase [58].

- OipA is a phase variable adhesin, and a virulence factor that induces IL-8. OipA initiates an OipA-specific pathway that is likely to produce a more sever inflammation in the host, and may also interact with CagA [59].

Another prominent virulence factor is the high temperature requirement A (HtrA). HtrA is a conserved protease that cleaves epithelial cadherin (E-cadherin), resulting in bacterial influx since the cell-to-cell junctions are opened. Furthermore, E-cadherin is a tumor suppressor protein involved in preventing gastric cancer which is interfered by HtrA [60, 61].

1.1.5 Disease and treatment

It is believed that *H. pylori* is orally transmitted within families, and colonization usually occurs in early childhood, causing disease in about 10-15% of the cases [31, 62, 63]. *H. pylori* infection usually causes asymptomatic chronic gastritis, which is a risk factor for developing gastric or duodenal ulcers, mucosa-associated lymphoid tissue (MALT) or gastric adenocarcinoma [26]. In some cases, the disease will be diagnosed as gastric cancer [64]. *H. pylori* was classified as a group 1 carcinogen in 1994 by The International Agency for Research on Cancer (a World Health Organization (WHO) agency) [65]. Plummer *et al.* estimated in 2015 that 6.2% of all cancers worldwide are caused by *H. pylori* [66].

Several studies have concluded that *H. pylori* vaccination would be cost-effective and the preferred option for infection management [31, 67-69]. Although *H. pylori* vaccine development started in the early 1990s, no effective vaccines exist. Several successful immunization experiments in rodents exist using *H. pylori* specific antigens like urease, CagA, and VacA. However, these *H. pylori*-specific vaccines are non-effective or giving a short-lived protection in human trials [31]. Salama *et al.* concluded that a successful human vaccine candidate would have to override the host immune response [31]. A phase III clinical trial with an oral recombinant *H. pylori* urease vaccine was successfully completed, and holds promise for a prophylactic treatment [70, 71]. In this trial, 70% protection was observed in children and the highest efficacy measured in the first year [71, 72].

H. pylori eradication is likely the best treatment until an effective vaccine is available [73]. Several countries recommend a triple therapy (proton-pump inhibitor (PPI), clarithromycin, and amoxicillin/metronidazole) for 14 days as a first-line treatment for *H. pylori* eradication, as illustrated in Figure 2 [72]. This triple therapy inhibits protein synthesis (clarithromycin), and either inhibits cell growth (amoxicillin) or inhibit nucleic acid synthesis (metronidazole). However, depending on previous antibiotic history other first-line therapies are prescribed to the patient, including the bismuth quadruple therapy (PPI, bismuth, tetracycline, and nitroimidazole), concomitant therapy (combinations of antibiotics used together with PPI) or sequential therapy (a 5+5 day treatment with varying antibiotics, *e.g.* amoxicillin the first five days, followed by clarithromycin for five days) [72]. Antibiotic resistance is an increasing problem worldwide. *H. pylori* is now on the WHO priority pathogens list for research and development of new antibiotics, especially due to the rise of clarithromycin-resistant strains [74].
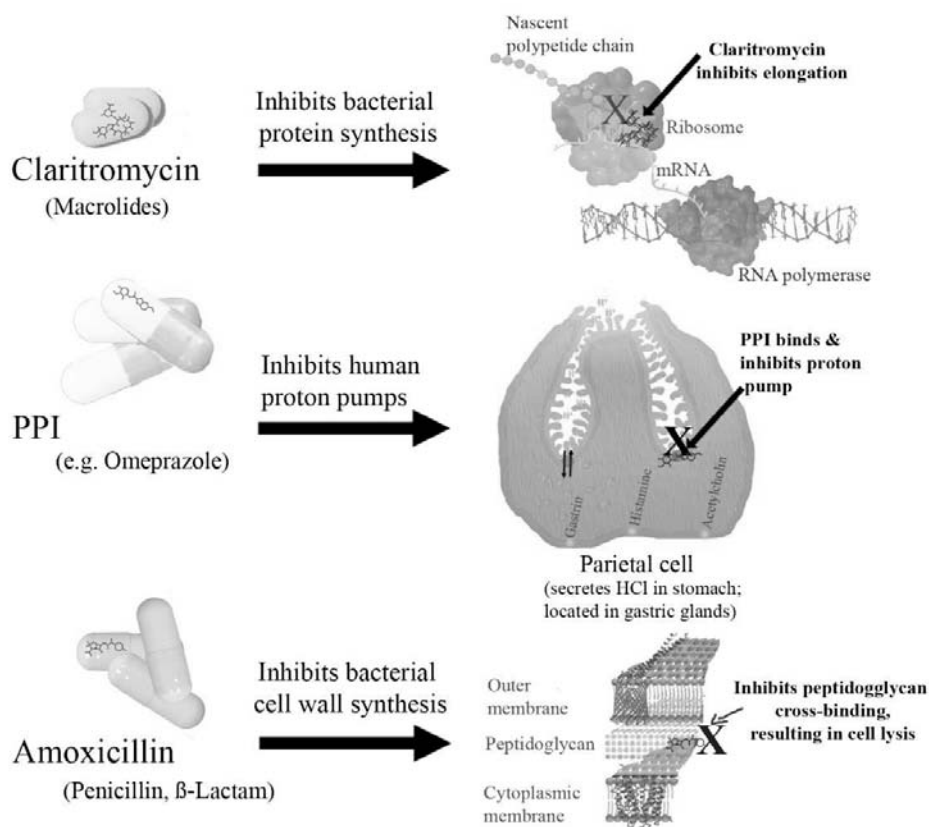
Figure 2: H. pylori triple therapy. This figure illustrates the three therapeutic components that may constitute a triple therapy (left-side) and how they work (right-side). Protein Data Bank (PDB) structures (drug IDs: DB01060, DB00338 and DB01211) were retrieved from DrugBank [75].

1.1.6 Genomic variation

Bacteria can thrive in a vast variety of environmental niches. They live in a continuously changing environment and rapidly adapt to temperature changes, nutrient limitation and the exposure to toxic substances. They may endure harsh conditions, *e.g.* acidic gastric mucosa. Genetic mechanisms that yield bacterial diversity enable rapid adaptation. These genetic mechanisms include rearrangements on larger and smaller genomic scale, *e.g.* recombination, horizontal gene transformation (HGT), inversion, mutation, insertion and deletion [76]. HGT is a common mechanism to obtain resistance and other genes needed for survival in a habitat [77]. *H. pylori* genes and operons are often involved in HGT, including the *cag*PAI operon [78]. Wiedenbeck discussed how HGT can contribute to adaptation in "Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches" [77].

H. pylori is known as a quasispecies due to high genetic variability [79]. DNA variability is caused by frequent mutations and natural transformation. Recombination contributes to *H. pylori* adaption and survival [80-82]. *H. pylori* quickly adapts to changes in the environment, and Salama *et al.* reviewed how this bacterium has evolved different mechanisms to escape the host immune system [31]. The wide range of places where *H. pylori* has been detected, reflects an adaptable bacterium that tolerates many obstacles, including respiratory bursts, acidic gastric environment, and other host defense mechanisms. *H. pylori* has, for example, evolved specialized DNA-repair systems and gene expression pathways that allow survival in harsh conditions, *e.g.* the acidic gastric mucosa [33, 83].

Despite the high sequence variation observed in *H. pylori*, 1237 common core genes were found among five *H. pylori* genomes. The encoding amino acid identities range between 65-100%. Among these core genes are housekeeping (HK) genes, essential for *H. pylori* survival, and the genetic variability in these genes remains very low [84]. This conservation is reflected in phylogenetic analyzes, where sequence analysis of HK genes has been used to trace human migration, indicating co-evolution between *H. pylori* and its host. Linz *et al.* traced *H. pylori* infection in humans to before their migration from Africa. This supports clinical data showing that ulcer and gastric cancer have occurred for thousands of years [85]. There are a few other genes found in the core genomes that show high conservation, *e.g. htrA* that encodes the HtrA protease is likely important for virulence, but is not required for cell maintenance or replication [61].

Phase variation is a mechanism used by the bacteria to control the transcription of a gene. Various genetic mechanism could be the source of phase variation, including HGT [76]. Through homopolymeric tracts, the bacteria may switch a gene ON or OFF by inserting or deleting an extra nucleotide. This will often lead to a missense mutation resulting in a premature STOP codon, terminating the transcription process too early [86, 87]. An allele resulting in a truncated protein might be dissolved by the bacterial repair system or it will have a dysfunctional role in the bacteria [86]. *H. pylori* contains phase variable OMPs that allow quickly adapting to an environment that is continuously changing: SabA, a phase variable adhesive OMP [59]; phase variable BabA and BabB OMPs that help escape immunoglobulin attacks [81]; and the phase variable OMPLA enzyme [88].

1.2 Bacterial OMPs

1.2.1 The OM barrier

Bacterial species can be divided into two main groups, Gram-positive and Gram-negative based on their cell wall composition [89]. Gram-negative bacteria have a cell envelope that consists of an OM and inner membrane (IM), separated by the periplasm, while Gram-positive bacteria lack an OM [90]. Despite vast differences in bacterial envelope structures, similar transport mechanisms for ions and nutrients have been identified. Many membrane transport mechanisms still remain unclear since only a small fraction of membrane proteins have been solved, and many functions and possible moonlighting functions still need to be unraveled [91-93].

Nikaido described the OM as a barrier that simultaneously expels toxins and allow nutrient passage [94]. The OM contains an inner (mainly phospholipids) and an outer leaflet (lipopolysaccharides; LPS). LPS belong to a family of glycolipids and the three components that constitute the amphipathic LPS layer are lipid A, core oligosaccharide, and O-antigen. This LPS layer is critical for survival of all Gram-negative bacteria [92]. There are generally two distinct protein classes in the OM; the lipoproteins (lipid molecules) embedded in the inner leaflet and transmembrane membrane β-barrel structures (OMPs). OMPs account for roughly 50% of the OM mass. Lin *et al.* discussed how proteins located in the OM, which include siderophores and efflux pumps, are involved in bacterial adaptation [95].

1.2.2 OMP structures

OMPs are found in mitochondria, chloroplasts and Gram-negative bacteria. These proteins in Gram-negative bacteria are embedded in the OM with a β-barrel motif. Other motifs are found in the OM of both chloroplasts and mitochondria. Gram-positive bacteria and *Archaea*, on the other hand, do not seem to have β-barrel proteins embedded in their membrane. Most of the β-barrel proteins with known 3D structures are found in the Proteobacteria phylum, but Yen *et al.* showed that β-barrel transporter proteins are also found in non-Proteobacteria species (*e.g. Deinococcus*, *Cyanobacteria* and *Mycobacteria*) [96]. The Gram-negative bacteria *Pseudomonas aeruginosa,* have evolved membranes lacking general diffusions porins. The channels present in the OM are specific, and allows a tighter regulation of molecules passing through the membrane, which includes blocking the entrance of toxic substances (*e.g.* antibiotics) [97, 98].

OMPs are predominantly composed of β-strands that are circularly connected to form a barrel motif as depicted in Figure 3. This motif is common despite OMPs vast variety of functions. The next sections include information derived from β-barrel review articles that discuss OMP subfamilies and functions from Bishop *et al.*, Fairman *et al.,* Galdiero *et al.*, Koebnik *et al.*, Lin *et al.*, Nikaido, Pages *et al.*, Schulz, Van den Berg, and Wimley [94, 95, 99-105].
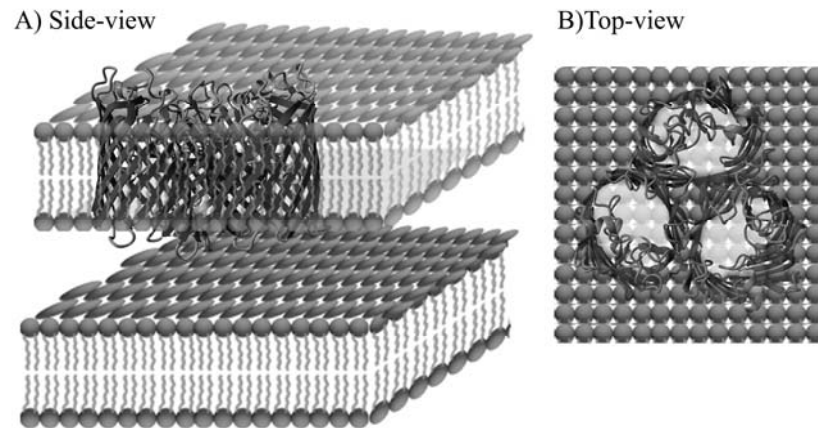
Figure 3: OMP structure. *OMPs are composed of a transmembrane motif that consists of ß-strands forming a ß-barrel. These barrels are connected by short turns in the periplasm and longer extracellular loops. The number of ß-strands, the length of the ß-strands, and the loop-lengths vary widely among OMPs. Here, the trimeric general diffusion porin (GDP) from Escherichia coli (PDB ID: 2J1N, [106]) is shown as an example. A) The side-view of the protein with the barrel embedded in the OM membrane. B) The top-view where the same structure is viewed in isolation from the extracellular side, and the pores are visible. Strands are colored red, while loop and turns are colored turquoise. Co-factors and water from the pdb file was removed for clarity.*

All bacterial transmembrane ß-barrels include an even number of anti-parallel ß-strands, and both the N- and C-terminus are usually found at the periplasmic side. These proteins are stabilized through extensive hydrogen bonding network and/or oligomerizations. The strands are usually connected with short turns at the periplasmic side and long loops at the extracellular side. OMPs are usually amphipathic proteins having a polar core and hydrophobic residues that interact with the lipid membrane. Figure 4 illustrates the distribution of hydrophobic and hydrophilic residues. The alternating hydrophobic/ hydrophilic side chains create a polar core with hydrophobic side chains facing the lipid membrane [107-109].
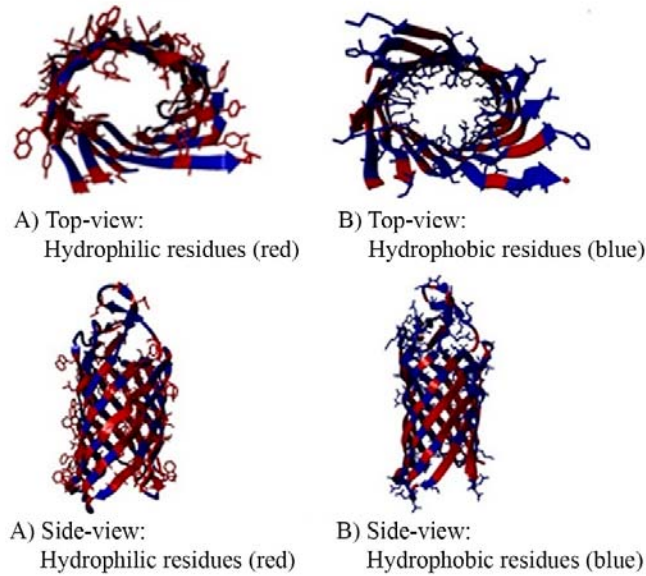
Figure 4: OMP hydrophobicity. *The distribution of hydrophobic and hydrophilic residues represented by the monomeric 12 ß-stranded NanC (PDB ID: 2WJR, [110]). Hydrophobic residues represented by red and hydrophilic residues are colored blue. The extracellular loops have been removed for clarity to illustrate the polar core in panel A and B. Only the hydrophilic side chains are visualized (red) in panel A and C. Panel B and D visualized only the hydrophobic residues (blue). This figure shows most blue hydrophilic residues located in the barrel core (for the transport of acidic sugars), while most the hydrophobic residues are facing the lipid membrane.*

The number of structures solved and deposited in PDB is continuously increasing, although not as much as in the late 90s to the early 2000s. These solved structures give us a great deal of information, but many questions remain unsolved. Electron Microscopy (EM), Nuclear magnetic resonance (NMR) and X-ray crystallography are methods used to solve protein structures. These techniques yield high-resolution 3D coordinates. Some protein families, like membrane proteins, are more difficult to solve, because they are difficult to isolate as they are embedded in a hydrophobic lipid-layer. Homology modeling techniques exist to help solve questions on those proteins that have not yet been answered (9).

A homology model is created from a protein sequence of an unknown 3D structure, based on a known 3D structure template (homologous to the model sequence). The higher the sequence similarity is between the template and the model, the better quality the model has, see Figure 5. The model is based on the alignment between the template sequence (from a solved protein structure) and a model sequence (with known sequence, but unknown structure). The quality of a model is therefore dependent on the sequence identity of the model and template, but also on the quality of the multiple sequence alignment and the quality of the template. A good alignment

can only be created when as many sequences as possible are used. It is important that the final alignment is not too diverse, yet not too similar. Usually, this is done by filtering out the high and low sequence identity.
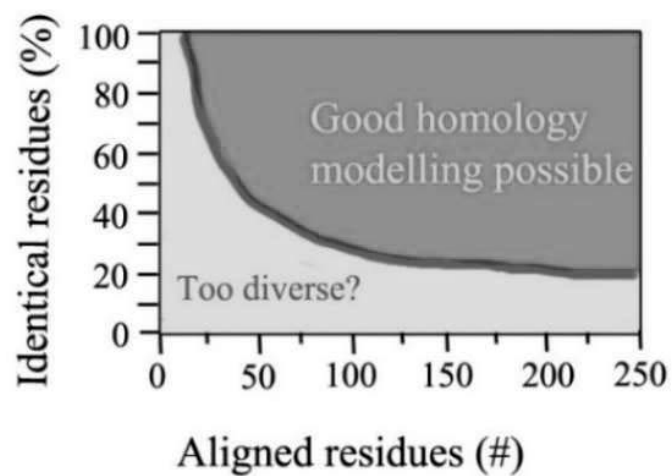
Figure 5: Homology modeling graph. *This graph shows the optimal sequence between model and template.*

1.2.3 OMP classes

OMPs can be classified into six main subfamilies: efflux pumps, enzymes, transporters, ushers, virulence factors, and porins. Structures representing each of the six bacterial OMP subfamilies are illustrated in Figure 6.
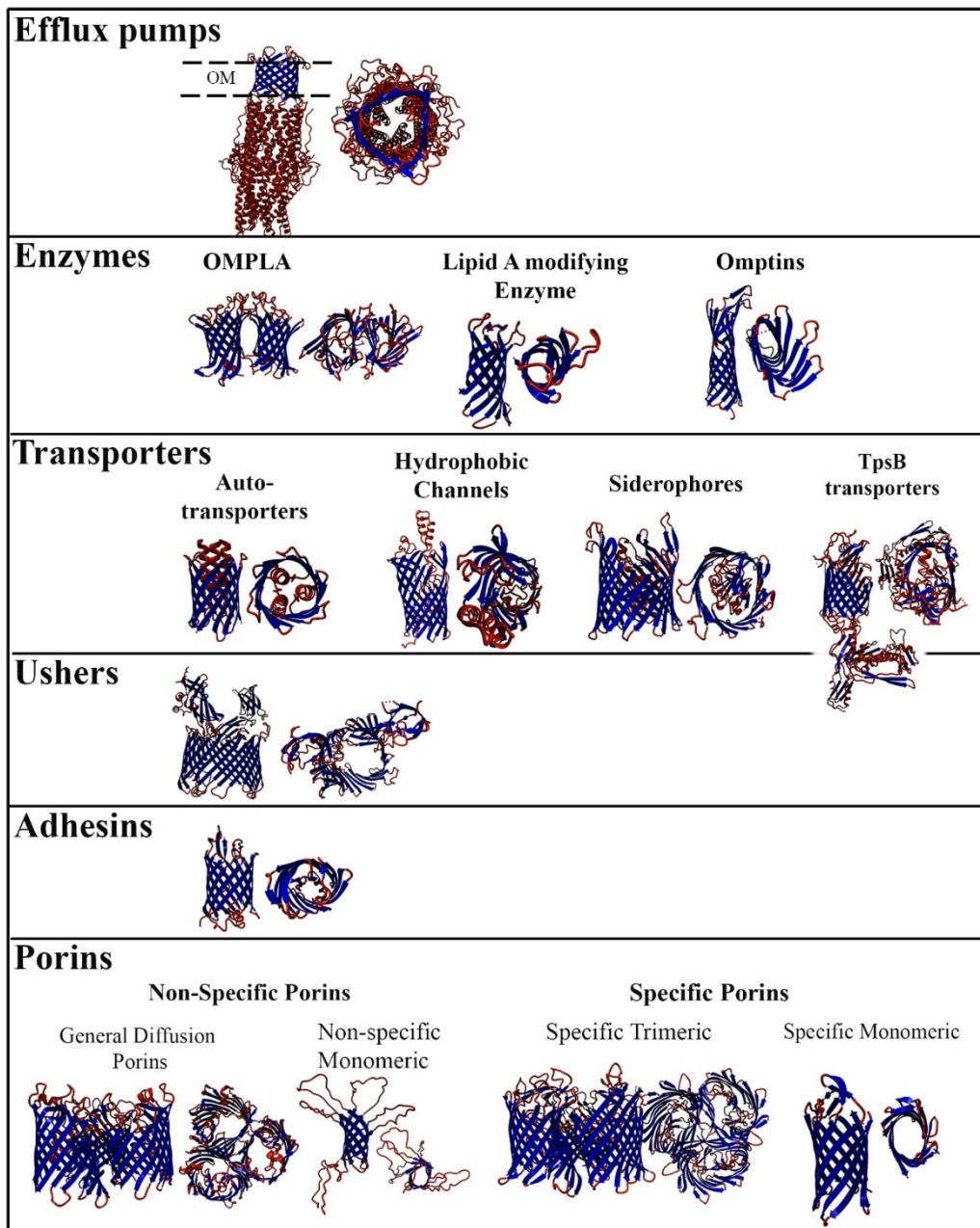
Figure 6: OMP classification. *OMP classification containing representative structures from each subfamily. Every structure is shown with a side-view (left) and top-view (right) representation. PDB ID (from left to right and from top to bottom): 3D5K, 1QD6, 3GP6, 2X55, 2GR8, 1T16, 2GUF, 4K3B, 3RFZ, 4E1S, 2J1N, 2K0L, 1AF6 and 2QJR. Depending on the literature, the multifunctional protein OmpA (PDB ID 2K0L) can be found in different OMP sub-families [111-114] .*

Efflux pumps, also known as multidrug resistance (MDR) pumps, comprise a tripartite system that spans from the inner membrane to the OM. Efflux pumps are involved in antibiotic resistance where bacteria evolve mechanisms to survive harmful drugs [115-117]. Although the first identified efflux pump operon was the *P. aeruginosa mexAB* operon in 1993, yet Protein Data Bank (PDB) holds currently only four different OM efflux pump proteins[1] [118].

Enzymes are mainly involved in maintaining membrane integrity. They provide the first line of defense against external attack by host immune system molecules or antibiotics that perturb the membrane. Currently, there exist six different OM enzymes with solved 3D-structures [100, 119]: OM phospholipase A (OMPLA) which is likely involved in membrane disruption processes, lipid-A modifying enzymes that may reduce endotoxicity (LpxR, PagP, and PagL), and omptins that may inactivate host plasminogen (OmpT and Pla). OM enzymes will be further elaborated in section 1.3.2.

Transporters are a large and diverse class of OMPs that can be further sub-classified into four main groups: autotransporters [120], hydrophobic transporters [121, 122], siderophores [111] and TpsB transporters [123]. *Autotransporters* belong to the TpsB subclass of transporters, consisting of a variable C-terminal multifunctional passenger domain [124-127]. Both Yen *et al.* and Quin *et al.* have reviewed the class of autotransporters [96, 128]. Among the 800 homologue autotransporters identified, is the *H. pylori* VacA virulence factor [129]. Among *hydrophobic transporters* are FadL and Tsx. These transporters ensure uptake of lipids and nucleosides, respectively [122]. *Siderophores* are iron-uptake transporters. Since iron is a micronutrient vital for humans and bacteria, bacteria have adapted different mechanism to scavenge human iron, which was reviewed by Sheldon *et al.* [130].

Ushers are the largest OMPs with 24-strand β-barrels [131]. They are responsible for the assembly of pili and their translocation to the bacterial surface [132]. Pili are adhesive protein fibers that mediate attachment to host cells and tissues, for colonization and infection [133].

Virulence factors, previously defined as molecules involved in disease, are often multifunctional OMPs that may be classified in several groups. This is a diverse group that vary in size and function. Adhesion factors, *e.g.* intimin, may attach bacteria to other cells or surfaces

---

[1] PDB search [http://www.rcsb.org/pdb], March 2017.

and may be involved in host invasion mechanisms [102]. Some adhesins are virulence factors that can interact with host signaling pathways (*e.g.* the toll-like receptor pathway) [134-136]. The extracellular domain of these virulence factors may mimic eukaryotic proteins to facilitate host binding [137].

Porins[2] form the largest OMP subfamily. They are involved in the selective uptake of molecules. They may also function as a bacteriocin [138, 139], and they can be involved in host-cell interactions [94]. Porins are up-regulated in the presence of nutrient molecules, and down-regulated in the presence of toxins or other harmful molecules (*e.g.* antibiotics, heavy metals, detergents, or bile salts [140, 141]. *Escherichia coli* amyloid secretion channel, CsgG, forms the largest known OMP channel with a 36-stranded β-barrel (9x4-stranded monomers) [142, 143].

1.3 Involvement of OMPs in niche adaptation

Since OMPs are in direct contact with the environment, they must quickly detect and adapt to environmental changes [95]. This section will focus on how enzymes and porins are involved in niche adaptation, since these are the most relevant OMPs for this thesis.

1.3.1 Porins

Porins are often multifunctional, implicated in host-cell interactions [144] and as receptors for bacteriocins [138, 139]. They are transmembrane barrels usually comprising between eight and 24 β-strands connected by short periplasmic turns and longer extracellular loops (see Figure 3). Many porins, e.g. *E. coli* OmpF form tightly packed trimers through loop interactions [145]. Achouak *et al.* and Delcour review the structure of these highly stable trimeric porins that are detergent resistant and only dissociated by extreme environmental conditions [140, 146]. Extracellular loops may also be involved in host evasion through loop variability, resulting in a potential threat not recognized by the host immune system. Thus, loop variability may be one pathogenic mechanism; however, conserved protein binding sites can also reveal cooperative interactions [147]. Currently, many different classification systems for porins exist in the literature. However, they are all usually divided into either specific or non-specific, and monomeric or trimeric porins.

_____

[2] For simplicity, all non-specific and specific diffusion channels for hydrophilic proteins will be defined as porins.

Achouak *et al.* and Pages *et al.* reviewed porin regulation at the gene level [140, 141]. Gene expression of porins may be regulated based on nutrient uptake and can reach 10^6 copies per cells. Although nutrients may upregulate porin expression, toxins (*e.g.*, antibiotics, heavy metals, detergents, or bile salts) result in a tighter membrane with fewer porins. Nutritional limits and osmolality changes result in porin expression that quickly will adapt to OM permeability. The *ompF* and *ompC* genes (encoding for the General Diffusion Porins (GDPs) OmpF and OmpC, respectively) are correlated and regulated by environmental conditions. OmpF predominates at low osmolality levels, while the OmpC porin is inversely correlated. This is probably due to the difference in pore size. During low nutrition with low glucose concentrations in the environment, OmpC levels increase while both OmpF and maltoporin (coded by the *lamB* gene) protein synthesis are turned completely off. The bacteria adjust maltoporin expression by inducing gene expression of *lamB* during high concentrations of maltose/dextrin concentrations [140].

OmpA is one of the major *E. coli* OMPs, and Koebnik *et al.* estimated that there are 100,000 OmpA copies per cell [101]. According to Krishan *et al.*, OmpA is conserved among the Gram-negative species [148]. Smith *et al.* reviewed the multifunctional OmpA as a "Swiss-army knife" since it can contribute to adhesiveness, invasiveness, evasion or be involved in biofilm formation and conjugation. OmpA is also targeted by the immune system and is a bacteriophage receptor. These features are mainly conferred by the exterior loops [149]. OmpA has probably adapted loop mutations to evade the immune system and allow bacterial survival in very hostile environments, such as the brain. *E. coli* OmpA loops adhere to the microvascular cell surface located in the human brain, and OmpA has therefore been suggested to have role in the pathogenesis of bacterial meningitis [150]. Studies indicate OmpA pore activity, although this has been debated [149, 151-157].

Bacteria living in nutrient-limited environments have replaced porins with substrate-specific transporters. Although this is a more effective mechanism for capturing the substrates, the membrane permeability is drastically diminished which influences uptake of antibiotics. Antibiotic resistance studies from the United States indicated that *P. aeruginosa* and A*cinetobacter baumannii* are among the most difficult hospital infections to treat, likely due to the tightening of the OM-gating. Most of the small water-soluble molecules diffuse through OM carboxylic acid channel (Occ) proteins. They are primary found in *Pseudomonads* and related species and show structural and functional differences [98]. OccK5 is a small basic

multifunctional Occ protein that contributes to membrane stability, EDTA resistance and likely polymyxin B resistance when bound to Mg^{2+} [158].

1.3.2 Enzymes

Modification of the OM is one of the mechanisms that bacteria have evolved to evade the host immune system. OM enzymes are multifunctional proteins that likely also are involved in pathogenesis. Bishop *et al.* reviewed OM enzymes, while Qiao *et al.* described reduced host immune response through lipid A modifying enzymes or direct attack on the innate immune system [100, 159].

The first OM β-barrel enzyme to be characterized was OMPLA. OMPLA belongs to the large and diverse lipolytic enzyme family which catalyzes hydrolysis of phospholipids to lysophospholipids [160]. These enzyme vary in substrate preference, mode of action and regulation [161]. The *E. coli* OMPLA prevents uncontrolled breakdown of the surrounding phospholipids [162]. Snijder and Dijkstra suggested that OMPLA activity is triggered by diverse events such as environmental changes (temperature shift or heat shock), toxin release (polyxmixin, phage-induced lysis or colicin release) or detection of unstable membrane. They concluded that activity is correlated with loss of membrane integrity in *E. coli* [100, 160]. OMPLA is likely to destabilize the outer leaflet of OM during self-assisted cell lysis [163]. However, this could also be part of a life-saving protection system where the needed compounds are recycled [100, 164-166]. Unwanted cell lysis is prevented by the bacteria through tight regulation [167]. The solved OMPLA structure from *E. coli* reveals a dimeric regulated activation of the transmembrane enzyme [168]. This serine hydrolase is activated by membrane perturbation and requires a calcium-induced dimerization [169].

The catalytic activity of another OMP, *Salmonella enterica* serovar Typhimurium (hereafter *Salmonella* Typhimurium) LpxR, is predicted to be similar to that of OMPLA. LpxR activity is also phase-dependent and only activated when the bacteria reach the stationary phase [170]. LpxR reduces lipid A bioactivity which is a host evasion mechanism [171]. LpxR removes the 3′-acyloxyacyl moiety of the lipid A portion of lipopolysaccharides (LPS). LpxR has homologues in *H. pylori*, *Vibrio cholerae, E. coli* and *Yersinia enterolitica.* LpxR is growth dependent in these homologues, except in *H. pylori* where the enzyme is constitutively active [170].

The lipid A 3-O-deacylase, PagL, was crystallized from *P. aeruginosa.* It displays a 30° tilting in the lipid bilayer. The binding pocket is in a hydrophobic groove that is positioned perpendicular to the membrane plane. This multifunctional enzyme confers polymyxin B resistance when activated and attenuates lipid A endotoxicity, thus, reducing the host immune response [172-174].

PagP is a lipid acetyltransferase that transfers a palmitate chain to lipid A [175]. This enzyme guards the permeability barrier and responds to membrane perturbations induced by a $Mg2^{+}$-limited environment. It depends on lateral diffusion of phospholipids [176-178]. Bacterial resistance to antimicrobial agents derived from the host is enhanced by the ability of PagP to evade the host immune system [177, 179]. Another function of PagP is as an apical sensory transducer to detect perturbations caused by lipid asymmetry [179].

Omptins are a group of proteases situated in the OM (including Pla and OmpT). Pla is also a virulence factor required for bubonic plague establishment. Pla proteolytically cleaves and thus inactivates plasmin inhibitor, enabling cell migration. Human zymogen plasminogen and complement proteins are also among its substrates. It can function as an adhesin that binds to laminin (a mammalian glycoprotein), which is subsequently degraded by another plasmin and mediates invasion into human endothelial cells. The omptin family has a variety of different functions that likely spread among Gram-negative bacteria through HGT, followed by genetic adaptation to their host [180]. The aspartic protease OmpT cleaves paired basic amino acids [181-184]. *E. coli* OmpT is a multifunctional protein that regulates the biogenesis of OM vesicles (OMV), inactivates toxic peptides and is likely to enhance colonization [185]. OMV are secreted from enterohemorrhagic *E. coli* (EHEC) and implicated in the pathogenesis of an infection (caused by *e.g.* food poisoning) [185].

1.4 *H. pylori* OMPLA

Literature findings suggest that *H. pylori* OMPLA could be involved in colonization in the human gastric ventricle [186-189]. Dorrell *et al.* found that *H. pylori pldA* mutant did not colonize in mice, neither two nor eight weeks after infection [186]. Xerry and Owen discussed OMPLA's role in colonization, and analyzed the *pldA* gene in 124 samples using Restriction Fragment Length Polymorphism (RFLP). They found it remained conserved independently of the geographical origin of the isolates [190]. In 2006, Istvan *et al.* reviewed phospholipase's role in Gram-negative bacterial pathogenies, including how *H. pylori* phospholipase activity is linked to the degradation of the mucosal barrier [161].[3]

Tannæs *et al.* characterized phase variation due to DNA slippage in the *pldA* gene that resulted in either a complete (OMPLA$_{ON}$) or truncated protein (OMPLA$_{OFF}$) [191]. DNA slippage is due to a homopolymeric tract in the gene, where G7-tract results in a truncated OMPLA and an G8-tract yields a full-length OMPLA. This homopolymeric tract was found in all clinical isolates of *H. pylori* sequenced by Tannæs *et al.* [191].

Only the variants with an intact OMPLA, OMPLA$_{ON}$-variants, survive prolonged acid exposure. At neutral pH, both OMPLA$_{ON}$-and OMPLA$_{OFF}$-variants may survive. However, OMPLA$_{ON}$-variants display an altered lipid composition. This is independent of enzyme activity, since OMPLA has an enzyme optimum at pH 7 and is inactive at pH 5 [88].

---

[3] Istvan discuss three *H. pylori* phospholipases (PLA1, A2 and C). However, only OMPLA is found in the *H. pylori* genome (gene search using GenBank files).

2. Aims of the Study

The overriding goal of this study was to construct a model that could best describe *H. pylori* OMPLA structure and function(s). All bacterial OM proteins have a common β-barrel motif. We intend to examine this motif with regard to OMP multifunctionality and niche adaptation, focusing on porin function. Our hypothesis was that *H. pylori* OMPLA is adapted to protect the bacteria in acidic environment of the gastric mucosa. The possibility that this multifunctional protein may vary its function depending on pH-level was explored.

Objective 1: Bacterial OMPs. All bacterial OMPs have the same β-barrel fold, but their size and function vary greatly. We wanted to examine the architecture of multifunctional OMPs from Gram-negative bacteria, focusing on how niche adapted function affected the structure. The β-barrel fold allows a hollow pore that often regulate molecules passing through the membrane. An important question here was whether the hollow *H. pylori* OMPLA structure could form a pore which would allow diffusion or transport of molecules.

Objective 2: Comparative *Helicobacter* OMPLA sequence analyzes. To better understand OMPLA's role in *H. pylori*, we wanted to study sequences available from different *Helicobacter* species. We would use the sequences of *pldA* and its neighboring genes, and compare gastric and enterohepatic *Helicobacters*.

Objective 3: *H. pylori* OMPLA 3D model structure. In order to study structure-function related questions regarding *H. pylori* OMPLA, we wanted to construct a theoretical 3D structure model by implementing the previously described objectives. Available structures from similar species and orthologous sequences, combined with literature findings and previous work from our group were included in this model.

This thesis aims are discussed in the papers listed below (hereby referred to by their Roman numerals):

Paper I: *In silico* structure and sequence analyzes of bacterial porins and specific diffusion channels for hydrophilic molecules: Conservation, multimericity and multifunctionality. Vollan HS, Tannæs T, Vriend G, Bukholm G. Int. J. of Mol. Sci. 2016; 17(4):599.

Paper II: *In silico* evolutionary analyzes of *Helicobacter pylori* outer membrane phospholipase A (OMPLA). Vollan HS, Tannaes T, Yamaoka Y, Bukholm G. BMC Microbiol. 2012; 12:206.

Paper III: Outer membrane phospholipase A's roles in *H. pylori* acid adaptation. Vollan HS, Tannæs T, Caugant DA, Vriend G, Bukholm G. Gut Pathog. 2017; 9:36.

3. Material and Methods

3.1 Softwares

NCBI, PubMed [192], Google Scholar [193], Thomson Reuters Web of Science™ [194], and Ovid MEDLINE [195] were used for literature searches. Outer membrane protein database (OMPdb) [196], Transporter classification database (TCDB) [197], Protein families' database (Pfam) [198], Class Architecture Topology Homology (CATH) [199] and Structural classification of proteins-extended (SCOPe) [200] were databases used to identify current OMP classifications. All sequences were collected using Basic Local Alignment Search Tool (BLAST) [201], *H. pylori* Multi Locus Sequence Typing (MLST) [202] [http://pubmlst.org/helicobacter/], MRS [203], Protein Information Resource (PIR) [http://pir.georgetown.edu] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [204]. Sequence retrieval was achieved using NCBI Batch Entrez [http://www.ncbi.nlm.nih.gov/sites/batchentrez] (when NCBI Gene or Protein ID was available). All structures were collected and downloaded from PDB [205]. BioEdit [206] was mainly used to extract DNA or protein regions and view sequences, CAIcal [207], EBI ClustalW / ClustalΩ [208-210], MAFFT [211], Gblocks, SWAAP 1.0.3 [212], TOPD/FMTS [213], and Sequence Manipulation Suite (reverse complement calculator) [214] were sequence tools used in Paper II. Most protein structure visualization and analyzes were performed using the YASARA-WHAT IF twinset [215, 216]; structure superposition was done in YASARA using MUSTANG pairwise motif aligner (or the MUSTANG multi-aligner for visualization purposes of each subfamily) [217]. Furthermore, homology models were constructed using either the WHAT IF servers [218] or running the Yasara Structure script hm_build.mcr [219]; homology model structures were optimized using the Yasara Structure scripts md_refine.mcr and md_runmembranefast.mcr [215, 219]; signal sequences prediction using NetSurfP [220]; pore size estimation using HOLE [221] and WHAT IF [216, 222]; pore channels were further characterized using PoreWalker [223]. Finally, FgenesB (Softberry Inc., Mount Kisco, NY, USA) and ProOpDB [224, 225] were used to predict operons in Paper III.

3.2 Protein structure analyzes (Papers I and III)

Protein sequences and structures used in Papers I and III were analyzed using the YASARA-WHAT IF twinset [215, 216]. YASARA is a visualization, modeling and simulation software used mainly to analyze protein structures, while WHAT IF is a molecular modelling package that includes tools to display and manipulate sequences used to construct both profile alignments, multiple sequence alignments (MSA) and studying these alignments through entropy-variability analyzes (EVA).

3.2.1 Data collection

Bioinformatics has become a large field of science in which biological data is interpreted. The amount of data from life sciences has increased greatly in the past decades and the analyzes of data resulting from the many novel high-throughput methods require specialized software and training. For example, there exist over 500,000 sequences belonging to the protein family of bacterial porins in the NCBI database. Collecting and filtering data are important steps in constructing a good alignment. All relevant structures and sequences were collected to make the highest quality possible of the homology model structure.

3.2.2 Profile-based structure alignment

Profile-based structure alignment is based on an iterative process where a vast amount of sequences, or all sequences available, are used. Low-identity, which are not relevant sequences, are filtered out after aligned to a profile (227). A structure-based profile alignment improves a regular sequence alignment because it is based on all available sequences and structures for a given protein family. This yields important information on vital residues needed for function (*e.g.* active site residues or calcium-binding sites).

The effect of an amino acid substitution varies depending on which amino acid it is mutated into. One would expect similar residues (*e.g.* Ile to Leu) to cause least harm due to quite similar side chains (see Appendix Figure 1). This is because these types of mutation would have little impact on the protein and how it functions, which would likely result in an overall unchanged fitness for the bacteria. Therefore, a sequence alignment that is comparing different protein sequences should punish similar residue substitutions less than those one would expect to cause dramatically different properties. However, this is just a general model that neither considers location nor function of the residue (*e.g.* lipid-interacting residue vs active site residue). A mutation in the active site could cause great harm for the protein, resulting in decreased fitness

of the organism. This will be discussed later in the profile alignment section. A generalized scoring matrix is used to weigh different mutations. The actual likelihood of each substitution occurring during evolution is impossible to estimate, but there exist several substitution matrices. Each approach has its advantages, depending on which protein family to be analyzed (*e.g.* membrane protein vs. globular cytoplasmic protein).

A profile based structure alignment starts with a residue exchange matrix where important features of a protein class are adjusted for in a profile alignment. Core structure motifs, *e.g.* β-strands or α-helices derived from template structure, are implemented in the scoring matrix. The template sequence is then aligned together with the collected sequences. This alignment procedure contains crucial structural information not implemented in standard sequence alignments.

The iterative steps in these analyzes aim to optimize the alignment until the best possible alignment between two sequences is generated. It is important that the profile represents the subclass one wants to analyze, so the position of both the template and model sequences are carefully monitored for iterative each step. The resulting MSA can be generated in mview-format where residues are colored according to side-chain properties (see Appendix Figure 1). This makes it easy to find conserved and variable regions in an alignment with thousands of sequences.

Homology models were constructed by uploading the aligned model sequence extracted from the MSA, the template sequence, and the template 3D structure files to the WHAT IF Homology Modelling server (see link under Softwares section).

3.2.3 Entropy-variability analyzes (EVA)

EVA in YASARA/WHAT IF twinset [215, 216] was used to study the variability in an alignment by developing an evolutionary model. EVA is based on well-established experimental methods from multiple, large protein families: globin chains, G protein-coupled receptors, Ras-like proteins, and serine proteases [226, 227]. Signal transduction residues have also been identified in other protein families, including the nuclear receptor family [227]. The information derived from the EVA can be mapped onto the template or model structure, and determine which regions are most conserved or highly variable

Variability patterns for individual residue positions in MSAs can be defined by either the Shannon entropy (E_i[4]) or by the number of amino acid types observed (more than 0.5%) at position i (V_i). The entropy, E_i, can be plotted against the variability, V_i, for all residue positions in the MSA in an Entropy-Variability (EV) plot. This EV plot can be divided into five boxes. Figure 7 is an empty EV plot used as an example. A real EV plot would be filled with dots where each dot would represent an amino acid position in the final MSA (*e.g.* Papers I and III).
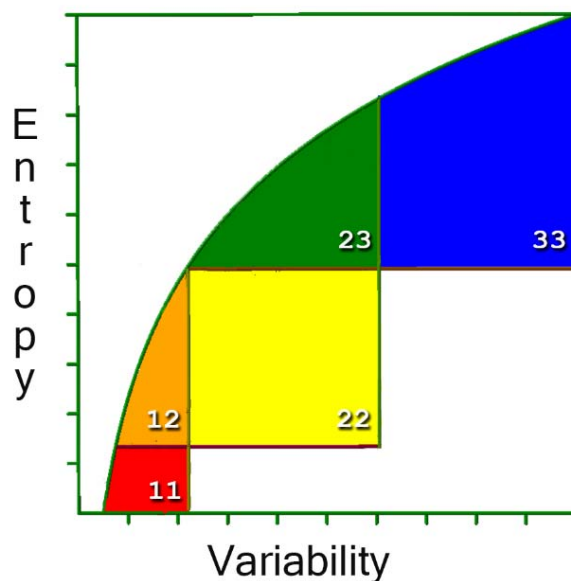
Figure 7: Entropy-Variability plot. This is an empty EV plot to illustrate the five boxes in an EVA.

Each of the five boxes, as illustrated in Figure 7, contains residues involved in mainly one broad functional category [226, 228]: Box 11 (low entropy and low variability, contains residues in the main active site, colored red), Box 12 (intermediate entropy and low variability, contains residues that support the structure of the main active site, colored orange and often situated next to the red residues), Box 22 (intermediate entropy and intermediate variability, contains residues involved in communication between the main active site and regulatory sites, colored yellow), Box 23 (high entropy and intermediate variability, contains residues involved in regulation of protein activity, colored green), and Box 33 (high entropy and high variability, contains residues for which no function is known, colored blue).

---

[4] E_i is defined as $\Sigma Plog(P)$, where P represents the frequency of occurrence of a given amino acid in the MSA at position i.

3.3 Pore size estimations (Paper I and III)

In Paper I, porin pore sizes were estimated for each subclass using HOLE [221]. This software uses Monte Carlo simulation to route through the channel. This software was created to characterize ion channels in 1993.

In Paper III, we calculate the pore sizes for a spherical probe using the WHAT IF software [216, 222]. All residues in the structure were mutated to alanine to ensure a realistic size not occluded by flexible side chains. The OMPLA pore was further characterized with PoreWalker [223]. This is an automated method that analyzes residue composition, pore lining and shape of the pore. The predicted path of a water molecule is estimated and embedded in the structure file.

3.4 Operon predictions (Paper III)

Available ProOpDB [224, 225] operon predictions were used (see Figure 8 in Paper III). In order to compare gastric and enterohepatic *Helicobacter pldA* operons FgenesB (Softberry Inc., Mount Kisco, NY, US) was used (see Table 3, Paper III). FgenesB is included in the package for bacterial annotation pipeline, and has been widely tested.

3.5 HGT analyzes (Paper II)

Genome instability is necessary for natural selection in a healthy population. HGT between members of a species in the same ecosystem facilitates transfer of genes useful in that environment. This may be genes coding for virulence, antibiotic resistance, etc. [229]. Bioinformatic analyzes may detect discrepancies in the AT: GC ratio (comparing one gene to the average genome) or through phylogenetic analyzes. A reference (HK genes or 16S rRNA) is often compared to a gene of interest. This will reveal the evolutionary path of the gene of interest and may indicate its origin if HGT did occur. Three different methods were used to test for HGT in Paper II: (1) DNA stability, (2) Codon Adaptation Index (CAI) and (3) Phylogenetic analyzes. Software used includes CAIcal [207], SWAAP 1.0.3 [212], and TOPD/FMTS [213]. Different methods were used to detect possible adaptive evolutionary sites by calculating the number of synonymous and nonsynonymous substitutions per site.

4. Summary of results

4.1 Overview of papers

Paper I: Porins are involved in the selective uptake of nutrients and form the largest OMP family. A new classification scheme was made to accommodate function and size of this protein family. Structure and sequence based analyzes revealed conserved interaction sites and variable loops. The high mutation rate observed in surface-exposed loops is likely an important mechanism for host immune system evasion. We observed a pattern for the trimeric proteins in all analyzed subclasses. This suggests that all proteins found in this family exist in a multimeric state.

Paper II: Literature has implicated the *pldA* gene, encoding OMPLA, in *H. pylori* colonization of the human gastric ventricle. We examined sequence variation in 227 isolates collected from Norway and Korea, and found biogeographic patterns. Our findings indicated a conserved *pldA* gene. The bacterium is preserving the function of OMPLA, although some sites are still being evolutionarily optimized. Bioinformatic analyzes did indicate a possible HGT. However, since the gene showed biogeographic patterns as the HK genes, the transfer likely occurred long ago.

Paper III: We constructed a 3D model of *H. pylori* OMPLA and examined how this protein is needed (but enzymatically inactive) in acidic environment. This led to the discovery of a polar core with an approximately 4 Å pore diameter, which could be involved in urea or ammonium diffusion across the membrane. A niche-specific extracellular loop was observed among acid-tolerant *Helicobacter* species. We propose a model of OMPLA multi-functionality in *H. pylori* that enables survival in acidic environment. Finally, *pldA* is in a conserved operon with two inner membrane transporters in gastric *Helicobacters*. This operon is lacking in enterohepatic species, indicating that these three genes are required for survival in the gastric mucosa. We constructed models for these inner membrane channels, and suggested that they could be involved in ammonium/ammonia efflux from cytosol to periplasm.

4.2 OMP subfamilies

Multifunctional OMPs with niche specific properties were of interest due to the common β-barrel motif located in the bacterial OM. We collected available OMP literature and structure data, with an emphasis on structure-function related specific niche information (*e.g.* analyzing known mutations or structural changes).

The 95 solved structures are listed in Table 1 with protein name, subfamily name, and size (number of β-strands found in the transmembrane β-barrel structure). The alignments discussed here are based on structure alignments of the entire proteins, which are used to compare structures within one subfamily (see Appendix for details). This information was used to better understand structure-functions of OMPs. Table 1 holds the most recent list over OMP structures deposited in the PDB database before 12th March 2017. The newest structures are highlighted in red, but these were not included in the structure analyzes found in the Appendix.

Table 1: OMPs. This table lists all solved OMP structures with protein name and size for each subfamily. A complete PDB list is found in the appendix. Column name "Size" lists the number of β-strands found in a β-barrel protein structure. The structures highlighted in red are included in the thesis, but have not been analyzed in Appendix. They were deposited between 2014 (first data collection) and 2017 (last data collection).

OMP Subfamily	OMP Class	Protein name	Size
Efflux pump		CusC, CmeC, VceC, OprM, OprN, and TolC	12 (trimer)
Enzymes		LpxR, OMPLA (EcOMPLA, StOMPLA), OmpT, PagL, PagP, and Pla	8-12
Transporters	Autotransporter	AlgE, EstA, EspP, HbP, Hia, and NalP	12-18
	Hydrophobic channel	COG4313, FadL (EcFadL, PaFadL), LptD (SeLptD, SfLptD, EcLptD, KpLptD, PaLptD, YpLptD), OmpW, OprG, TbuX, TodX, and Tsx	8-14
	Siderophore	Cir, BtuB, FauA, FecA, FepA, FhuA, FptA, FpvA, FusA, HasR, HpuA, PirA (PaPirA, AbPirA), PiuA, ShuA, and ZnuD	22
	TpsB transporter	FhaC, BamA (NgBamA, HdBamA), SusD, TamA and TbpA	16-22
	Other specific transporters	PorB (use host ATP)	18
Ushers		FimD and PapC	22
Adhesins	Virulence/Niche factor	Ail, Intimin, Invasin, NspA, OmpX, Opa$_{60}$, OpcA, Pallilysin, Wzi (and PorB).	8-18
Porins	Specific	CymA, CsgG, EcMaltoporin, NanC, KdgM, Occ[5], OprP, OprO, OprB, PgaA, ScrY, and StMaltoporin,	8-36
	Non-specific	CarO, GDP (RcGDP, RbGDP), Omp32, OmpA (EcOmpA, KpOmpA), OmpC (EcOmpC, StOmpC, OmpK36, OmpE36), OmpF (EcOmpF, StOmpF), OmpG, PhoE	8-18
Others	Unknown function	RPA1785, TtoA, and UPF0311	8-10

---

[5] The Occ channels include 14 solved *P. aeruginosa* porin structures listed in Appendix Table 8.

4.3 Porins: structure-function relationship

OMP are gatekeepers that hold a hollow transmembrane structure that is often used for molecules diffusion/transport. The largest OMP family is porins (see 1.2.3 for definition). Pore specificity is determined by amino acid composition and pore size [95, 101].

The gating mechanism of the larger porins is achieved through a constriction loop. This is usually the largest extracellular loop, loop L3, that is located inside the barrel. Only small structural changes in the porin constriction site can result in different substrate preferences. For example, the PhoE constriction site has a charged Lysine residue where OmpF has a Glycine residue. This mutation results in anionic selectivity of PhoE, while OmpF preferers cationic substrates. *E. coli* PhoE expression occurs during phosphate starvation, the phosphoporin is only anion-selective, lacking any particular phosphate-specificity [94]. The constriction site is just big enough for a glucose molecule to pass through without any steric hindrance that would affect selectivity [230].

Another example of pore specificity includes OmpF and OmpC porin homologs, *Klebsiella pneumoniae* OmpK35 and OmpK36, respectively [140, 231]. Although OmpK36 has lower conductance than OmpC, it has a higher cation selectivity than OmpF. The most variable regions between OmpF and OmpK36 were found in the loop regions. OmpK36 protein expression is preferred at high osmotic strength; this may be explained by the increase in charge density that would change the electro-physical properties of the osmoporin pore [140, 231].

The *Salmonella* Typhimurium ScrY sugar-specific porin is structurally similar to *E. coli* maltoporin, but ScrY is less occluded because of shorter exterior loops. Both structures contain aromatic residues that form the 'greasy slide' in the binding sites [232], but three residues differ in the constriction site. This results in ScrY specificity that does not allow the passing of maltodextrin. Sucrose molecules inside the maltoporin pore get stuck and block ion passage [233]. Furthermore, high sequence variation is found in the exterior loops. This allows niche-specific adaptation while conserving important structural elements.

The OmpG porin differs from the other porins because it is a monomeric 14-stranded barrel and only small amounts are expressed in *E. coli*. It facilitates sugar-uptake if either ScrY or the maltoporin gene (LamB) has been inactivated or if the genes are not present [234-236]. It controls membrane permeability in a pH-dependent manner where one of the exterior loops closes the channel in acidic conditions [237, 238]. The rest of the protein remains unaffected,

demonstrating its stability even in acidic environments [237]. The loops are flexible and they adopt different conformations when the pH is changed; four of the loops break the hydrogen bonding network, resulting in greater flexibility [239]. The inner face of OmpG contains a lining of aromatic residues that guides the sugar molecules through the channel and into the periplasm; furthermore, OmpG lacks the common L3 constriction loop [235, 240].

The versatile OmpA protein (see introduction) is a small protein with 8 β-strands and long extracellular loops. These loops are vital for OmpA functions, *e.g.* differences in loop regions L2 observed between OmpA1 and OmpA2 may be the difference of an invasive and less invasive *E. coli* strain [241]. OmpA homologues include the psychrophilic fish pathogen *Moritella viscosa*. MvOmp1 is predicted containing a barrel with highly variable extracellular loops (making vaccine development difficult) [242]. The *P. aeruginosa* OprF virulence factor has functionally and structurally similarities to OmpA [243]. In *K. pneumoniae* OmpA homologue, KpOmpA, has adapted to a pathogenic survival in the hostile lung environment using OmpA as an evasin. Structural dynamics analysis of *K. pneumoniae* KpOmpA demonstrated that the longest loops, L1 and L3, are highly mobile. The remaining L2 and L4 loops, the most conserved loops, display restricted motions. These results relate dynamics and mobility with sequence variations in the loops [244].

Thorough investigation of porin literature and protein structures analyzes resulted in a new classification system shown in Table 2 (see Table 3 in Paper I for more details). This is a system based on size and specificity of the porins, but does not take multimericity into account. The WHAT IF estimated pore sizes of these eight subclasses ranges from 1.5 Å to 10 Å: 1.5 Å (Class 1), 5.5 Å (Class 3B), 5.9 Å (Class 4A), 7.4 Å (Class 5A), 8.5 Å (Class 5B), 8.0 Å (Class 5C), 8.1 Å (Class 6B) and 10.0 Å (Class 6C).

Table 2: The non-specific porin and specific diffusion channel family classification system. The non-specific porin and specific diffusion channel family distributed among six classes and further divided into eight subclasses. Examples of each subclass is listed under the "Protein name"-column. This classification system is based on size and specificity, where empty classes and subclasses (annotated "-") are reserved for future structures. See Paper I for more details.

Class #	Size	Subclass #	Subclass name	Protein name
1	8	1A	Non-specific, petite porin	OmpA
2	10	2A	Non-specific, mini porin	–
3	12	3A	Non-specific, small porin	–
		3B	Oligogalacturonate-specific, small channel	NanC
4	14	4A	Non-specific, intermediate porin	OmpG
5	16	5A	Non-specific, medium porin	OmpC
		5B	Sugar-specific, medium channel	OprB
		5C	Phosphate-specific, medium channel	OprP
6	18	6A	Non-specific, large porin	–
		6B	Sugar-specific, large channel	ScrY
		6C	Carboxyl-specific, large channel	OccK1

Six new structures have been added to the PDB since Paper I was published: CarO (Class 1A, 8-stranded non-specific porin), CymA (Class 4B, 14-stranded specific porin), OmpE36 (Class 4A, 14-stranded non-specific porin), PgaA (Class 4C, 14-stranded dPNAG[5]-specific porin), OprO (Class 5C, 16-stranded phosphate-specific porin), CsgG (Class 15B; 36-stranded specific porin). There are now 7 different porin classes, and 11 subclasses.

---

[5] dPNAG: N-acetylated poly-β-1,6-N-acetyl-d-glucosamine

4.4 *H. pylori* OMPLA model

A 3D model of *H. pylori* OMPLA structure was constructed based on an alignment with a known structure (see Paper III). The homologous protein structure from *E. coli* was used as a template. OMPLA sequences from a wide variety of species were used in the multiple sequence analyzes, see Figure 8. The OMPLA$_{ON}$ transcript may form functional OMP structures that may function as either porins in acidic environment or phospholipase enzyme in neutral environments, likely stabilized as trimers.
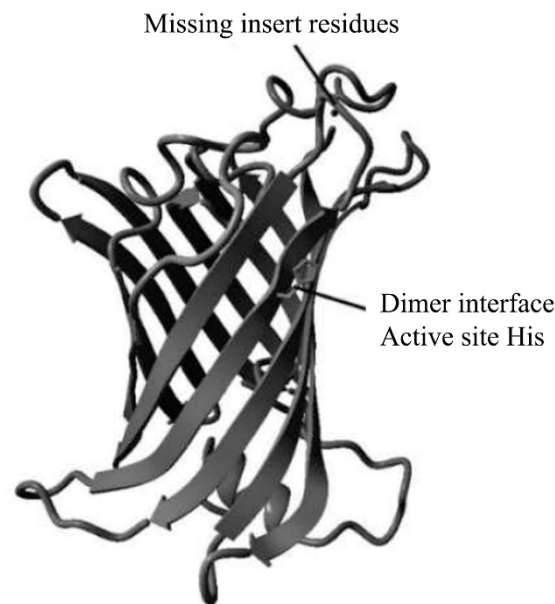
Figure 8: H. pylori OMPLA model. The H. pylori model structure was constructed based on the multiple sequence alignment with E. coli OMPLA (PDB ID: 1QD5) as a template structure. This model is visualized as a monomer for clarity, however, we predict a trimer model will create more stability.

Pore size estimation can be calculated by finding the largest sphere around a protein by using the WHAT IF software [216, 222]. A pseudo density map is contoured to create a molecular surface for a probe of a certain size, and that size can be adjusted to correspond to the pore. The probe size is simply by trial and error iterated till it nearly closes the pore. These pore calculations are based on spheres, so elliptical pores like OMPLA may let through non-spherical probes that are larger than the estimated radius. The *H. pylori* OMPLA pore size was estimated to be approximately 4 Å, see Figure 9.
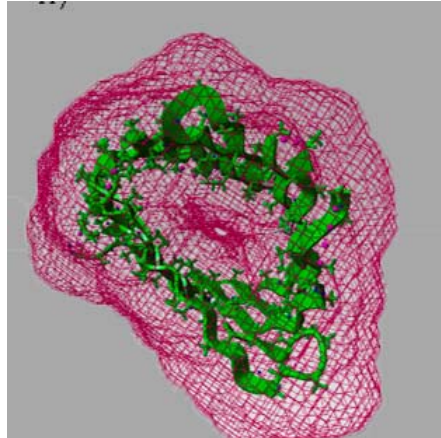
Figure 9: OMPLA pore function prediction. *YASARA's visualization of the WHAT IF calculated excluded volume around the poly-alanine mutated H. pylori OMPLA*

The OMPLA$_{OFF}$ strains lack OMPLA activity which could be due to a protein lacking the barrel-core, resulting in neither pore activity nor enzyme activity in the putative trimer interface. A putative model is shown in Figure 10, although the fold might differ since it might deviate from the template structure. However, it is more likely that OMPLA$_{OFF}$ transcripts might be detected as a potential deleterious polypeptides (premature proteins) and tagged for degradation [86].
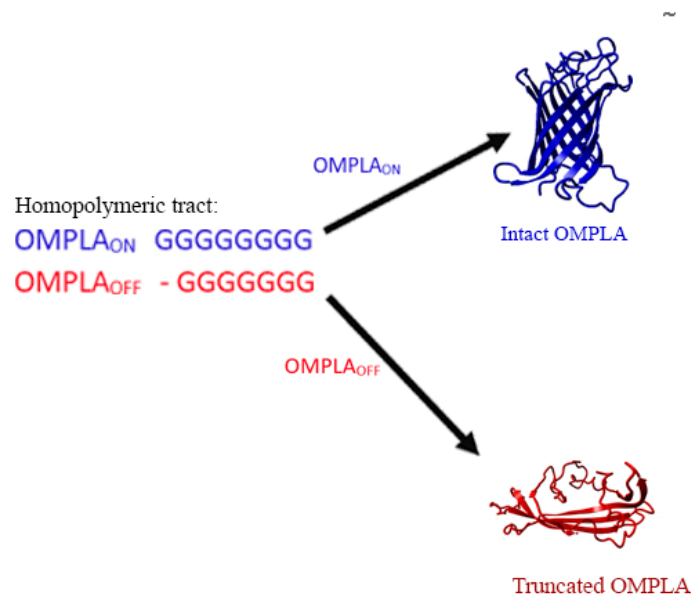
Figure 10: The pldA phase variation model. *The pldA gene encodes the OMPLA protein, however, due to a homopolymeric G-tract (illustrated to the left) it may be transcribed as a pldA$_{ON}$ (intact, active OMPLA) or pldA$_{OFF}$ (truncated). The truncated OMPLA model seem to lack pore and enzyme function, and it is likely that it is degraded. The active OMPLA can be enzymatically active with a pH optimum at 7, or may serve as a pore in acidic environments. Previous studies indicate that porin are most likely to form at trimeric conformation as it stabilizes the structure, but is visualized as a monomer for clarity in this figure.*

4.5 *pldA* operon

In Paper III, *pldA* operons are compared to different species. The ProOpDB was used for the initial comparison of COG2829 (Outer membrane phospholipase A), while the FgenesB software [http://www.softberry.com/img/help/pipelines/fgenesB-A.png] (Softberry Inc., Mount Kisco, NY, US) was used to predict *pldA* operons in species not found in the ProOpDB. The FgenesB software is recommended for the automatic annotation pipeline for bacterial genome annotation, but requires manual curation of possible gene operons. Table 3 in Paper III, compares gastric *Helicobacters* with enterohepatic *Helicobacters.* Our result showed a unique *pldA* operon in all gastric *Helicobacters*. The operon prediction results from *Helicobacter* genus show variations, but all gastric bacteria have two transporters and one OMPLA clustered in one operon. Looking at all *H. pylori* sequences, the operon includes DNA replication genes (the sliding clamp and GyrB). Some operons also contain enzymes implicated in the R-M system. The operon organization show similar results to the length of the non-modeled loop when comparing gastric *Helicobacters* to enterohepatic *Helicobacters*. See Figure 11 for visualization of *H. pylori pldA* operon.
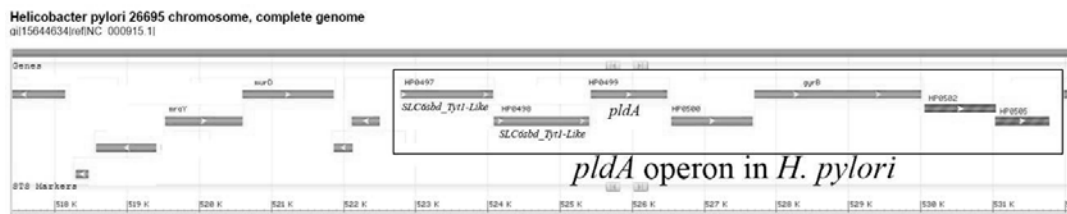
Figure 11: pldA operon. *This is an example of the seven genes likely co-regulated with pldA in H. pylori 26695. The three first genes (2 SLC6sbd_Tyr1-Like genes and pldA) are conserved in gastric Helicobacters. This graph is extracted from the NCBI website[6].*

Throughout these analyses there are two gastric *Helicobacter* outliers (*H. mustelae* and *H. himalayensis*). They behave like enterohepatic *Helicobacters* in all analyses, and are likely to have evolved a different tolerance mechanism. This is further elaborated in Paper III.

---

[6] [https://www.ncbi.nlm.nih.gov/projects/sviewer/?id=NC_000915.1&v=525424:526491]; Accessed: 05/06/2017

Figure 12 show an example using the FgenesB output for the predicted *H. pylori* 26695 *pldA* operon bacterial genome annotation.

```
Prediction of potential genes in microbial  genomes
Time:    Tue Jan  1 00:00:00 2005
Seq name: AE000511.1 Helicobacter pylori 26695 Positions 522514 to 530008
Length of sequence - 7495 bp
Number of predicted genes - 5
Number of transcription units - 1, operons - 1
    N      Tu/Op  Conserved  S           Start       End    Score
                  pairs(N/Pv)
    1     1 Op  1      .      +   CDS       229 -      1557   1706
    2     1 Op  2      .      +   CDS      1568 -      2896   1349
    3     1 Op  3      .      +   CDS      2911 -      3978   1233
    4     1 Op  4      .      +   CDS      4036 -      5160   1470
    5     1 Op  5      .      +   CDS      5173 -      7494   2681
Predicted protein(s):
>GENE      1       229  -     1557   1706      442 aa, chain +
MGNHFSKLGFVLAALGSAIGLGHIWRFPYMTGVSGGGAFVLLFLFLSLSVGAAMFIAEML
LGQSTQKNVTEAFKELDINPKKRWKYAGLLLVSGPLILTFYGTILGWVLYYLVSVSFNLP
NNIQESEQIFTQTLQSIGLQSIGLFSVLLITGWIVSRGIKEGIEKLNLVLMPLLFATFFG
LLFYAMSMDSFSKAFHFMFDFKPKDLTSQVFTYSLGQVFFSLSIGLGINITYAAVTDKTQ
NLLKSTIWVVLSGILISLVAGLMIFTFVFEYGANVSQGTGLIFTSLPVVFGQMGAIGILV
SILFLLALAFAGITSTVALLEPSVMYLTERYQYSRFKVTWGLVALIFVVGVVLIFSLHKD
YKDYLTFFEKSLFDWLDFASSTIIMPLGGMATFIFMGWVLKKEKLRLLSVHFLGPKLFAT
WYFLLKYITPLIVFSIWLSKIY
>GENE      2      1568  -     2896   1349      442 aa, chain +
MGKFSKLGFILATLGSSIGLGHIWRFPYMVGHNGGSAFVLLYLVLTLSLGIAMLLVEMLI
GNLGKKDVVSNYQILDPKRKKYYPFTSFFILGGPLILSFYAVVLGWVLYYLFVVTFDLPK
DLEQAKMQFSMLQNGSLIWPVIGFSACLLPTIWFVSRGIEEGIEKLNVVLMPLLFVIFIG
LLIYAMTLESMPKALHFLFNFEIQKIDFKVVMDALGQMFFSLSLGVGTIITYSAFTPKKE
NLFKSSLFIVLPGILISLIAGVMIFTFVFEYHADVSQGPGLVFISLPLTFAKMGMSGQIV
SLFFFMALVFAGITSTVSLIEPLALYLINRFNFSRLQASLWIGVVVYVLGVLVILSMNER
YAKFLSFAHKSVFGWLDFITSSFLMPLGGGLFSVLFIGWILNKKRSFLATKHFFNANAFKA
WHFSVRFIAPVVILAIFILQFK
>GENE      3      2911  -     3978   1233      355 aa, chain +
MKSILLFMIFVVCQLEGKKFSQDNFKVDYNYYLRKQDLHIIKTQNDLSNSWYLPPQKAPK
EHSWVDFAKKYLNMMDYLGTYFLPFYHSFTPIFQWYHPNINPYQRNEFKFQISFRVPVFR
HILWTKGTLYLAYTQTDWFQIYNDPQSAPMRMMNFMPELIYVYPINFKPFGGKIGNFSEI
WIGWQHISNGVGGAQCYQPFNKEGNPENQFPGQPVIVKDYNGQKDVRWGGCRSVSAGQRP
VFRLVWEKGGLKIMVAYWPYVPYDQSNPNLIDYMGYGNAKIDYRRGRHHFELQLYDIFTQ
YWRYDRWHGAFRLGYTYRINPFVGIYAQWFNGYGDGLYEYDVFSNRIGVGIRLNP
>GENE      4      4036  -     5160   1470      374 aa, chain +
MKISVSKNDLENALRYLQAFLDKKDASSIASHIHLEVIKEKLFLKASDSDIGLKSYIFTQ
SSDKEGVGTINGKKFLDIISCLKDSNIILETKDDSLAIKQNKSSFKLPMFDADEFPEFPV
IDPKVSIEVNAPFLVDAFKKIAPVIEQTSHKRELAGILMQFDQKHQTLSVVGTDTKRLSY
TQLEKISIHSTEEDISCILPKRALLEILKLFYENFSFKSDGMLAVIENEMHTFFTKLIDG
NYPDYQKILPKEYISSFTLGKEEFKESIKLCSSLSSTIKLTLEKNNALFESLDSEHSETA
KTSVEIEKGLDIEKAFHLGVNAKFFLEALNALGTTQFVLRCNEPSSPFLIQESLDEKQSH
LNAKISTLMMPITL
>GENE      5      5173  -     7494   2681      773 aa, chain +
MQNYQSHSIKVLKGLEGVRKRPGMYIGDTNVGGLHHMVYEVVDNAVDESMAGFCDTINIT
LTDEGSCIVEDNGRGIPVDIHPTEKIPACTVVLTILHAGGKFDNDTYKVSGGLHGVGVSV
VNALSKRLIMTIKKEGQIYRQEFEKGIPTSELEIIGKTKSAKESGTTIEFFPDESVMEVV
EFQAGILQKRFKEMAYLNDGLKISFKEEKTQLQETYFYEDGLKQFVKDSAKKELLTPIIS
FKSMDEETRTSIEVALAYADDYNENTLSFVNNIKTSEGGTHEAGFKMGLSKAILQYIGNN
IKTKESRPISEDIKEGLIAVVSLKMSEPLFEGQTKSKLGSSYARALVSKLVYDKIHQFLE
ENPNEAKIIANKALLAAKAREASKKARELTRKKDNLSVGTLPGKLADCQSKDPLESEIFL
VEGDSAGGSAKQGRDRVFQAILPLKGKILNVEKSHLSKILKSEEIKNMITAFGCGIQESF
DIERLRYHKIIIMTDADVDGSHIQTLLMTFFYRYLRPLIEQGHVYIAQAPLYKYKKGKTE
IYLKDSVALDHFLIEHGINSVDIEGIGKNDLMNLLKVARHYRYALLELEKRYNLLEILRF
LIETKDALSLDMKVLEKSILEKLEGLNYQILRSFATEESLHLHTQTPKGLVEFNLDDNLF
KEVLFEEANYTYQKLMEYNLDFLENKDILAFLEEVENHAKKGANIQRYKGLGEMNPNDLW
ETTMHKENRSLIKLKIEDLEKTDAVFSLCMGDEVEPRRAFIQAHAKDVKQLDV|
```

Figure 12: Operon prediction using FgenesB. Softberry's FgenesB operon prediction for H. pylori 26695 sequence positions 522514 to 530008. FgenesB predicts one operon and one transcription unit that consist of five genes: Gene 1 and 2 are "sodium- and chloride dependent transporters" (COG0733), Gene 3 is OMPLA, Gene 4 is DNA polymerase sliding clamp and Gene 5 is GyrB.

4.6 AmCI and AmCII

In Paper III, we constructed models of the proteins encoded by the two genes downstream of the *pldA* gene, conserved among gastric *Helicobacters*. These two genes both belong to the *SLC6sbd_Tyt1-Like* genes, which is part of the "Na+-dependent transporters (channels) of the SNF family". These channels will herby be referred to as Ammonium Channels I and II (AmCI and AmCII), and their models were constructed as described for OMPLA.

AmCI and AmCI had approximately 50% sequence identity to each other. The protein sequences have highest homology to the MhsT structure, from *Bacillus halodurans* belonging to the same family as AmCI and AmCII. The MhsT template had between 30-35% sequence identity to the two model sequences. 3D models were made to better study possible function of these proteins, and our results indicate that they could be involved in ammonium (and possibly ammonia) efflux. Visual inspection of the two channels yielded quite similar proteins. However, AmCI seemed to contain more polar pore residues when compared to AmCII. This could yield different substrate preference.

5. Discussion

5.1 Overview

In this thesis, *H. pylori* OMPLA protein functions were explored through *in silico* analyzes. We found that there exist many multifunctional OMPs, which was confirmed by an in-depth study of bacterial porins. In addition to a pore function, OMPs may have variable loops likely involved in host evasion (see Paper I). Further knowledge regarding OMPLA functions was achieved by analyzing OMPLA sequences (Papers II and III). The conserved *H. pylori pldA* gene is preserved in the bacteria, where only a few sites are likely to be under positive selection pressure. Sequence analyzes were also used to find which genes, if any, are co-regulated with *pldA*. Our analyzes indicated that gastric *Helicobacter* species have a conserved *pldA* operon with two IM transporters that are not found elsewhere (Paper III). We generated a theory, supporting previously collected laboratory data with *in silico* studies, where *H. pylori* OMPLA multifunctionality could explain the observed highly conserved gene sequences. A protein structure model of *H. pylori* OMPLA was generated and we found that size and shape of the OMPLA enzyme structure could indicate a pore function. We hypothesize that urea could

diffuse into the cell through OMPLA (low pH), while ammonium exit through OMPLA (pH >6). Therefore, we have included OMPLA in a new model for *H. pylori* acid adaptation.

5.2 OMPs involved in niche adaptation

In order to quickly detect and adapt to a changing environment, the gating of molecules entering the cells is regulated through OMPs. Bacterial OMPs protect the OM and control membrane permeability. OMPs specific requirements may include substrate specificity and co-factor binding. These traits have been selected through evolution where the bacteria's ability to adapt is an essential survival mechanism [245]. This ability for genetic niche adaption to hostile environments poses an increased challenge for combating bacterial diseases [246].

There has been an increase in the number of solved OMP structures, yet the structures of many OMPs remain unsolved, and very few structures of OMP complexes have been determined [102]. Currently, only 0.1% of the protein structures deposited in the PDB belong to the OMP family. Nearly 40% of the solved OMP structures are porins, indicating their importance in the bacterial protein field. Many questions regarding membrane proteins still remain to be answered, particularly OMP gating [247].

In this thesis, a niche-adapted OMP review was used to better understand this protein superfamily. OmpA (see Introduction, and Class 1 porins in Paper I) [149] is a well-studied and multifunctional OMP porin with long flexible loops, and its expression depends upon different environmental signals. Another impressive example of niche-adapted OMPs is the siderophores that are resistant to lytic enzymes, *e.g.* peptidase, because they are composed of unusual amino acids (D-amino acids and ornithines) [248]. These unusual residues enable survival in the host. Another astonishing example is the multifunctional *Neisseria meningitidis* PorB; it is both a virulence factor and a solute transporter, and may contribute to antibiotic resistance. This multi-selective voltage-dependent channel has three separate translocation pathways for sugars, cations and anions. PorB targets host's mitochondria where it is required for pathogenesis [102, 249-253]. The L3 loop constricts the pore and is likely involved in gating of the different translocation pathways due to the exposed positively charged residues [102].

We found many examples where small changes in a protein yield drastic changes in virulence. For example, small loop changes in the Pla protein of *Yersina pestis* yield dramatic clinical effects ranging from the milder coughs symptoms to the plague. Pla is an omptin (see OM enzymes) and a human plasminogen activator, resulting in bacterial adherence and invasion

[254]. The Thr259Leu mutation in extracellular loop L5 results in an efficient plasminogen activator [255]. This is a mutation that goes from a polar to an aliphatic residue with great consequence for virulence. Pla is an example of a horizontally transferred gene where small changes had lethal consequences for humans [180].

The non-specific OmpG channel that is involved in the uptake of oligosaccharides has a pH-gated loop, L6, that can open and close the pore depending on environmental pH. Mutating one of the two His residues in this loop, removes the pH-sensitivity and opens the gate. Thus, changing a positive (histidine) to negatively (glutamate) charged loop residue (H231E) changes the pore function [256].

Protein stabilization is a key factor in how the proteins can tolerate and adapt to changes in the OM, which explains the solid barrel structure with flexible loops quickly capable of adapting to any environmental change [140]. Membrane proteins display a great tolerance and their capability of adopting new conformations enables the bacteria to thrive under harsh conditions and changing environments. This includes an intact OMPLA required for acidic survival (see Introduction, Papers I and III). OMP loops are needed to gate substrates through the channel [94, 98, 257, 258]. *In silico* studies indicate that these are conserved loop residues (often found inside the pore, *e.g.* loop L3 in porins), while the variable loop-residues are involved in host evasion (see Paper I) [94, 101]. The function of a possible *H. pylori* OMPLA loop and its role in acidic adaptation will be discussed later.

Sequence conservation yields information about protein evolution, and mapping this information onto a structure shows where these regions are, which in turn gives insights into why residues could be highly conserved or variable. Adding literature findings to *in silico* analyzes yield a powerful tool to analyze proteins. For example, evolutionary analyzes of the *ompF* gene sequences revealed that the evolutionary forces contribute differently within *Yersinia* species. These sequences display adaptive sites in the surface loops, reflecting niche adaptation [259]. OmpF facilitates the diffusion of small polar molecules through voltage and pH dependency [260].

The high loop variability observed in many of the porins analyzed (especially Class 1 and Class 4 porins) supports the high mutation rate needed for events such as host evasion (see Paper I). According to the literature, the exterior loops of the porins are continuously changing to avoid detection by the host immune system, phage invasion, and as a response to ecological pressure.

The exterior loops are the most variable regions in porins, which reflects the adaptive traits accomplished through mutation or recombination [140, 261, 262]. According to Stenkova *et al.*, porin structures reflect any previous interactions with environmental challenges. By frequently changing surface-exposed residues, the bacterium can go undetected by the host immune system.

During the analyzes of the porin structures and sequences, we found porin oligomerization likely for all eight porin classes (see Paper I). The idea of oligomerized porins have been discussed in the literature. Classic monomeric porins actually oligomerized in the right conditions, *e.g.* at lower temperatures or with less detergents than in the standard protocol [239, 263-269]. Thus, unconventional methods might be necessary to detect porin oligomerization. Our results support porin oligomerization since the conserved residues are most likely due to important protein-protein interactions. However, this conservation may not be observed in a normal environment since it could be a feature necessary for porins to stabilize in harsh conditions. Porins were divided into different groups based on specificity, size and oligomeric state. The many subdivisions existing in literature were considered, but the analysis resulted in eight porin classes based on pore size and function (see Table 3 in Paper I). The analyzes would reveal the conserved regions important for porins, *e.g.* during substrate translocation. The main strength and weakness of our analyzes lie in the number of sequences used in each porin class. Vast number of sequences in a class contribute to strengthen the analyzes, *e.g.* 1394 sequences used to construct Class 8 MSA. Likewise, the 50 sequences in Class 3 is less than optimal number of sequence. Thus, the main strength and weakness of our analyzes lies in the total number of sequences used in each porin class. The overall conclusion is more reliable since there were no contradicting results. All porin classes displayed conserved core residues, involved in substrate specificity, and conserved residues facing the lipid membrane that are likely involved in oligomerization.

H. pylori has a robust cell wall with a unique membrane composition [189, 270]. Cullen *et al.* discovered that *H. pylori* lipid A dephosphorylation, *e.g.* by LpxR, is required for mammalian colonization [271]. *H. pylori* can also steal and modify host cholesterol situated in lipid rafts that are cholesterol rich micro-domains [272]. The bacteria incorporate the glycosylated cholesterol into its own OM. This contributes to both pathogenicity and antibiotic resistance [273].

5.3 *H. pylori pldA* sequence analyzes

The *pldA* gene, encoding the OMPLA protein*,* is found in several Gram-negative species. The wide distribution of *pldA* implies a physiological role of importance [274]. Nevertheless, the sequence identity among the genes was quite low (*e.g.* the sequence identity between *E. coli* and *H. pylori pldA* is 46% on gene-level and 28% on protein-level). The sequence analyzes showed high inter-species variation (analyzing multiple species), but a very low intra-species variability for *H. pylori*.

The *H. pylori pldA* phase variation distinguishes between two variants: the $OMPLA_{ON}$-variant (intact OMPLA sequence) and the $OMPLA_{OFF}$-variant (truncated OMPLA sequence). The switch between these two variants is a genomic instability that is likely to occur spontaneously in nature, but our collected data (the published *pldA* sequences) contain mostly $OMPLA_{ON}$ sequences. The phase variation is caused by a homopolymeric tract which could be challenging to sequence (due to the number of equal bases in a row) [275-277]. The technical limitations of sequencing were of course kept in mind when sequencing and annotating the *H. pylori pld*A sequences at our laboratory (using Sanger sequencing), reducing the chance of errors. The sequencing technology has greatly improved in recent years; Sanger sequencing is more precise than *e.g.* whole-genome draft sequences from the last decades [277]. The $OMPLA_{ON}$-variant dominated in our entire data material (including the collected NCBI sequences), indicating that this is the most common variant. Even though some sequences could be erroneous, we strongly doubt this would affect the majority of our data. These findings revealed a very high sequence identity, which also support the significant role this protein must have in *H. pylori*.

Phylogenetic analyzes of the $OMPLA_{ON}$ sequences were compared to a reference set of *H. pylori* HK genes. Despite the high sequence variation observed in *H. pylori*, conserved core genes exist, including HK genes which are essential for survival [84, 278]. The conserved HK genes have been used to trace human migration through phylogenetic analyzes, demonstrating co-evolution between *H. pylori* and its host. *H. pylori* infection has been traced in humans before their migration from Africa through sequence analysis [85, 278]. Due to the many available analyzes platforms, the reference set was a control that enabled us to study geographical phylogeny in the *pldA* dataset. Nearly all of the *pldA* sequences clustered according to place of origin (with Asian, European and African clusters), as was observed for the reference (HK genes). The two *pldA* outliers observed in our Norwegian dataset turned out to be of African origin (samples from African patients taken at a Norwegian hospital). Although

HGT was detected by three different methods (codon bias, GC content, and phylogenetic conflict), the biogeography of the *pldA* sequences indicated that the transfer was ancient. It should be kept in mind that most sequences were of European origin, which could introduce a bias. However, *H. pylori* biogeographic clustering has been observed in the literature, including East Asian clustering at genome-level [85, 279].

The high nucleotide sequence identity observed among the *pldA* sequences were further analyzed for selection pressure. Purifying selection was detected at the clear majority of residues. This highly conserved gene in *H. pylori* is likely to have an evolutionarily stable function. However, some probable interaction sites were found to be under positive selection. The positively selected sites were mainly found in the surface-exposed loops or in the predicted signal sequence. Positive selection is often observed in hypervariable loops of OMPs likely to be involved in host evasion [280, 281].

The survival study of laboratory strains (OMPLA$_{ON}$ and OMPLA$_{OFF}$) in acidic environment (pH=3.5) resulted in different survival pattern compared to that observed at neutral pH [282]. Both variants survive neutral pH, however only OMPLA$_{ON}$-variants (with an intact OMPLA) survived in acidic environment. Tannæs *et al.* demonstrated that enzyme activity is not needed for survival in acidic environment, although the *pldA* gene is required [88].

Solving the structure of membrane proteins is difficult because of the hydrophobic membrane the proteins are embedded in (see Introduction). Since *E. coli* OMPLA structure was solved, the next step was to construct model. Looking at our data and the OMPLA structure, a pore-theory could explain the conservation of a protein not often used (as it would destabilize the membrane when active), but still dependent on it (expelling protons out) in highly acidic environment as is observed in the stomach. The sequence would be preserved and found in all strains since *H. pylori* might need it in certain harsh conditions.

5.4 *H. pylori* OMPLA model

A homology model structure of *H. pylori* OMPLA was constructed based on *E. coli* template (see Paper III). A multiple sequence alignment was made based on 660 sequences, which was used in EVA of OMPLA. These EVA yielded conserved residues situated at the putative trimer interface and in extracellular loop region. This indicates a conserved interface necessary for activity since the residues are conserved in our alignment. Calcium-binding site is found in the loop regions conserved, but our results indicated further interactions important for OMPLA.

Although loop residues form stabilizing interactions with LPS, these conserved loop residues are likely involved in protein-protein interactions [283]. The number of sequences used in the MSA was large enough and contained enough variability to construct a good alignment.

The sequence identity between *H. pylori* and *E. coli* OMPLA is low, but above the optimal threshold for homology modeling (when removing a gastric *H. pylori* specific insert discussed below), so the insert was not implemented in the model. Side-view picture of the generated *H. pylori* model structure is found in Figure 8. The low sequence identity between *E. coli* and *H. pylori* raises a question of why *H. pylori* has an enzyme that varies so much from the other species.

There are many unsolved questions regarding the insert observed in non-*Helicobacter* species (including the *E. coli* OMPLA template), yet some hypothesis regarding acid adaptation can be made due to laboratory results. Those *Helicobacter* species known to survive the harsh condition of the gastric mucosa cluster together and have a long insert that is lacking in other species, including enterohepatic *Helicobacters* (see Table 2 in Paper III). Literature has shown that mutating one charged extracellular loop residue can have deleterious effect on acid survival in bacteria and humans [284-289]. There is a possibility that this stretch of residues constitute an extra β-strand and an extracellular loop not found in the template. This would result in a larger pore; however, this must be determined by analyzing a solved *H. pylori* OMPLA structure.

The conserved cysteines found in this insert, predicted to be an extracellular loop, are likely stabilizing the structure through disulfide binding. This could be crucial for acid survival since it is found in all gastric *Helicobacters* (see Paper III), perhaps a pH-gated loop like the OmpG porin [237]. Missense mutations important for acidic adaptations were mapped onto our model in order to look for possible explanations. One of these mutations was found in the predicted loop insert. Therefore, we created the hypothesis that this insert could be niche-acquired. The *H. pylori* strain euBZ (with P157S mutation) was modeled to be in the active site and putative trimeric interface indicating that either enzyme activity or trimerization is of importance for acid survival.

Harboring phase variable genes might seem energetically unfavorable due to the cost of transcribing, translating and degrading truncated transcripts. There might be a functional role for OMPLA$_{OFF}$ transcripts (see OMPLA$_{OFF}$ model in Figure 10). However, it is more likely that

phase variation is an important mechanism where OMPLA is only fully transcribed when needed. Or more precisely, those strains containing an intact OMPLA will survive in different environments than those expressing a truncated OMPLA. Further studies on this subject are needed, but this theory does seem probable considering the extreme variation in habitats where *H. pylori* may be found. For example, it would not make sense for an OMPLA pore active at pH < 3 in an environment lacking urea, and perhaps *H. pylori* living in neutral environments would benefit from the complete absence of OMPLA (OMPLA$_{OFF}$)

The estimated *H. pylori* OMPLA pore size of 4 Å could be larger since these calculations estimate the size of spherical probes. The estimated OMPLA pore size is larger than the Class 1 porins, but smaller than the estimated pore size for Class 2 porins. OMPLA structure and model have elliptic shaped pores that could allow larger substrates than estimated. Although, *E. coli* OMPLA structure and *H. pylori* model have estimated pores in both WHAT IF and PoreWalker softwares, they do vary in shape and amino acid composition. In WHATIF, all OMPLA residues were mutated into Ala to estimate the maximum β-barrel size so that the pore size that is not restricted by flexible side chains. This might explain variation in the estimated pore size.

5.5 *pldA* operon

Genes are often co-expressed in bacteria to enable quick changes necessary for survival in a changing environment. ProOpDB can compare a limited number of species, but is useful as an initial comparison while we analyzed genes neighboring *pldA* and examined possible operons, using FgenesB. *H. pylori* was compared to 20 different species. There is a difference between gastric and enterohepatic *Helicobacters*. Most of the gastric species seem to co-regulate *pldA* with two inner membrane transporters belonging to COG0377 protein family. These are *H. pylori, H. acinonychis, H. bizzozeronii, H. cetorum, H. heilmannii, H. suis,* and *H. felis.* The enterohepatic species, and two gastric species (*H. mustelae* and *H. himalayensis*) lack this operon or have different genes clustered together. The two outliers either lack urease or *pldA* genes, indicating they have evolved a different survival mechanism (which is true for *H. mustelae* that has a second unique urease).

5.6 AmCI and AmCII

The gene regulation with the three genes clustering together seem unique to gastric species, and thus likely relevant for acid protection. AmCI and AmCII belong to the SLC6-family of channel proteins that transport small substances [290, 291], e.g. glycine, serotonin, dopamine, and norepinephrine [292]. They all have 5+5 helix motif in common, although there are some structures in this protein family that contain additional helices [290]. Based on the helical motif of AmCI and AmCII, they are predicted to be located in the IM since helical proteins seldom are found in the bacterial OM [293].

These channels may have a variety of different functions, however, since they're likely co-regulated with OMPLA in gastric *Helicobacters*, we hypothesize they are involved in acid tolerance. Since there are currently no ammonium/ammonia channels, we believe the two COG0733 are likely candidates. The differences observed between the AmCI and AmCII channels, could be explained by different substrate specificities. AmCI seem to have a more polar pore upon visual inspection, and could be a NH_4^+ channel. Visual inspection shows a more hydrophobic AmCII pore which indicate a core for NH_3 efflux.

6. Conclusions

The common feature of all OMPs is that they are specialized to overcome environmental challenges that bacteria may encounter. These robust transmembrane proteins are composed of a common β-strand motif, but their function varies greatly. The evolutionary patterns of porins reveal interactions that should be further explored. We believe *H. pylori* OMPLA has acquired an additional feature to survive the harsh condition found in the gastric mucosa.

Literature has shown that *H. pylori* has a more robust OM than other Gram-negative bacteria since it is less leaky. Although we know that the periplasmic pH is lower than the exterior pH, there are not enough findings explaining how the bacteria can maintain this level and buffer the acidity. An OMPLA with a pore activity expelling protons is illustrated in Figure 13 (perhaps stabilized as a trimer). Further studies are needed to understand the impact of the insert that seem to be acidic-specific. The presence of this non-modeled insert raises many questions, including why it is located in the phase variable region.

In conclusion, our analyzes indicated that OMPLA may be a niche-adapted protein. Although it once was likely horizontally acquired, the *pldA* sequences are highly conserved. Sequence analyzes indicate positive selection occurring at sites needed for possible pathogenic interactions. Current literature does not explain how the OM is involved in maintaining higher periplasmic pH level compared to the acidic outside environment. Urea could pass through OMPLA when the environment is acidic, while ammonia (NH_3/NH_4^+) could exit when periplasmic pH is high, as shown in Figure 13.
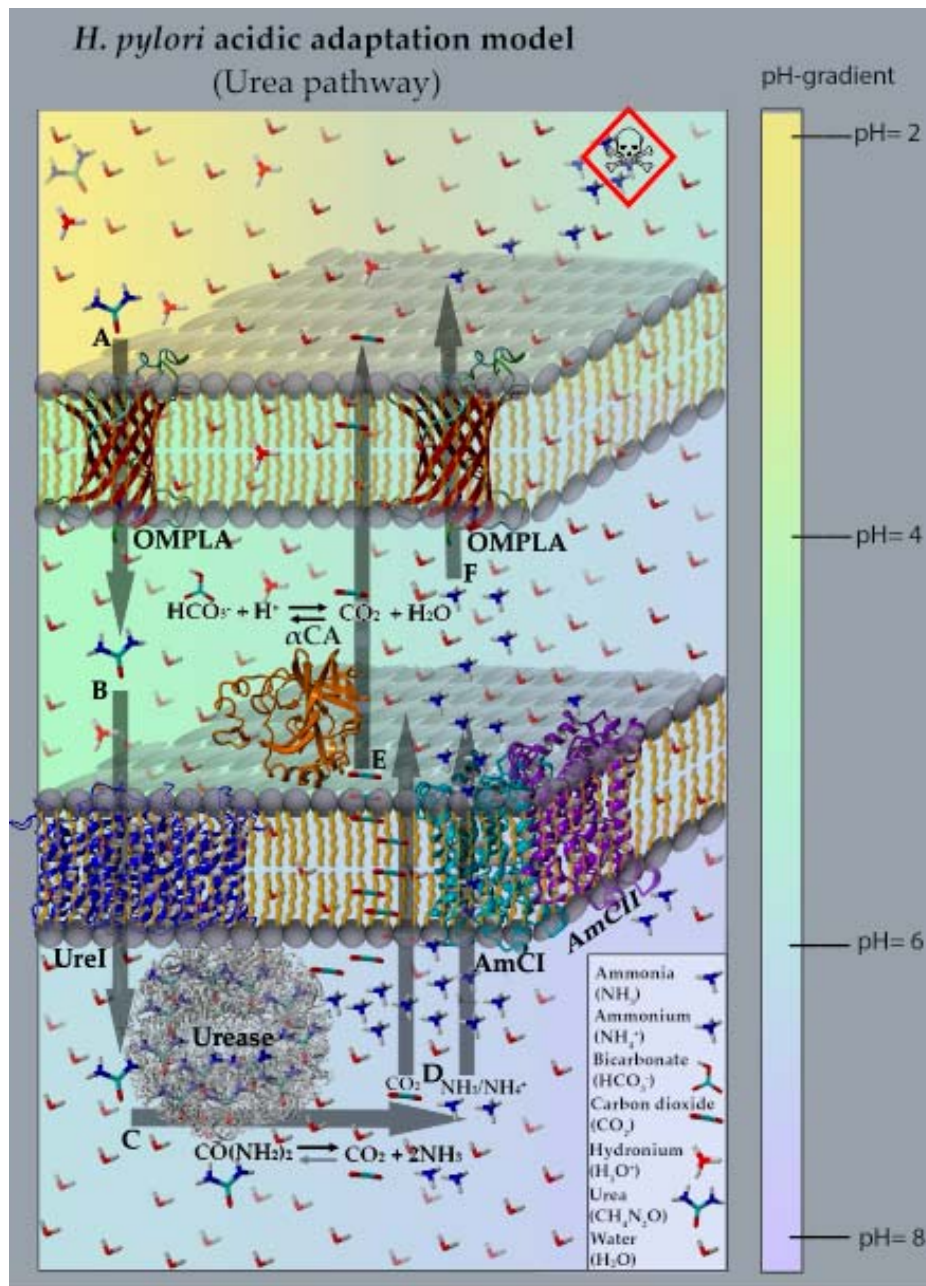

Figure 13: H. pylori acid tolerance model. *This illustration shows possible H. pylori* OMPLA *function required for acid tolerance (copied from Paper III).*

7. Future perspectives

The overall goal with this project was to develop a model explaining how *H. pylori* OMPLA has been inferred in survival in the gastric mucosa. In order to confirm OMPLA's roles in acid tolerance, pore-activity needs to be confirmed through *in vitro* and *in vivo* experiments.

Molecular dynamic simulations would allow better estimates of substrates for these channels. However, it would be quite difficult to show substrate passing through at different pH-levels. Thus, solving the 3D structure of OMPLA, AmC1 & AmCII would allow better structure-function assessments that could aid further experimental design and characterizations of these proteins which we believe participate in the urea-pathway. Furthermore, model organisms could be useful in experiments to better understand pore-activity. Xenopus oocytes (frog eggs) have previously been used to characterize other protein channels [294, 295].

Promoter and transcription factor analyzes combined with experimental phase variation analyzes could identify when the *pldA* operon is activated and when the genes are in active or truncated form. It would be of great interest to study which niche the truncated OMPLA$_{OFF}$ preferred. Our data imply that this is not in the gastric mucosa, and we hypothesize that OMPLA is truncated during transmission.

This thesis shed light on possible roles of *H. pylori* OMPLA, but the field is still in its early stages on the road to understanding *H. pylori* survival in the acidic stomach. This creates an opportunity for many exciting new discoveries in the future.

8. References

1. Zimmer C, Emlen DJ: Evolution: making sense of life. Roberts and Company Publishers; 2013.
2. Moran L: What is evoltuion? [http://www.talkorigins.org/faqs/evolution-definition.html] Accessed Date: December 2016 2016.
3. Westminster-Abbey: Our history: Charles Darwin [http://www.westminster-abbey.org/our-history/people/charles-darwin] Accessed Date: 24th September 2015.
4. van Wyhe J: Darwin's field notes on the Galapagos: 'a little world within itself' [http://darwin-online.org.uk/EditorialIntroductions/Chancellor_Keynes_Galapagos.html] Accessed Date: August 2016.
5. Darwin CR: The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 6 edn. London; 1876.
6. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. Am Biol Teach. 1973; 35:125-129.
7. Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. J Clin Invest. 2009; 119:2475-2487.
8. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2009; 37:D26-31.
9. Bahadori A, Esmaeillu M, Sadighbayan KH, Bahadori F, Jabbari MV, Attarhosseani M, Mobaiyen H, Hossein Z. Evaluation of non *H. pylori* spiral organisms in human gastric biopsies by using PCR and microscop methods in Iran (first report). ESJ. 2013; 9:506-516.
10. Kersulyte D, Rossi M, Berg DE. Sequence divergence and conservation in genomes of *Helicobacter cetorum* strains from a dolphin and a whale. PLoS One. 2013; 8:e83177.
11. Nagamine CM, Shen Z, Luong RH, McKeon GP, Ruby NF, Fox JG. Co-infection of the Siberian hamster (*Phodopus sungorus*) with a novel *Helicobacter* sp. and *Campylobacter* sp. J Med Microbiol. 2015; 64:575-581.
12. Menard A, Pere-Vedrenne C, Haesebrouck F, Flahou B. Gastric and enterohepatic *Helicobacters* other than *Helicobacter pylori*. Helicobacter. 2014; 19 Suppl 1:59-67.
13. Dailidiene D, Dailide G, Ogura K, Zhang M, Mukhopadhyay AK, Eaton KA, Cattoli G, Kusters JG, Berg DE. *Helicobacter acinonychis:* genetic and rodent infection studies of a *Helicobacter pylori*-like gastric pathogen of cheetahs and other big cats. J Bacteriol. 2004; 186:356-365.
14. Baele M, Decostere A, Vandamme P, Ceelen L, Hellemans A, Mast J, Chiers K, Ducatelle R, Haesebrouck F. Isolation and characterization of *Helicobacter suis* sp. nov. from pig stomachs. Int J Syst Evol Microbiol. 2008; 58:1350-1358.
15. Thomson JM, Hansen R, Berry SH, Hope ME, Murray GI, Mukhopadhya I, McLean MH, Shen Z, Fox JG, El-Omar E, Hold GL. Enterohepatic *Helicobacter* in ulcerative colitis: potential pathogenic entities? PLoS One. 2011; 6:e17184.
16. Schauer DB: Enterohepatic *Helicobacter* species. In *Helicobacter pylori*: Physiology and Genetics. Mobley HLT, Mendz GL, Hazell SL, editors. Washington (DC): ASM Press; 2001: 43.
17. Mikkonen TP, Karenlampi RI, Hanninen ML. Phylogenetic analysis of gastric and enterohepatic *Helicobacter species* based on partial HSP60 gene sequences. Int J Syst Evol Microbiol. 2004; 54:753-758.
18. Nobelprize.org: Press release: the 2005 Nobel Prize in physiology or medicine to Barry J. Marshall and J. Robin Warren Accessed Date: 15 Sep 2015 2015.

19. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, et al. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterology. 2017.

20. Bui D, Brown HE, Harris RB, Oren E. Serologic evidence for fecal-oral transmission of *Helicobacter pylori*. Am J Trop Med Hyg. 2016; 94:82-88.

21. Parsonnet J, Shmuely H, Haggerty T. Fecal and oral shedding of *Helicobacter pylori* from healthy infected adults. Jama. 1999; 282:2240-2245.

22. Aziz RK, Khalifa MM, Sharaf RR. Contaminated water as a source of *Helicobacter pylori* infection: a review. J Adv Res. 2015; 6:539-547.

23. Krajden S, Fuksa M, Anderson J, Kempston J, Boccia A, Petrea C, Babida C, Karmali M, Penner JL. Examination of human stomach biopsies, saliva, and dental plaque for *Campylobacter pylori*. J Clin Microbiol. 1989; 27:1397-1398.

24. Winiecka-Krusnell J, Wreiber K, von Euler A, Engstrand L, Linder E. Free-living amoebae promote growth and survival of *Helicobacter pylori*. Scand J Infect Dis. 2002; 34:253-256.

25. Siavoshi F, Saniee P. Vacuoles of *Candida* yeast as a specialized niche for *Helicobacter pylori*. World J Gastroenterol. 2014; 20:5263-5273.

26. Kusters JG, van Vliet AHM, Kuipers EJ. Pathogenesis of *Helicobacter pylori* Infection. Clin Microbiol Rev. 2006; 19:449-490.

27. Anand PS, Kamath KP, Anil S. Role of dental plaque, saliva and periodontal disease in *Helicobacter pylori* infection. World J Gastroenterol. 2014; 20:5639-5653.

28. Nardone G, Compare D. The human gastric microbiota: is it time to rethink the pathogenesis of stomach diseases? United European Gastroenterol J. 2015; 3:255-260.

29. Nasrollahzadeh D, Malekzadeh R, Ploner A, Shakeri R, Sotoudeh M, Fahimi S, Nasseri-Moghaddam S, Kamangar F, Abnet CC, Winckler B, et al. Variations of gastric corpus microbiota are associated with early esophageal squamous cell carcinoma and squamous dysplasia. Sci Rep. 2015; 5:8820.

30. Eun CS, Kim BK, Han DS, Kim SY, Kim KM, Choi BY, Song KS, Kim YS, Kim JF. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. Helicobacter. 2014; 19:407-416.

31. Salama NR, Hartung ML, Muller A. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. Nat Rev Microbiol. 2013; 11:385-399.

32. Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, Ewoldt RH, McKinley GH, So P, Erramilli S, Bansil R. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. Proc Natl Acad Sci U S A. 2009; 106:14321-14326.

33. van der Veen S, Tang CM. The BER necessities: the repair of DNA damage in human-adapted bacterial pathogens. Nat Rev Microbiol. 2015; 13:83-94.

34. Lund P, Tramonti A, De Biase D. Coping with low pH: molecular strategies in neutralophilic bacteria. FEMS Microbiol Rev. 2014; 38:1091-1125.

35. Pflock M, Finsterer N, Joseph B, Mollenkopf H, Meyer TF, Beier D. Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. J Bacteriol. 2006; 188:3449-3462.

36. Huang JY, Sweeney EG, Sigal M, Zhang HC, Remington SJ, Cantrell MA, Kuo CJ, Guillemin K, Amieva MR. Chemodetection and destruction of host urea allows *Helicobacter pylori* to locate the epithelium. Cell Host Microbe. 2015; 18:147-156.

37. McNulty R, Ulmschneider JP, Luecke H, Ulmschneider MB. Mechanisms of molecular transport through the urea channel of *Helicobacter pylori*. Nat Commun. 2013; 4:2900.

38.   Krulwich TA, Sachs G, Padan E. Molecular aspects of bacterial pH sensing and homeostasis. Nat Rev Micro. 2011; 9:330-343.

39.   Sachs G, Weeks DL, Wen Y, Marcus EA, Scott DR, Melchers K. Acid acclimation by *Helicobacter pylori*. Physiology (Bethesda). 2005; 20:429-438.

40.   Luecke H. Structure, function, and inhibitors of the acid-gated *Helicobacter pylori* urea channel, an essential component for acid survival and chronic infection. FASEB J. 2015; 29:239.231.

41.   Posselt G, Backert S, Wessler S. The functional interplay of *Helicobacter pylori* factors with gastric epithelial cells induces a multi-step process in pathogenesis. Cell Commun Signal. 2013; 11:77.

42.   Eaton KA, Gilbert JV, Joyce EA, Wanken AE, Thevenot T, Baker P, Plaut A, Wright A. In vivo complementation of ureB restores the ability of *Helicobacter pylori* to colonize. Infect Immun. 2002; 70:771-778.

43.   Mine T, Muraoka H, Saika T, Kobayashi I. Characteristics of a clinical isolate of urease-negative *Helicobacter pylori* and its ability to induce gastric ulcers in Mongolian gerbils. Helicobacter. 2005; 10:125-131.

44.   Liechti G, Goldberg JB. Outer membrane biogenesis in *Escherichia coli*, *Neisseria meningitidis*, and *Helicobacter pylori*: paradigm deviations in *H. pylori*. Front Cell Infect Microbiol. 2012; 2:29.

45.   Yamaoka Y. Mechanisms of disease: *Helicobacter pylori* virulence factors. Nat Rev Gastroenterol Hepatol. 2010; 7:629-641.

46.   Roesler BM, Rabelo-Gonçalves EMA, Zeitune JMR. Virulence factors of *Helicobacter pylori*: a review. Clin Med Insights Gastroenterol. 2014; 7:9-17.

47.   Matsuo Y, Kido Y, Yamaoka Y. *Helicobacter pylori* outer membrane protein-related pathogenesis. Toxins (Basel). 2017; 9.

48.   Su YL, Huang HL, Huang BS, Chen PC, Chen CS, Wang HL, Lin PH, Chieh MS, Wu JJ, Yang JC, Chow LP. Combination of OipA, BabA, and SabA as candidate biomarkers for predicting *Helicobacter pylori*-related gastric cancer. Sci Rep. 2016; 6:36442.

49.   Sallas ML, Melchiades JL, Zabaglia LM, do Prado Moreno JR, Orcini WA, Chen ES, Smith MdAC, Payão SLM, Rasmussen LT. Prevalence of *Helicobacter pylori* vacA, cagA, dupA and oipA genotypes in patients with gastric disease. 2017.

50.   Delgado-Rosado G, Dominguez-Bello MG, Massey SE. Positive selection on a bacterial oncoprotein associated with gastric cancer. Gut Pathog. 2011; 3:18.

51.   Furuta Y, Yahara K, Hatakeyama M, Kobayashi I. Evolution of *cagA* oncogene of *Helicobacter pylori* through recombination. PLoS One. 2011; 6:e23499.

52.   Ta LH, Hansen LM, Sause WE, Shiva O, Millstein A, Ottemann KM, Castillo AR, Solnick JV. Conserved transcriptional unit organization of the Cag Pathogenicity Island among *Helicobacter pylori* strains. Front Cell Infect Microbiol. 2012; 2.

53.   Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, Haas R. Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol. 2001; 42:1337-1348.

54.   Yamaoka Y, Kato M, Asaka M. Geographic differences in gastric cancer incidence can be explained by differences between *Helicobacter pylori* strains. Intern Med. 2008; 47:1077-1083.

55.   Ling SS, Khoo LH, Hwang LA, Yeoh KG, Ho B. Instrumental role of *Helicobacter pylori* γ-glutamine transpeptidase in VacA-dependent vacuolation in gastric epithelial cells. PLoS One. 2015; 10:e0131460.

56. Ling SSM, Yeoh KG, Ho B. Helicobacter pylori γ-glutamyl transpeptidase: a formidable virulence factor. World J Gastroenterol. 2013; 19:8203-8210.

57. McGovern KJ, Blanchard TG, Gutierrez JA, Czinn SJ, Krakowka S, Youngman P. γ-glutamyltransferase is a Helicobacter pylori virulence factor but is not essential for colonization. Infect Immun. 2001; 69:4168-4173.

58. Wang M-y, Chen C, Shao C, Wang S-b, Wang A-c, Yang Y-c, Yuan X-y, Shao S-h. Intact long-type DupA protein in Helicobacter pylori is an ATPase involved in multifunctional biological activities. Microbial pathogenesis. 2015; 81:53-59.

59. Yamaoka Y, Ojo O, Fujimoto S, Odenbreit S, Haas R, Gutierrez O, El-Zimaity HM, Reddy R, Arnqvist A, Graham DY. *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. Gut. 2006; 55:775-781.

60. Schmidt TP, Perna AM, Fugmann T, Bohm M, Jan H, Haller S, Gotz C, Tegtmeyer N, Hoy B, Rau TT, et al. Identification of E-cadherin signature motifs functioning as cleavage sites for *Helicobacter pylori* HtrA. Sci Rep. 2016; 6:23264.

61. Tegtmeyer N, Moodley Y, Yamaoka Y, Pernitzsch SR, Schmidt V, Traverso FR, Schmidt TP, Rad R, Yeoh KG, Bow H, et al. Characterisation of worldwide *Helicobacter pylori* strains reveals genetic conservation and essentiality of serine protease HtrA. Mol Microbiol. 2016; 99:925-944.

62. Axon A, Forman D. *Helicobacter* gastroduodenitis: a serious infectious disease. Bmj. 1997; 314:1430-1431.

63. Bergman M, Del Prete G, van Kooyk Y, Appelmelk B. *Helicobacter pylori* phase variation, immune modulation and gastric autoimmunity. Nat Rev Microbiol. 2006; 4:151-159.

64. Graham DY. History of *Helicobacter pylori*, duodenal ulcer, gastric ulcer and gastric cancer. World J Gastroenterol. 2014; 20:5191-5204.

65. IARC Working Group: *Helicobacter pylori* eradication as a strategy for preventing gastric cancer. vol. 8. France: "International Agency for Research on Cancer"; 2014.

66. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to *Helicobacter pylori*. Int J Cancer. 2015; 136:487-490.

67. Rupnow MF, Chang AH, Shachter RD, Owens DK, Parsonnet J. Cost-effectiveness of a potential prophylactic *Helicobacter pylori* vaccine in the United States. J Infect Dis. 2009; 200:1311-1317.

68. Aebischer T, Bumann D, Epple HJ, Metzger W, Schneider T, Cherepnev G, Walduck AK, Kunkel D, Moos V, Loddenkemper C, et al. Correlation of T cell response and bacterial clearance in human volunteers challenged with Helicobacter pylori revealed by randomised controlled vaccination with Ty21a-based Salmonella vaccines. Gut. 2008; 57:1065-1072.

69. Venerito M, Goni E, Malfertheiner P. *Helicobacter pylori* screening: options and challenges. Expert Rev Gastroenterol Hepatol. 2016; 10:497-503.

70. Malfertheiner P, Megraud F, O'Morain CA, Gisbert JP, Kuipers EJ, Axon AT, Bazzoli F, Gasbarrini A, Atherton J, Graham DY, et al. Management of *Helicobacter pylori* infection: the Maastricht v/Florence consensus report. Gut. 2017; 66:6-30.

71. Zeng M, Mao XH, Li JX, Tong WD, Wang B, Zhang YJ, Guo G, Zhao ZJ, Li L, Wu DL, et al. Efficacy, safety, and immunogenicity of an oral recombinant *Helicobacter pylori* vaccine in children in China: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2015; 386:1457-1464.

72. Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG clinical guideline: treatment of *Helicobacter pylori* infection. Am J Gastroenterol. 2017; 112:212-239.

73. Graham DY. *Helicobacter pylori* update: gastric cancer, reliable therapy, and possible benefits. Gastroenterology. 2015; 148:719-731.e713.

74. WHO Media centre: WHO publishes list of bacteria for which new antibiotics are urgently needed [http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/] Accessed Date: 27/02 2017.

75. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014; 42:D1091-1097.

76. Darmon E, Leach DR. Bacterial genome instability. Microbiol Mol Biol Rev. 2014; 78:1-39.

77. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev. 2011; 35:957-976.

78. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, Suerbaum S, Achtman M, Linz B. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. PLoS Genet. 2010; 6:e1001069.

79. Blaser MJ. The versatility of *Helicobacter pylori* in the adaptation to the human stomach. J Physiol Pharmacol. 1997; 48:307-314.

80. Nedenskov-Sorensen P, Bukholm G, Bovre K. Natural competence for genetic transformation in *Campylobacter pylori*. J Infect Dis. 1990; 161:365-366.

81. Hanada K, Yamaoka Y. Genetic battle between *Helicobacter pylori* and humans: the mechanism underlying homologous recombination in bacteria, which can infect human cells. Microbes Infect. 2014; 16:833-839.

82. Bjorkholm B, Sjolund M, Falk PG, Berg OG, Engstrand L, Andersson DI. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. Proc Natl Acad Sci U S A. 2001; 98:14607-14612.

83. Krebes J, Morgan RD, Bunk B, Sproer C, Luong K, Parusel R, Anton BP, Konig C, Josenhans C, Overmann J, et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res. 2014; 42:2415-2432.

84. McClain MS, Shaffer CL, Israel DA, Peek RM, Jr., Cover TL. Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. BMC Genomics. 2009; 10:3.

85. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007; 445:915-918.

86. Janssen BD, Hayes CS. The tmRNA ribosome rescue system. Adv Protein Chem Struct Biol. 2012; 86:151-191.

87. Hui MP, Foley PL, Belasco JG. Messenger RNA degradation in bacterial cells. Annu Rev Genet. 2014; 48:537-559.

88. Tannæs T, Dekker N, Bukholm G, Bijlsma JJE, Appelmelk BJ. Phase variation in the *Helicobacter pylori* phospholipase A gene and Its role in acid adaptation. Infect Immun. 2001; 69:7334-7340.

89. Gram HC. Über die isolierte Färbung der *Schizomyceten* in Schnitt- und Trockenpräparaten (English translation: "The differential staining of *Scbizomycetes* in tissue sections" by Brock T.D., ASM Press. 1998, p210). Fortschritte der Medizin. 1884; 2:185-189.

90. Gupta RS. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. Antonie Van Leeuwenhoek. 2011; 100:171-182.

91. Islam ST, Lam JS. Topological mapping methods for α-helical bacterial membrane proteins: an update and a guide. Microbiologyopen. 2013; 2:350-364.

92. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect Biol. 2010; 2:a000414.

93. Dramsi S, Magnet S, Davison S, Arthur M. Covalent attachment of proteins to peptidoglycan. FEMS Microbiol Rev. 2008; 32:307-320.

94. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev. 2003; 67:593-656.

95. Lin J, Huang S, Zhang Q. Outer membrane proteins: key players for bacterial adaptation in host niches. Microbes Infect. 2002; 4:325-331.

96. Yen MR, Peabody CR, Partovi SM, Zhai Y, Tseng YH, Saier MH. Protein-translocating outer membrane porins of Gram-negative bacteria. Biochim Biophys Acta. 2002; 1562:6-31.

97. Masi M, Pagès JM. Structure, function and regulation of outer membrane proteins involved in drug transport in *Enterobactericeae*: the OmpF/C – TolC case. Open Microbiol J. 2013; 7:22-33.

98. Eren E, Vijayaraghavan J, Liu J, Cheneke BR, Touw DS, Lepore BW, Indic M, Movileanu L, van den Berg B. Substrate specificity within a family of outer membrane carboxylate channels. PLoS Biol. 2012; 10:e1001242.

99. Wimley WC. The versatile β-barrel membrane protein. Curr Opin Struct Biol. 2003; 13:404-411.

100. Bishop RE. Structural biology of membrane-intrinsic β-barrel enzymes: sentinels of the bacterial outer membrane. Biochim Biophys Acta. 2008; 1778:1881-1896.

101. Koebnik R, Locher KP, Van Gelder P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol. 2000; 37:239-253.

102. Fairman JW, Noinaj N, Buchanan SK. The structural biology of β-barrel membrane proteins: a summary of recent reports. Curr Opin Struct Biol. 2011; 21:523-531.

103. Schulz GE. β-barrel membrane proteins. Curr Opin Struct Biol. 2000; 10:443-447.

104. van den Berg B. Lateral gates: β-barrels get in on the act. Nat Struct Mol Biol. 2013; 20:1237-1239.

105. Galdiero S, Galdiero M, Pedone C. β-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. Curr Protein Pept Sci. 2007; 8:63-82.

106. Basle A, Rummel G, Storici P, Rosenbusch JP, Schirmer T. Crystal structure of osmoporin OmpC from *E. coli* at 2.0 Å. J Mol Biol. 2006; 362:933-942.

107. Gromiha MM, Ahmad S, Suwa M. Application of residue distribution along the sequence for discriminating outer membrane proteins. Comput Biol Chem. 2005; 29:135-142.

108. Gromiha MM, Suwa M. Current developments on β-barrel membrane proteins: sequence and structure analysis, discrimination and prediction. Curr Protein Pept Sci. 2007; 8:580-599.

109. Martin J, de Brevern AG, Camproux AC. *In silico* local structure approach: a case study on outer membrane proteins. Proteins. 2008; 71:92-109.

110. Wirth C, Condemine G, Boiteux C, Berneche S, Schirmer T, Peneff CM. NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM porin family. J Mol Biol. 2009; 394:718-731.

111. Sandrini S, Masania R, Zia F, Haigh R, Freestone P. Role of porin proteins in acquisition of transferrin iron by enteropathogens. Microbiology. 2013; 159:2639-2650.

112. Kleinschmidt JH: Folding and stability of monomeric β-Barrel membrane proteins, in protein-lipid Interactions: from membrane domains to cellular networks In Protein–Lipid Interactions: From Membrane Domains to Cellular Networks. Tamm LK, editors. Weinheim, FRG. : Wiley-VCH Verlag GmbH & Co. KGaA; 2006: 27-282.

113.   Schirmer T. General and specific porins from bacterial outer membranes. J Struct Biol. 1998; 121:101-109.
114.   Faraldo-Gómez JD, Sansom MS. Acquisition of siderophores in Gram-negative bacteria. Nat Rev Mol Cell Biol. 2003; 4:105-116.
115.   Bhardwaj AK, Mohanty P. Bacterial efflux pumps involved in multidrug resistance and their Inhibitors: rejuvinating the antimicrobial chemotherapy. Recent Pat Antiinfect Drug Discov. 2012; 7:73-89.
116.   Li XZ, Nikaido H. Efflux-mediated drug resistance in bacteria: an update. Drugs. 2009; 69:1555-1623.
117.   Blair JM, Piddock LJ. Structure, function and inhibition of RND efflux pumps in Gram-negative bacteria: an update. Curr Opin Microbiol. 2009; 12:512-519.
118.   Poole K, Krebes K, McNally C, Neshat S. Multiple antibiotic resistance in *Pseudomonas aeruginosa*: evidence for involvement of an efflux operon. J Bacteriol. 1993; 175:7363-7372.
119.   Schulz GE. The structure of bacterial outer membrane proteins. Biochim Biophys Acta. 2002; 1565:308-317.
120.   Loveless BJ, Saier MH, Jr. A novel family of channel-forming, autotransporting, bacterial virulence factors. Mol Membr Biol. 1997; 14:113-123.
121.   Ye J, van den Berg B. Crystal structure of the bacterial nucleoside transporter Tsx. EMBO J. 2004; 23:3187-3195.
122.   Hearn EM, Patel DR, Lepore BW, Indic M, van den Berg B. Transmembrane passage of hydrophobic compounds through a protein channel wall. Nature. 2009; 458:367-370.
123.   Jacob-Dubuisson F, Villeret V, Clantin B, Delattre AS, Saint N. First structural insights into the TpsB/Omp85 superfamily. Biol Chem. 2009; 390:675-684.
124.   Navarro-Garcia F, Elias WP. Autotransporters and virulence of enteroaggregative *E. coli*. Gut Microbes. 2011; 2:13-24.
125.   Lyskowski A, Leo JC, Goldman A. Structure and biology of trimeric autotransporter adhesins. Adv Exp Med Biol. 2011; 715:143-158.
126.   Scarselli M, Serruto D, Montanari P, Capecchi B, Adu-Bobie J, Veggi D, Rappuoli R, Pizza M, Arico B. *Neisseria meningitidis* NhhA is a multifunctional trimeric autotransporter adhesin. Mol Microbiol. 2006; 61:631-644.
127.   van Ulsen P, van Alphen L, ten Hove J, Fransen F, van der Ley P, Tommassen J. A *Neisserial* autotransporter NalP modulating the processing of other autotransporters. Mol Microbiol. 2003; 50:1017-1030.
128.   Qin W, Wang L, Lei L. New findings on the function and potential applications of the trimeric autotransporter adhesin. Antonie Van Leeuwenhoek. 2015; 108:1-14.
129.   Dautin N, Bernstein HD: Protein secretion in Gram-negative bacteria via the autotransporter pathway. In Annu Rev Microbiol. Volume 61; 2007: 89-112: Annual Review of Microbiology].
130.   Sheldon JR, Laakso HA, Heinrichs DE. Iron acquisition strategies of bacterial pathogens. Microbiology spectrum. 2016; 4.
131.   Volkan E, Kalas V, Pinkner JS, Dodson KW, Henderson NS, Pham T, Waksman G, Delcour AH, Thanassi DG, Hultgren SJ. Molecular basis of usher pore gating in *Escherichia coli* pilus biogenesis. Proc Natl Acad Sci U S A. 2013; 110:20741-20746.
132.   Lillington J, Waksman G. Ordered and ushered; the assembly and translocation of the adhesive type I and P pili. Biology. 2013; 2:841-860.
133.   Palomino C, Marin E, Fernandez LA. The fimbrial usher FimD follows the SurA-BamB pathway for its assembly in the outer membrane of *Escherichia coli*. J Bacteriol. 2011; 193:5222-5230.

134. Fairman JW, Dautin N, Wojtowicz D, Liu W, Noinaj N, Barnard TJ, Udho E, Przytycka TM, Cherezov V, Buchanan SK. Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. Structure. 2012; 20:1233-1243.

135. Tsai JC, Yen MR, Castillo R, Leyton DL, Henderson IR, Saier MH, Jr. The bacterial intimins and invasins: a large and novel family of secreted proteins. PLoS One. 2010; 5:e14403.

136. McClean S. Eight stranded β-barrel and related outer membrane proteins: role in bacterial pathogenesis. Protein Pept Lett. 2012; 19:1013-1025.

137. Whitelegge J. Gas-phase structure of the *E. coli* OmpA dimer. Structure. 2014; 22:666-667.

138. Zakharov SD, Eroukova VY, Rokitskaya TI, Zhalnina MV, Sharma O, Loll PJ, Zgurskaya HI, Antonenko YN, Cramer WA. Colicin occlusion of OmpF and TolC channels: outer membrane translocons for colicin import. Biophys J. 2004; 87:3901-3911.

139. Housden NG, Wojdyla JA, Korczynska J, Grishkovskaya I, Kirkpatrick N, Brzozowski AM, Kleanthous C. Directed epitope delivery across the *Escherichia coli* outer membrane through the porin OmpF. Proc Natl Acad Sci U S A. 2010; 107:21412-21417.

140. Achouak W, Heulin T, Pages JM. Multiple facets of bacterial porins. FEMS Microbiol Lett. 2001; 199:1-7.

141. Pages JM, James CE, Winterhalter M. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. Nat Rev Microbiol. 2008; 6:893-903.

142. Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I, Troupiotis-Tsailaki A, Jonckheere W, Pehau-Arnaudet G, Pinkner JS, Chapman MR, et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature. 2014; 516:250-253.

143. Zahn M, D'Agostino T, Eren E, Basle A, Ceccarelli M, van den Berg B. Small-molecule transport by CarO, an abundant eight-stranded β-barrel outer membrane protein from *Acinetobacter baumannii*. J Mol Biol. 2015; 427:2329-2339.

144. Galdiero S, Falanga A, Cantisani M, Tarallo R, Della Pepa ME, D'Oriano V, Galdiero M. Microbe-host interactions: structure and role of Gram-negative bacterial porins. Curr Protein Pept Sci. 2012; 13:843-854.

145. Ziervogel BK, Roux B. The binding of antibiotics in OmpF porin. Structure. 2013; 21:76-87.

146. Delcour AH. Structure and function of pore-forming β-barrels from bacteria. J Mol Microbiol Biotechnol. 2002; 4:1-10.

147. Criss AK, Seifert HS. A bacterial siren song: intimate interactions between *Neisseria* and neutrophils. Nat Rev Microbiol. 2012; 10:178-190.

148. Krishnan S, Prasadarao NV. Outer membrane protein A and OprF: versatile roles in Gram-negative bacterial infections. FEBS J. 2012; 279:919-931.

149. Smith SGJ, Mahon V, Lambert MA, Fagan RP. A molecular Swiss army knife: OmpA structure, function and expression. FEMS Microbiol Lett. 2007; 273:1-11.

150. Shin S, Lu G, Cai M, Kim KS. *Escherichia coli* outer membrane protein A adheres to human brain microvascular endothelial cells. Biochem Biophys Res Commun. 2005; 330:1199-1204.

151. Pautsch A, Schulz GE. Structure of the outer membrane protein A transmembrane domain. Nat Struct Biol. 1998; 5:1013-1017.

152. Pautsch A, Schulz GE. High-resolution structure of the OmpA membrane domain. J Mol Biol. 2000; 298:273-282.

153. Koebnik R. Structural and functional roles of the surface-exposed loops of the β-barrel membrane protein OmpA from *Escherichia coli*. J Bacteriol. 1999; 181:3688-3694.

154. Arora A, Abildgaard F, Bushweller JH, Tamm LK. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. Nat Struct Biol. 2001; 8:334-338.

155. Bond PJ, Faraldo-Gomez JD, Sansom MS. OmpA: a pore or not a pore? Simulation and modeling studies. Biophys J. 2002; 83:763-775.

156. Bond PJ, Sansom MS. The simulation approach to bacterial outer membrane proteins. Mol Membr Biol. 2004; 21:151-161.

157. Hong H, Szabo G, Tamm LK. Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. Nat Chem Biol. 2006; 2:627-635.

158. Macfarlane ELA, Kwasnicka A, Ochs MM, Hancock REW. PhoP-PhoQ homologues in *Pseudomonas aeruginosa* regulate expression of the outer-membrane protein OprH and polymyxin B resistance. Mol Microbiol. 1999; 34:305-316.

159. Qiao S, Luo Q, Zhao Y, Zhang XC, Huang Y. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. Nature. 2014.

160. Snijder HJ, Dijkstra BW. Bacterial phospholipase A: structure and function of an integral membrane phospholipase. Biochim Biophys Acta. 2000; 1488:91-101.

161. Istivan TS, Coloe PJ. Phospholipase A in Gram-negative bacteria and its role in pathogenesis. Microbiology. 2006; 152:1263-1274.

162. Dekker N. Outer-membrane phospholipase A: known structure, unknown biological function. Mol Microbiol. 2000; 35:711-717.

163. Cavard D, Rampini C, Barbu E, Polonovski J. [Phospholipase activity and other modifications in metabolism of the phospholipids consequent to the action of the colicins on *E. coli*]. Bull Soc Chim Biol (Paris). 1968; 50:1455-1471.

164. Dirusso CC, Black PN. Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. J Biol Chem. 2004; 279:49563-49566.

165. Hellion P, Landry F, Subbaiah PV, Proulx P. The uptake and acylation of exogenous lysophosphatidylethanolamine by *Escherichia coli* cells. Can J Biochem. 1980; 58:1381-1386.

166. Homma H, Nishijima M, Kobayashi T, Okuyama H, Nojima S. Incorporation and metabolism of 2-acyl lysophospholipids by *Escherichia coli*. Biochim Biophys Acta. 1981; 663:1-13.

167. Buchanan SK. β-barrel proteins from bacterial outer membranes: structure, function and refolding. Curr Opin Struct Biol. 1999; 9:455-461.

168. Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR, Dekker N, Dijkstra BW. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. Nature. 1999; 401:717-721.

169. Baaden M, Meier C, Sansom MS. A molecular dynamics investigation of mono and dimeric states of the outer membrane enzyme OMPLA. J Mol Biol. 2003; 331:177-189.

170. Knirel Y, Valvano M: Bacterial lipopolysaccharides: structure, chemical synthesis, biogenesis and Interaction with host cells. Springer; 1st Edition. edition; 2011.

171. Kawasaki K, Teramoto M, Tatsui R, Amamoto S. Lipid A 3'-O-deacylation by *Salmonella* outer membrane enzyme LpxR modulates the ability of lipid A to stimulate Toll-like receptor 4. Biochem Biophys Res Commun. 2012; 428:343-347.

172. Kawasaki K, Ernst RK, Miller SI. 3-O-deacylation of lipid A by PagL, a PhoP/PhoQ-regulated deacylase of *Salmonella typhimurium*, modulates signaling through Toll-like receptor 4. J Biol Chem. 2004; 279:20044-20048.

173. Rutten L, Geurtsen J, Lambert W, Smolenaers JJ, Bonvin AM, de Haan A, van der Ley P, Egmond MR, Gros P, Tommassen J. Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A. 2006; 103:7071-7076.

174. Kawasaki K, China K, Nishijima M. Release of the lipopolysaccharide deacylase PagL from latency compensates for a lack of lipopolysaccharide aminoarabinose modification-dependent resistance to the antimicrobial peptide polymyxin B in *Salmonella enterica*. J Bacteriol. 2007; 189:4911-4919.

175. Evanics F, Hwang PM, Cheng Y, Kay LE, Prosser RS. Topology of an outer-membrane enzyme: measuring oxygen and water contacts in solution NMR studies of PagP. J Am Chem Soc. 2006; 128:8256-8264.

176. Khan MA, Neale C, Michaux C, Pomes R, Prive GG, Woody RW, Bishop RE. Gauging a hydrocarbon ruler by an intrinsic exciton probe. Biochemistry. 2007; 46:4565-4579.

177. Bishop RE. The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. Mol Microbiol. 2005; 57:900-912.

178. Khan MA, Bishop RE. Molecular mechanism for lateral lipid diffusion between the outer membrane external leaflet and a β-barrel hydrocarbon ruler. Biochemistry. 2009; 48:9745-9756.

179. Ahn VE, Lo EI, Engel CK, Chen L, Hwang PM, Kay LE, Bishop RE, Prive GG. A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. EMBO J. 2004; 23:2931-2941.

180. Suomalainen M, Haiko J, Ramu P, Lobo L, Kukkonen M, Westerlund-Wikstrom B, Virkola R, Lahteenmaki K, Korhonen TK: Using every trick in the book: the pla surface protease of *Yersinia pestis*. In Genus Yersinia: From Genomics to Function. Volume 603. Perry RD, Fetherston JD; 2007: 268-278: Advances in Experimental Medicine and Biology].

181. Haiko J, Suomalainen M, Ojala T, Lahteenmaki K, Korhonen TK. Invited review: breaking barriers: attack on innate immune defences by omptin surface proteases of enterobacterial pathogens. Innate Immun. 2009; 15:67-80.

182. Rupprecht KR, Gordon G, Lundrigan M, Gayda RC, Markovitz A, Earhart C. omp T: *Escherichia coli* K-12 structural gene for protein A (3b). J Bacteriol. 1983; 153:1104-1106.

183. Sugimura K, Nishihara T. Purification, characterization, and primary structure of *Escherichia coli* protease VII with specificity for paired basic residues: identity of protease VII and OmpT. J Bacteriol. 1988; 170:5625-5632.

184. Kukkonen M, Korhonen TK. The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. Int J Med Microbiol. 2004; 294:7-14.

185. Premjani V, Tilley D, Gruenheid S, Le Moual H, Samis JA. Enterohemorrhagic *Escherichia coli* OmpT regulates outer membrane vesicle biogenesis. FEMS Microbiol Lett. 2014; 355:185-192.

186. Dorrell N, Martino MC, Stabler RA, Ward SJ, Zhang ZW, McColm AA, Farthing MJ, Wren BW. Characterization of *Helicobacter pylori* PldA, a phospholipase with a role in colonization of the gastric mucosa. Gastroenterology. 1999; 117:1098-1104.

187. Ziprin RL, Young CR, Byrd JA, Stanker LH, Hume ME, Gray SA, Kim BJ, Konkel ME. Role of *Campylobacter jejuni* potential virulence genes in cecal colonization. Avian Dis. 2001; 45:549-557.

188. Berstad AE, Berstad K, Berstad A. pH-activated phospholipase A2: an important mucosal barrier breaker in peptic ulcer disease. Scand J Gastroenterol. 2002; 37:738-742.

189. Bukholm G, Tannaes T, Nedenskov P, Esbensen Y, Grav HJ, Hovig T, Ariansen S, Guldvog I. Colony variation of *Helicobacter pylori*: pathogenic potential is correlated to cell wall lipid composition. Scand J Gastroenterol. 1997; 32:445-454.

190. Xerry J, Owen RJ. Conservation and microdiversity of the phospholipase A (*pldA*) gene of *Helicobacter pylori* infecting dyspeptics from different countries. FEMS Immunol Med Microbiol. 2001; 32:17-25.

191. Tannaes T, Bukholm IK, Bukholm G. High relative content of lysophospholipids of *Helicobacter pylori* mediates increased risk for ulcer disease. FEMS Immunol Med Microbiol. 2005; 44:17-23.

192. NCBI Resource Coordinators. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2015; 43:D6.

193. Jacob-Dubuisson F, Striker R, Hultgren SJ. Chaperone-assisted self-assembly of pili independent of cellular energy. J Biol Chem. 1994; 269:12447-12455.

194. Daniels R, Normark S. Twin ushers guide pili across the bacterial outer membrane. Cell. 2008; 133:574-576.

195. Huang Y, Smith BS, Chen LX, Baxter RH, Deisenhofer J. Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. Proc Natl Acad Sci U S A. 2009; 106:7403-7407.

196. Tsirigos KD, Bagos PG, Hamodrakas SJ. OMPdb: a database of β-barrel outer membrane proteins from Gram-negative bacteria. Nucleic Acids Res. 2011; 39:D324-331.

197. Saier MH, Jr., Reddy VS, Tamang DG, Vastermark A. The transporter classification database. Nucleic Acids Res. 2014; 42:D251-258.

198. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014; 42:D222-230.

199. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, et al. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. 2015; 43:D376-381.

200. Fox NK, Brenner SE, Chandonia JM. SCOPe: structural classification of proteins-extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res. 2014; 42:D304-309.

201. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36:W5-9.

202. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010; 11:595.

203. Hekkelman ML, Vriend G. MRS: a fast and compact retrieval system for biological data. Nucleic Acids Res. 2005; 33:W766-769.

204. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999; 27:29-34.

205. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235-242.

206. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999; 41.

207. Puigbò P, Bravo I, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biol Direct. 2008; 3.
208. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. ClustalW and ClustalX version 2. Bioinformatics. 2007; 23:2947–2948.
209. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539.
210. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010; 38:W695-699.
211. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008; 9:286-298.
212. Pride D. SWAAP Version 1.0: sliding windows alignment analysis program: a tool for analyzing patterns of substitutions and similarity in multiple alignments. Distributed by the author. 2000.
213. Puigbo P, Garcia-Vallve S, McInerney JO. TOPD/FMTS: a new software to compare phylogenetic trees. Bioinformatics. 2007; 23:1556-1558.
214. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 2000; 28:1102, 1104.
215. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA: a self-parameterizing force field. Proteins. 2002; 47:393-402.
216. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph. 1990; 8:52-56, 29.
217. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: a multiple structural alignment algorithm. Proteins. 2006; 64:559-574.
218. Vriend G: WHAT IF server [http://swift.cmbi.ru.nl/servers/html/index.html] Accessed Date: 2007-2017.
219. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins. 2009; 77 Suppl 9:114-122.
220. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. . BMC Struct Biol. 2009; 9:51.
221. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph. 1996; 14:354-360, 376.
222. Voorintholt R, Kosters MT, Vegter G, Vriend G, Hol WG. A very fast program for visualizing protein surfaces, channels and cavities. J Mol Graph. 1989; 7:243-245.
223. Pellegrini-Calace M, Maiwald T, Thornton JM. PoreWalker: A Novel Tool for the Identification and Characterization of Channels in Transmembrane Proteins from Their Three-Dimensional Structure. PLoS Comput Biol. 2009; 5.
224. Taboada B, Ciria R, Martinez-Guerrero CE, Merino E. ProOpDB: prokaryotic operon database. Nucleic Acids Res. 2012; 40:D627-631.
225. Taboada B, Verde C, Merino E. High accuracy operon prediction method based on STRING database scores. Nucleic Acids Res. 2010; 38:e130.
226. Oliveira L, Paiva PB, Paiva AC, Vriend G. Identification of functionally conserved residues with the use of entropy-variability plots. Proteins. 2003; 52:544-552.

227. Folkertsma S, van Noort P, Van Durme J, Joosten HJ, Bettler E, Fleuren W, Oliveira L, Horn F, de Vlieg J, Vriend G. A family-based approach reveals the function of residues in the nuclear receptor ligand-binding domain. J Mol Biol. 2004; 341:321-335.

228. Vollan HS, Tannæs T, Vriend G, Bukholm GS. *In Silico* structure and sequence analysis of bacterial porins and specific diffusion channels for hydrophilic molecules: conservation, multimericity and multifunctionality. International Journal of Molecular Sciences. 2016; 17:599.

229. Stecher B, Maier L, Hardt WD. 'Blooming' in the gut: how dysbiosis might contribute to pathogen evolution. Nat Rev Microbiol. 2013; 11:277-284.

230. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. Crystal structures explain functional properties of two *E. coli* porins. Nature. 1992; 358:727-733.

231. Dutzler R, Rummel G, Alberti S, Hernandez-Alles S, Phale P, Rosenbusch J, Benedi V, Schirmer T. Crystal structure and functional characterization of OmpK36, the osmoporin of *Klebsiella pneumoniae*. Structure. 1999; 7:425-434.

232. Forst D, Welte W, Wacker T, Diederichs K. Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. Nat Struct Biol. 1998; 5:37-46.

233. Van Gelder P, Dutzler R, Dumas F, Koebnik R, Schirmer T. Sucrose transport through maltoporin mutants of *Escherichia coli*. Protein Eng. 2001; 14:943-948.

234. Anbazhagan V, Qu J, Kleinschmidt JH, Marsh D. Incorporation of outer membrane protein OmpG in lipid membranes: protein-lipid interactions and β-barrel orientation. Biochemistry. 2008; 47:6189-6198.

235. Yildiz O, Vinothkumar KR, Goswami P, Kuhlbrandt W. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. EMBO J. 2006; 25:3702-3713.

236. Conlan S, Zhang Y, Cheley S, Bayley H. Biochemical and biophysical characterization of OmpG: a monomeric porin. Biochemistry. 2000; 39:11845-11854.

237. Damaghi M, Bippes C, Koster S, Yildiz O, Mari SA, Kuhlbrandt W, Muller DJ. pH-dependent interactions guide the folding and gate the transmembrane pore of the β-barrel membrane protein OmpG. J Mol Biol. 2010; 397:878-882.

238. Novikova OD, Solovyeva TF. Non-specific porins of the outer membrane of Gram-negative bacteria: structure and functions. Biol Membrany. 2009; 26:6-20.

239. Mari SA, Koster S, Bippes CA, Yildiz O, Kuhlbrandt W, Muller DJ. pH-induced conformational change of the β-barrel-forming protein OmpG reconstituted into native *E. coli* lipids. J Mol Biol. 2010; 396:610-616.

240. Fajardo DA, Cheung J, Ito C, Sugawara E, Nikaido H, Misra R. Biochemistry and regulation of a novel *Escherichia coli* K-12 porin protein, OmpG, which produces unusually large channels. J Bacteriol. 1998; 180:4452-4459.

241. Power ML, Ferrari BC, Littlefield-Wyer J, Gordon DM, Slade MB, Veal DA. A naturally occurring novel allele of *Escherichia coli* outer membrane protein A reduces sensitivity to bacteriophage. Appl Environ Microbiol. 2006; 72:7930-7932.

242. Bjornsson H, Marteinsson V, Frijonsson OH, Linke D, Benediktsdottir E. Isolation and characterization of an antigen from the fish pathogen *Moritella viscosa*. J Appl Microbiol. 2011; 111:17-25.

243. Reusch RN. Insights into the structure and assembly of *Escherichia coli* outer membrane protein A. European Journal of Biochemistry. 2012; 279:894-909.

244. Iordanov I, Renault M, Reat V, Bosshart PD, Engel A, Saurel O, Milon A. Dynamics of *Klebsiella pneumoniae* OmpA transmembrane domain: the four extracellular loops

display restricted motion behavior in micelles and in lipid bilayers. Biochimica Et Biophysica Acta-Biomembranes. 2012; 1818:2344-2353.

245. Sanders CR, Mittendorf KF. Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. Biochemistry. 2011; 50:7858-7867.

246. Bower S, Rosenthal KS. The bacterial cell wall: the armor, artillery, and achilles heel. IDCP. 2006; 14:309-317.

247. Karuppiah V, Berry JL, Derrick JP. Outer membrane translocons: structural insights into channel formation. Trends Microbiol. 2011; 19:40-48.

248. Schalk IJ, Mislin GL, Brillet K. Structure, function and binding selectivity and stereoselectivity of siderophore-iron outer membrane transporters. Curr Top Membr. 2012; 69:37-66.

249. Bauer FJ, Rudel T, Stein M, Meyer TF. Mutagenesis of the *Neisseria gonorrhoeae* porin reduces invasion in epithelial cells and enhances phagocyte responsiveness. Mol Microbiol. 1999; 31:903-913.

250. Derrick JP, Urwin R, Suker J, Feavers IM, Maiden MCJ. Structural and evolutionary inference from molecular variation in *Neisseria* porins. Infect Immun. 1999; 67:2406-2413.

251. Tanabe M, Nimigean CM, Iverson TM. Structural basis for solute transport, nucleotide regulation, and immunological recognition of *Neisseria meningitidis* PorB. Proc Natl Acad Sci U S A. 2010; 107:6811-6816.

252. Massari P, Ram S, Macleod H, Wetzler LM. The role of porins in neisserial pathogenesis and immunity. Trends Microbiol. 2003; 11:87-93.

253. Wetzler LM. Innate immune function of the neisserial porins and the relationship to vaccine adjuvant activity. Future Microbiol. 2010; 5:749-758.

254. Kukkonen M, Suomalainen M, Kyllonen P, Lahteenmaki K, Lang H, Virkola R, Helander IM, Holst O, Korhonen TK. Lack of O-antigen is essential for plasminogen activation by *Yersinia pestis* and *Salmonella enterica*. Mol Microbiol. 2004; 51:215-225.

255. Haiko J, Kukkonen M, Ravantti JJ, Westerlund-Wikstrom B, Korhonen TK. The single substitution I259T, conserved in the plasminogen activator Pla of pandemic *Yersinia pestis* branches, enhances fibrinolytic activity. J Bacteriol. 2009; 191:4758-4766.

256. Yu M, Sun P, He Y, Xiao L, Sun D, Zhang L, Tian C. Mutation of the critical pH-gating residues histidine 231 to glutamate increase open probability of outer membrane protein G in planar lipid bilayer. Protein Cell. 2013; 4:803-806.

257. Besya AB, Mobasheri H, Ejtehadi MR. Gating and conduction of nano-channel forming proteins: a computational approach. J Biomol Struct Dyn. 2012.

258. Chan WC, Schirmer T, Ferenci T. Combinatorial mutagenesis analysis of residues in the channel constriction loop L3 and neighbouring β-strands in the LamB glycoporin of *Escherichia coli*. Mol Membr Biol. 1996; 13:41-48.

259. Stenkova AM, Isaeva MP, Shubin FN, Rasskazov VA, Rakin AV. Trends of the major porin gene (ompF) evolution: insight from the genus *Yersinia*. PLoS One. 2011; 6:e20546.

260. Basle A, Qutub R, Mehrazin M, Wibbenmeyer J, Delcour AH, Bayley H. Deletions of single extracellular loops affect pH sensitivity, but not voltage dependence, of the *Escherichia coli* porin OmpF. Protein Eng Des Sel. 2004; 17:665-672.

261. LeClerc JE, Li B, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science. 1996; 274:1208-1211.

262. Easton D. Bacterial outer membrane porins: are they decoy agents? Australian Journal of Medical Science. 2005; 26.

263. Wang H, Andersen KK, Vad BS, Otzen DE. OmpA can form folded and unfolded oligomers. Biochim Biophys Acta. 2013; 1834:127-136.

264. Dunton TA, Goose JE, Gavaghan DJ, Sansom MS, Osborne JM. The free energy landscape of dimerization of a membrane protein, NanC. PLoS Comput Biol. 2014; 10:e1003417.

265. Shrivastava R, Basu B, Godbole A, Mathew MK, Apte SK, Phale PS. Repression of the glucose-inducible outer-membrane protein OprB during utilization of aromatic compounds and organic acids in *Pseudomonas putida* CSV86. Microbiology. 2011; 157:1531-1540.

266. Biswas S, Mohammad MM, Movileanu L, van den Berg B. Crystal structure of the outer membrane protein OpdK from *Pseudomonas aeruginosa*. Structure. 2008; 16:1027-1035.

267. Catel-Ferreira M, Nehme R, Molle V, Aranda J, Bouffartigues E, Chevalier S, Bou G, Jouenne T, De E. Deciphering the function of the outer membrane protein OprD homologue of *Acinetobacter baumannii*. Antimicrob Agents Chemother. 2012; 56:3826-3832.

268. Ishida H, Garcia-Herrero A, Vogel HJ. The periplasmic domain of *Escherichia coli* outer membrane protein A can undergo a localized temperature dependent structural transition. Biochim Biophys Acta. 2014; 1838:3014-3024.

269. Anand A, LeDoyt M, Karanian C, Luthra A, Koszelak-Rosenblum M, Malkowski MG, Puthenveetil R, Vinogradova O, Radolf JD. Bipartite topology of *Treponema pallidum* repeat proteins C/D and I: Outer membrane insertion, trimerization and porin function require a C-terminal β-barrel domain. J Biol Chem. 2015; 290:12313-12331.

270. Tannaes T, Bukholm G. Cholesteryl-6-O-acyl-alpha-D-glucopyranoside of *Helicobacter pylori* relate to relative lysophospholipid content. FEMS Microbiol Lett. 2005; 244:117-120.

271. Cullen TW, Giles DK, Wolf LN, Ecobichon C, Boneca IG, Trent MS. *Helicobacter pylori* versus the host: remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. PLoS Pathog. 2011; 7:e1002454.

272. Lai CH, Hsu YM, Wang HJ, Wang WC. Manipulation of host cholesterol by *Helicobacter pylori* for their beneficial ecological niche. BioMedicine. 2013; 3:27-33.

273. McGee DJ, George AE, Trainor EA, Horton KE, Hildebrandt E, Testerman TL. Cholesterol enhances *Helicobacter pylori* resistance to antibiotics and LL-37. Antimicrob Agents Chemother. 2011; 55:2897-2904.

274. Brok RG, Boots AP, Dekker N, Verheij HM, Tommassen J. Sequence comparison of outer membrane phospholipases A: implications for structure and for the catalytic mechanism. Res Microbiol. 1998; 149:703-710.

275. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012; 13:341.

276. Metzker ML. Emerging technologies in DNA sequencing. Genome Res. 2005; 15:1767-1776.

277. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008; 9:387-402.

278. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, et al. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003; 299:1582-1585.

279. Kawai M, Furuta Y, Yahara K, Tsuru T, Oshima K, Handa N, Takahashi N, Yoshida M, Azuma T, Hattori M, et al. Evolution in an oncogenic bacterial species with

extreme genome plasticity: *Helicobacter pylori* East Asian genomes. BMC Microbiol. 2011; 11:104.

280. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. Genes under positive selection in *Escherichia coli*. Genome Res. 2007; 17:1336-1343.

281. Perez LJ, Diaz de Arce H, Perera CL, Rosell R, Frias MT, Percedo MI, Tarradas J, Dominguez P, Nunez JI, Ganges L. Positive selection pressure on the B/C domains of the E2-gene of classical swine fever virus in endemic areas under C-strain vaccination. Infect Genet Evol. 2012; 12:1405-1412.

282. Esbensen Y, Vollan HS, TM T. A functional outer membrane phospholipase A (OMPLA) is required for survival of *Helicobacter pylori* at pH 3.5 [abstract]. . Helicobacter. 2011; 16:97-98.

283. Wu EL, Fleming PJ, Yeom MS, Widmalm G, Klauda JB, Fleming KG, Im W. *E. coli* outer membrane and interactions with OMPLA. Biophys J. 2014; 106:2493-2502.

284. Sandoz G, Douguet D, Chatelain F, Lazdunski M, Lesage F. Extracellular acidification exerts opposite actions on TREK1 and TREK2 potassium channels via a single conserved histidine residue. Proc Natl Acad Sci U S A. 2009; 106:14628-14633.

285. Niemeyer MI, Gonzalez-Nilo FD, Zuniga L, Gonzalez W, Cid LP, Sepulveda FV. Neutralization of a single arginine residue gates open a two-pore domain, alkali-activated K^+ channel. Proc Natl Acad Sci U S A. 2007; 104:666-671.

286. Wakabayashi S, Hisamitsu T, Pang T, Shigekawa M. Mutations of Arg^{440} and Gly^{455}/Gly^{456} oppositely change pH sensing of Na^+/H^+ exchanger 1. J Biol Chem. 2003; 278:11828-11835.

287. Hisamitsu T, Yamada K, Nakamura TY, Wakabayashi S. Functional importance of charged residues within the putative intracellular loops in pH regulation by Na^+/H^+ exchanger NHE1. FEBS J. 2007; 274:4326-4335.

288. Li CC, Ho TY, Kao CH, Wu SL, Liang JA, Hsiang CY. Conserved charged amino acid residues in the extracellular region of sodium/iodide symporter are critical for iodide transport activity. J Biomed Sci. 2010; 17:89.

289. Liu N, Delcour AH. Inhibitory effect of acidic pH on OmpC porin: wild-type and mutant studies. FEBS Lett. 1998; 434:160-164.

290. Bala PA, Foster J, Carvelli L, Henry LK. SLC6 Transporters: Structure, Function, Regulation, Disease Association and Therapeutics. Mol Aspects Med. 2013; 34:197-219.

291. Malinauskaite L, Quick M, Reinhard L, Lyons JA, Yano H, Javitch JA, Nissen P. A mechanism for intracellular release of Na^+ by neurotransmitter/sodium symporters. Nat Struct Mol Biol. 2014; 21:1006-1012.

292. Singh SK, Pal A. Biophysical Approaches to the Study of LeuT, a Prokaryotic Homolog of Neurotransmitter Sodium Symporters. Methods Enzymol. 2015; 557:167-198.

293. Vollan HS, Tannæs T, Vriend G, Bukholm G. *In silico* structure and sequence analysis of bacterial porins and specific diffusion channels for hydrophilic molecules: Conservation, multimericity and multifunctionality. IJMS. 2016; 17:599.

294. Gradogna A, Gaitán-Peñas H, Laparra-Cuervo L, Solsona C, Fernández-Dueñas V, Barrallo-Gimeno A, Ciruela F, Lakadamyali M, Estévez R, Pusch M. Novel Properties of LRRC8-Mediated VRAC Currents. Biophysical Journal. 2017; 112:416a-417a.

295. Le Caherec F, Bron P, Verbavatz JM, Garret A, Morel G, Cavalier A, Bonnec G, Thomas D, Gouranton J, Hubert JF. Incorporation of proteins into (Xenopus) oocytes by proteoliposome microinjection: functional characterization of a novel aquaporin. J Cell Sci. 1996; 109 (Pt 6):1285-1295.

Appendix

Table of Contents

List of Appendix Figures:

1 Abbreviations

AbPirA	*Acinetobacter baumannii* PirA
EcBamA	*E. coli* BamA
EcFadL	*E. coli* FadL
EcLptD	*E. coli* LptD
EcMaltoporin	*E. coli* maltoporin
EcOmp	*E. coli* Outer membrane protein (e.g. EcOmpA, EcOmpC, and EcOmpF)
EcOMPLA	*E. coli* Outer membrane phospholipase A
GDP	General Diffusion Porin
H-bonds	Hydrogen bonds
HdBamA	*Haemophilus ducreyi* BamA
KpLptD	*Klebsiella pneumoniae* LptD
KpOmpA	*Klebsiella pneumoniae* OmpA
L	Large
LPS	Lipopolysaccharide
M	Medium
NgBamA	*Neisseria gonorrhoeae* BamA
Occ	Outer membrane carboxylate channel (*e.g.* OccD1-D3 and OccK1-K11)
OMP	Outer membrane protein (*e.g.* Omp32, OmpA, OmpC, OmpF, and OmpX)
OMPLA	Outer membrane phospholipase A
PaFadL	*Pseudomonas aeruginosa* FadL
PaLptD	*Pseudomonas aeruginosa* LptD
PaPirA	*Pseudomonas aeruginosa* PirA
PDB	Protein Data Bank
POTRA	Polypeptide-TRansport-Associated
RbGDP	*Rhodopseudomonas blastica* GDP
RcGDP	*Rhodobacter capsulatus* GDP
S	Small
SeLptD	*Salmonella enterica* LptD
SfLptD	*Shigella flexneri* LptD
StMaltoporin	*Salmonella Typhimurium* Maltoporin
StOmp	*Salmonella Typhimurium* outer membrane protein (OmpC, OmpF)
StOMPLA	*Salmonella Typhimurium* OMPLA
YpLptD	*Yersinia pestis* LptD

2 Secondary structure preference

Amino acids have secondary structure preference that is energetically favorable in a protein fold. Appendix Figure 1 illustrates which amino acids prefer helices, β -strands, or turns. This figure also lists amino acid hydrophobicity, size and chemical structure.

Appendix Figure 1: Building blocks of Life. *This figure lists numerous important residue properties. From left to right: the amino acids (1-letter code, 3-letter code and complete name), size of side chain (S, small; M, medium; L, large), polar residues are marked with +/- signs, secondary structure preference (α-helix colored blue, β-strands colored red and loops colored green), and the chemical structure.*

We will focus on the two most common secondary structure motifs: right handed α-helix (plural form: α-helices) and β-sheet (plural form: β-strands). Interactions that stabilize the protein in the secondary structure include hydrogen bonds. An α-helix is a spiral of residues with ideally 3.6 residues per turn (*e.g.* the periplasmic efflux pump helices in Appendix Figure 23). The hydrogen bonds (H-bonds) run parallel to the helix axis. It is a stable structure. Due to these H-bonds, a helix structure can retain its fold even if some bonds are broken. Like the α-helix, the β-strand is highly stable too. The antiparallel β-sheet conformation is found in Appendix Figure 1 (antiparallel is when the strands are in opposite direction).

3   Structure alignment OMPs (Mustang motif aligner)

Data collection

The structures found in Table 1 and Appendix Table 8 were used to analyze each OMP subfamily. However, the following proteins were published after the time of the data collection for this study (9th March 2014), and have not been analyzed in this thesis. They are listed in Table 1 and highlighted red in Appendix Table 8: CmeC, OprN, StOMPLA, COG4313, KpLptD, PaLptD, YpLptD, FusA, HpuA, PaPirA, AbPirA, PiuA, ZnuD, SusD, TamA, Opa60, Pallilysin, CymA, CsgG, Occ, OprO, PgaA, CarO, and OmpE36.

Efflux pumps

The solved transmembrane and periplasmic efflux pump structures have quite conserved motifs, but high sequence variability (an average of nearly 23% sequence identity of the residues aligned, see Appendix Table 1). Only the monomeric molecules were aligned together (see Appendix Figure 2). One monomeric structure consists of three β-strands and a trimeric structure is needed for a 12 β-stranded barrel. The pair-wise Mustang motif aligner results are found in Appendix Table 1.
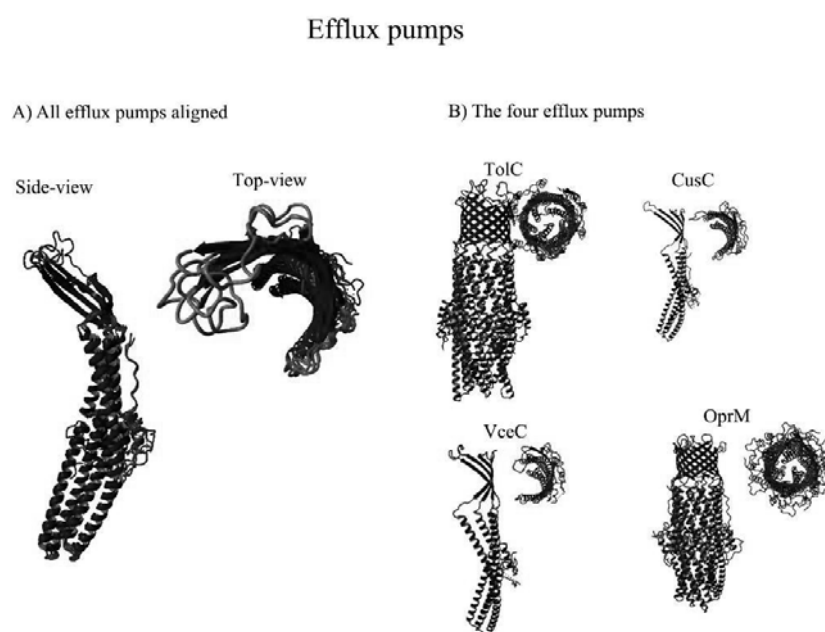
Appendix Figure 2: Efflux pumps. *Panel A) Side-view and top-view of the monomeric transmembrane efflux pump structure aligned using Mustang multiple object aligner. Panel B) E. coli TolC and CusC, V. cholarea VceC and P. aeruginosa OprM (PDB ID: 1EK9, 4K7R, 1YC9 and 3D5K). TolC and OprM are shown in trimeric state; the remaining are monomeric structures. See Appendix Table 8 for PDB IDs.*

Object 1	Object 2	RMSD (Å)	% Aligned Residues (smallest structure)	Sequence identity
CusC	OprM	1.30	94.17%	45.54%
CusC	VceC	1.55	89.78%	27.91%
VceC	OprM	1.51	89.05%	27.05%
OprM	TolC	1.70	78.97%	21.01%
VceC	TolC	1.78	81.75%	20.54%
CusC	TolC	1.42	82.01%	17.09%
Average		1.54	85.96%	26.52%

Enzymes

PDB currently holds three OM enzyme subfamilies with six solved structures (see Appendix Table 2 and Appendix Figure 3). These are the outer membrane phospholipase A (OMPLA), Lipid-A modifying enzyme (reducing endotoxicity) and omptins (inactivates host plasminogen) [4,5].

Appendix Table 2: Enzymes. *This table holds information from currently available OM enzyme structures derived from PDB (Protein Name, Subfamily, Function, PDB ID, Bacteriae and size). These six structures are visualized in Appendix Figure 3.*

Protein Name	Protein Subfamily	Enzymatic function	PDB ID	*Bacteriae*	Size
LpxR	Lipid-A-modifying	Deacylase	3FID	*Salmonella* Typhimurium	12
OMPLA	OM phospholipase	Phospholipase	1QD6	*Escherichia coli*	12
OmpT	Omptin	Protease	1I78	*Escherichia coli*	10
Pla	Omptin	Protease	2X55	*Yersinia pestis*	10
PagL	Lipid A modifying	Deacylase	2ERV	*Pseudomonas aeruginosa*	8
PagP	Lipid A modifying	Transferase	3GP6	*Escherichia coli*	8

OM Enzymes

A) LpxR   B) OMPLA   C) OmpT

D) PagL   E) PagP   F) Pla
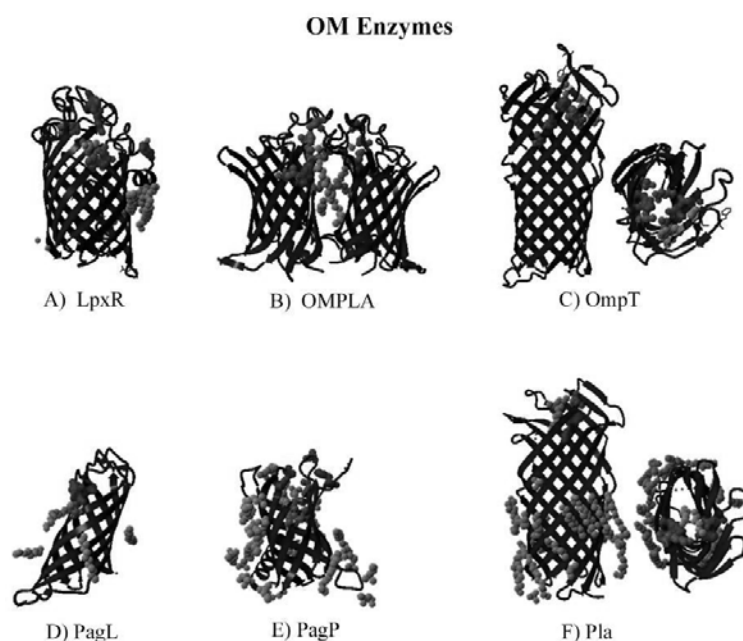
Appendix Figure 3: Enzymes. *The PDB currently holds six bacterial outer membrane (OM) enzymes from three subfamilies; lipid A modifying enzymes (Panel A, D, and E), OMPLA (Panel B) and omptins (Panel C and F). See Appendix Table 8 for more information regarding these structures. All structures are shown from a side-view orientation. Theo omptins subfamily is also visualized with a top-view orientation as the active site is found inside the structure (see Panel C and F). The structures are blue with residues important to enzyme activity highlighted; red represents the catalytic site, orange is important for substrate specificity, translocation, calcium binding, etc., and finally the yellow residues highlighted important residues with unknown functions detected through mutagenesis (OmpT have residues moderately affecting the enzyme activity). Panel D and E show monomeric enzyme with a tilted orientation found in PagL and PagP. PDB ID going left to right top to bottom: 3FID, 1QD6, 1I78, 2ERV, 3GP6 and 2X55).*

Although both *E. coli* OMPLA and *Salmonella* Typhimurium LpxR are composed of 12 β-stranded barrels, they vary greatly: only 44 residues could be structurally aligned with 1.91 Å RMSD and 6.82% sequence identity (using Mustang aligner; data not shown). The structural alignment between PagL and PagP yields quite different structure as shown in Appendix Figure 4A. The alignment yields a 1.52 Å RMSD for the 56.67% aligned structure with 8.24% sequence identity (in total, 85 residues were aligned with a total sequence length is 150 and 155, respectively). The 12-stranded LpxR was not compared to these eight stranded barrels (but OMPLA, see above). The omptin structures, OmpT and Pla, show the highest structure and sequence conservation among the enzyme structures with 1.08 Å RMSD over 256 aligned residues, see Appendix Figure 4B. The 53.91% sequence identity does indicate variations that could be related to niche specific adaptation.

LpxR (A) and omptins (B)

A) LpxR (PagL and PagP)   B) Omptins (OmpT and PlaA)

Side-view   Top-view   Side-view   Top-view
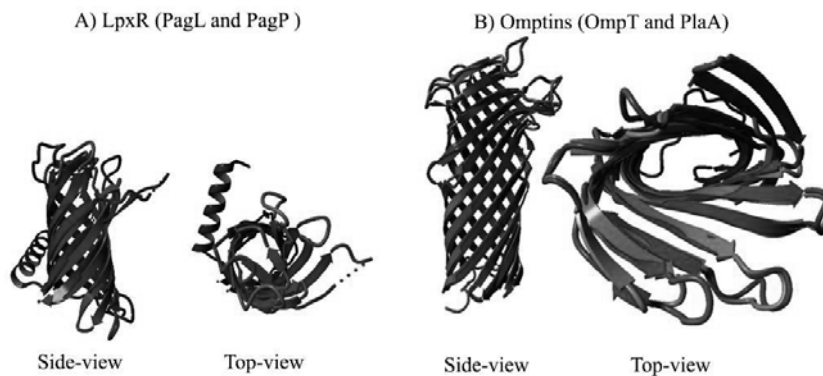
Appendix Figure 4: LpxR and omptins. *Panel A) PagL (blue) and PagP (purple) structurally aligned. B) The side and top-view (left to right) of the structural aligned omptins structures OmpT and Pla. See Appendix Table 8 for PDB IDs.*

Transporters

Transporters listed in Table 1 depict a vast variety of function and size. Bacteria have a wide range of different transporters and the number of different OMP transporters varies to withstand environmental conditions.

3.1.1   *Autotransporters*

Autotransporters vary in size, from 12 to 18 β-strands. EspP and Hbp (both isolated from *E. coli*) show the most similarity, while they have little in common with the larger AlgE autotransporter, see Appendix Table 3 and Appendix Figure 5. Hia (2GR8 from *H. influenzae*), EspP (2QOM from *E. coli*), EstA (3KVN from *P. aeruginosa*), Hbp (3AEH from *E. coli*), AlgE (3RBH from *P. aeruginosa*) and NalP (1UYN from N. meningitis) constitute the solved autotransporter structures in the PDB. In addition, the solved passenger domain of IscsA (3ML3 from *S. flexeri*) was excluded from further analyzes since it lacks the β-barrel domain. Two recent structures were solved (TamA 4L00 and 2LME YadA are not included in these analyzes). Autotransporters from the same species might have specialized functions, like EstA and AlgE from *P. aeruginosa*, that are of different size and not very conserved (less than 10% sequence identity where less than 23% of structure could be aligned).

Appendix Table 3: Autotransporters. *Mustang structure alignment results for autotransporters, using same parameters as described for Appendix Table 1. See Appendix Table 8 for a complete list of protein details, including PDB IDs.*

Object 1	Object 2	RMSD (Å)	Aligned (%)	Sequence identity (%)
EspP	Hbp	0.522	88.68%	65.53%
Hia	AlgE	1.384	26.50%	17.74%
EspP	NalP	1.664	72.83%	16.58%
EstA	NalP	1.532	87.00%	15.77%
Hbp	AlgE	2.111	3.88%	15.38%
Hbp	EstA	1.506	58.48%	14.81%
Hia	EspP	1.647	42.31%	12.12%
Hia	NalP	1.87	57.26%	11.94%
Hbp	NalP	1.669	59.86%	10.18%
NalP	AlgE	2.075	25.81%	9.72%
EstA	AlgE	1.693	22.74%	9.52%
Hia	Hbp	1.947	45.30%	9.43%
Hia	EstA	1.646	51.71%	9.09%
EspP	EstA	1.447	60.38%	7.50%
EspP	AlgE	1.403	29.81%	5.06%

Autotransporters
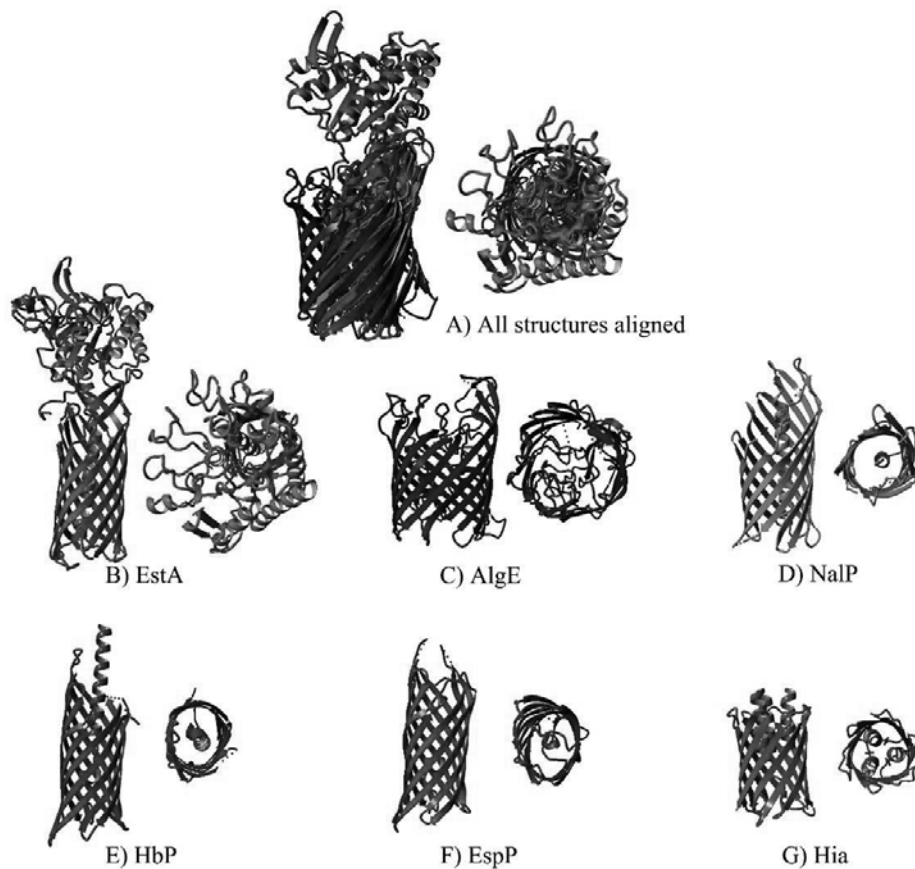
Appendix Figure 5: Autotransporters. *Top and side-view of all six autotransporters aligned using structure-based Mustang alignment (on all structures; panel A). Panel B-G shows the structures with the protein name listed below. See Appendix Table 8 for PDB IDs.*

3.1.2 Hydrophobic channels

The LptD family transports lipopolysaccharide (LPS). The LPS is transported from the periplasm to the outer membrane by LptD transporters (often in complex with other proteins, including LptE) [6]. The three solved structures are found in Appendix Figure 6, and Appendix Table 4 lists the results from a conserved family (after Mustang alignment).

LptD-family

A) All structures aligned

B) SeLptD   C) SfLptD   D) EcLptD
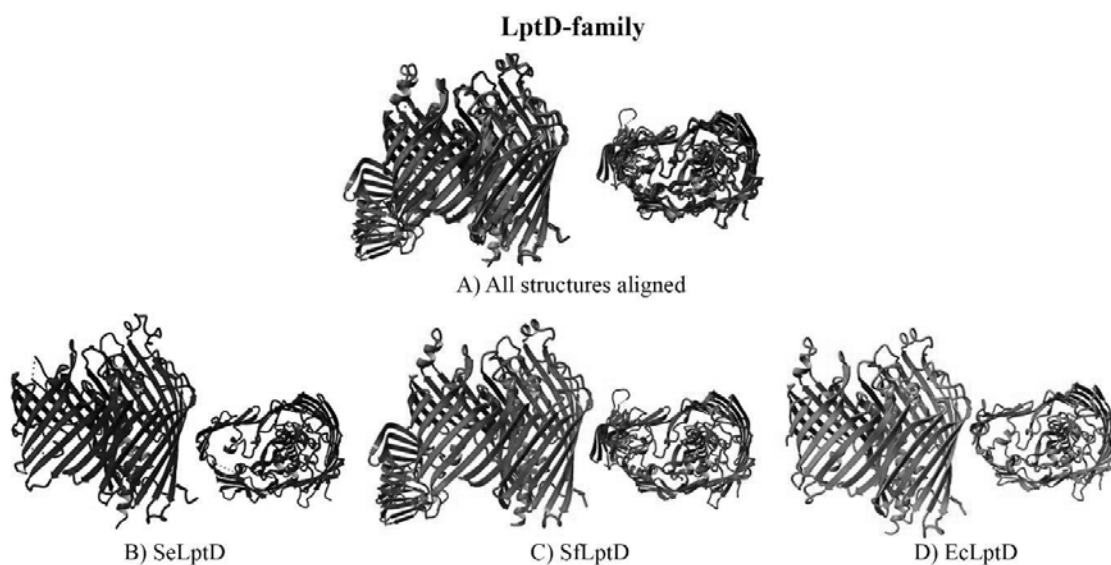
Appendix Figure 6: LptD-family of hydrophobic transporters. *Top and side-view of all three autotransporters aligned using structure-based Mustang alignment (on all structures; panel A). Panel B-D shows the structures with the protein name listed below. See Appendix Table 8 for PDB IDs.*

Appendix Table 4: LptD-family of hydrophobic transporters. *Mustang structure alignment results for LptD-family of hydrophobic transporters, using same parameters as described for Appendix Table 1. See Appendix Table 8 for a complete list of protein details, including PDB IDs.*

Object 1	Object 2	RMSD (Å)	Aligned (%)	Sequence identity (%)
SeLptD	SfLpD	0.76	97.2 %	88.40 %
SeLptD	EcLptD	0.79	97.4 %	88.87 %
SfLptD	EcLptD	0.72	99.6 %	99.57 %

The solved structures for the *FadL-family* of hydrophobic transporters are aligned are shown in Appendix Figure 7 and Appendix Table 5.
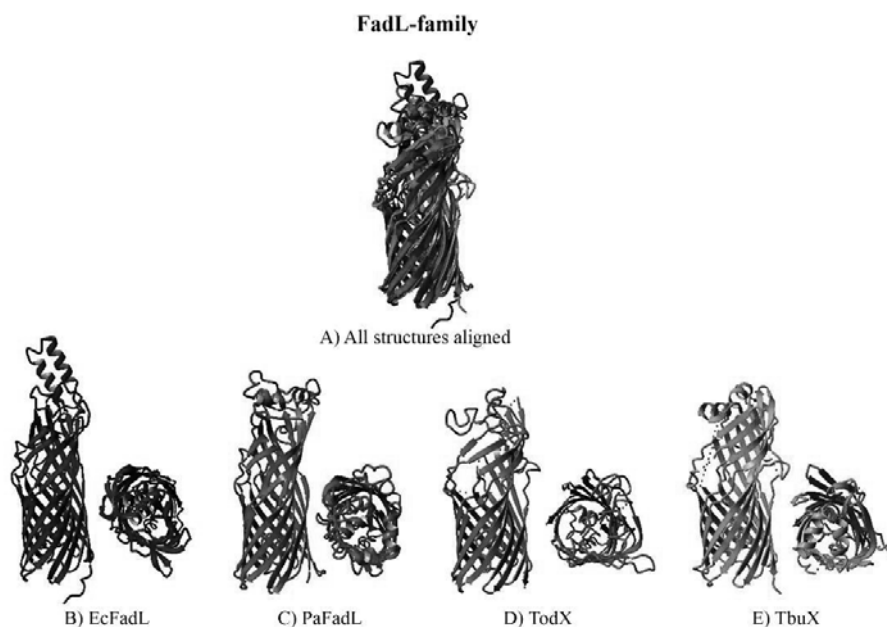
Appendix Figure 7: FadL hydrophobic transporters. *Top and side-view of all four FadL-transporters aligned using structure-based Mustang alignment (on all structures; panel A). Panel B-E shows the structures with the protein name listed below. See Appendix Table 8 for PDB IDs.*

Appendix Table 5: FadL-family of hydrophobic channels. *Mustang structure alignment results for the FadL-family of hydrophobic transporters, using same parameters as described for Appendix Table 1. See Appendix Table 8 for a complete list of protein details, including PDB IDs.*

Object 1	Object 2	RMSD (Å)	Aligned (%)	Sequence identity (%)
EcFadL	TbuX	1,77	74.4 %	19.7 %
EcFadL	TodX	1,91	70.6 %	17.8 %
EcFadL	PaFadl	1,10	79.0 %	25.7 %
TbuX	TodX	1,37	88.7 %	52.6 %
TodX	PaFadl	1,82	69.6 %	24.2 %

Appendix Table 5 show a diverse protein family with sequence identities ranging between ~18% to ~53%. This is likely due to differ rent substrate specificities, and the FadL family can be further divided into subclasses based on substrate specificity; long-chain fatty acid (FadL) and hydrocarbons (TodX and Tbux).

The structural alignment between TbuX and TodX yields a more conserved alignment of 1.4 Å RMSD and 52.6% sequence identity when ~89% of the Tbux structure is aligned (which is 15 residues shorter than TodX), as shown in Appendix Figure 8.

Fatty acid transporters

A) Side-view   B) Top-view
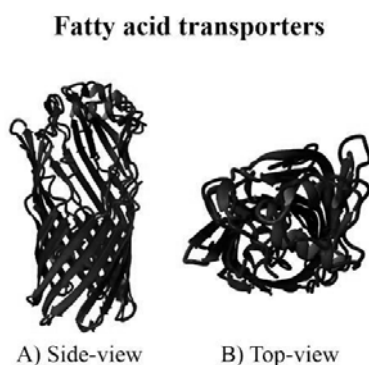
Appendix Figure 8: Fatty acid transporters. The pairwise Mustang alignment of the 14- stranded TbuX (blue) and TodX (purple) with side-view (left) and top-view (right). See Appendix Table 8 for PDB IDs.

The structural alignment between two non-FadL family members of hydrophobic channels, OmpW and OprG, have a slightly more conserved structure compared to that of TodX/Tbux. With a 1.09 Å RMSD with 52.35% sequence identity over 93.40% of the OmpW structure (the shortest sequence of the two) as shown in Appendix Figure 9 next to the Tsx nucleoside transporter.

Small hydrophobic transporters
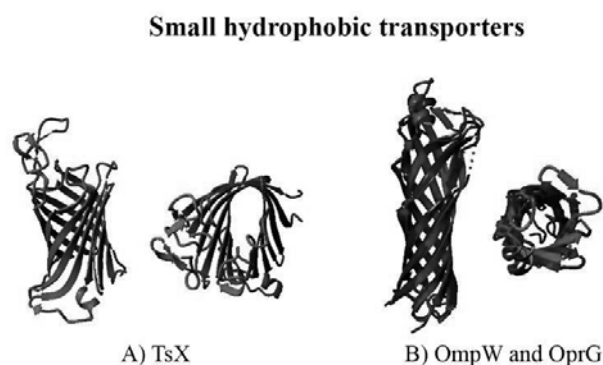
A) TsX   B) OmpW and OprG

Appendix Figure 9: Hydrophobic transporters. Panel A) Tsx nucleoside transporter (12 stranded β-barrel). Panel B) The 8-stranded OmpW and OprG structurally aligned with Mustang aligner. See Appendix Table 8 for PDB IDs.

3.1.3   Siderophores

The TonB-dependent structures consist of a 22 β-barrel structures. The structures contain plug domains essential for substrate transport. Since iron and heme are scarce they are difficult to obtain and the bacterium develops mechanisms to overcome these difficulties. TonB-dependent transporters contain plug domains that define their specificity. The sequence identity of TonB-dependent transporters showed yet another diverse protein family ranging from 8-38%. Over 500 different siderophores have been identified [7], but the ten[1] solved unique siderophore structures (Cir, BtuB, FauA, FecA, FepA, FhuA, FptA, FpvA, HasR, and ShuA) are shown in Appendix Figure 10 and pair-wise analyses results listed in Appendix Table 6.

Appendix Table 6: Siderophores. Mustang structure alignment results for siderophores of hydrophobic transporters, using same parameters as described for Appendix Table 1. See Appendix Table 8 for a complete list of protein details, including PDB IDs. See Appendix Table 8 for a complete list of protein details, including PDB IDs.

Object 1	Object 2	RMSD (Å)	% Aligned protein	Sequence identity (%)
Cir	BtuB	1.43	79.49%	26.71%
Cir	FauA	1.75	71.68%	17.32%
Cir	FecA	1.46	71.24%	21.36%
Cir	FepA	1.34	81.44%	37.78%
Cir	FhuA	1.75	70.90%	22.17%
Cir	FptA	1.74	70.57%	19.91%
Cir	FpvA	1.57	71.74%	22.14%
Cir	HasR	1.67	69.73%	20.38%
Cir	ShuA	1.71	79.10%	26.43%
BtuB	FauA	1.65	74.41%	18.29%
BtuB	FecA	1.69	76.04%	20.53%
BtuB	FepA	1.57	75.50%	22.84%
BtuB	FhuA	1.53	78.77%	21.66%
BtuB	FptA	1.54	77.13%	19.06%
BtuB	FpvA	1.55	80.22%	20.36%

---

[1] There exist 16 solved structures today, but these analyses were carried out in 2015

BtuB	HasR	1.66	72.60%	20.25%
BtuB	ShuA	1.54	78.40%	20.83%
FauA	FecA	1.77	70.98%	19.46%
FauA	FepA	1.91	65.21%	19.03%
FauA	FhuA	1.45	78.85%	23.73%
FauA	FptA	1.21	83.39%	37.32%
FauA	FpvA	1.26	85.66%	42.45%
FauA	HasR	1.73	68.01%	18.77%
FauA	ShuA	1.73	70.80%	18.77%
FecA	FepA	1.68	66.87%	19.00%
FecA	FhuA	1.65	72.62%	18.33%
FepA	FhuA	1.92	59.71%	17.73%
FepA	FptA	1.88	61.07%	19.00%
FepA	FpvA	1.82	61.91%	19.95%
FepA	HasR	1.68	63.09%	20.75%
FepA	ShuA	1.58	73.27%	19.78%
FhuA	FptA	1.49	78.63%	24.66%
FhuA	FpvA	1.50	72.70%	23.74%
FhuA	HasR	2.36	43.00%	17.11%
FhuA	ShuA	1.73	70.85%	19.09%
FptA	FpvA	1.31	88.55%	37.59%
FptA	HasR	1.74	64.43%	17.06%
FptA	ShuA	1.65	69.40%	17.87%
FpvA	HasR	1.71	56.71%	18.74%
FpvA	ShuA	1.61	71.50%	19.37%
HasR	ShuA	1.61	79.55%	27.53%
Average		*1.64*	*72.09%*	*22.31%*

Siderophores

A) All structure aligned

B) Cir

C) BtuB

D) FauA

E) FecA

F) FepA

G) FhuA

H) FptA

I) FpvA

J) HasR

K) ShuA
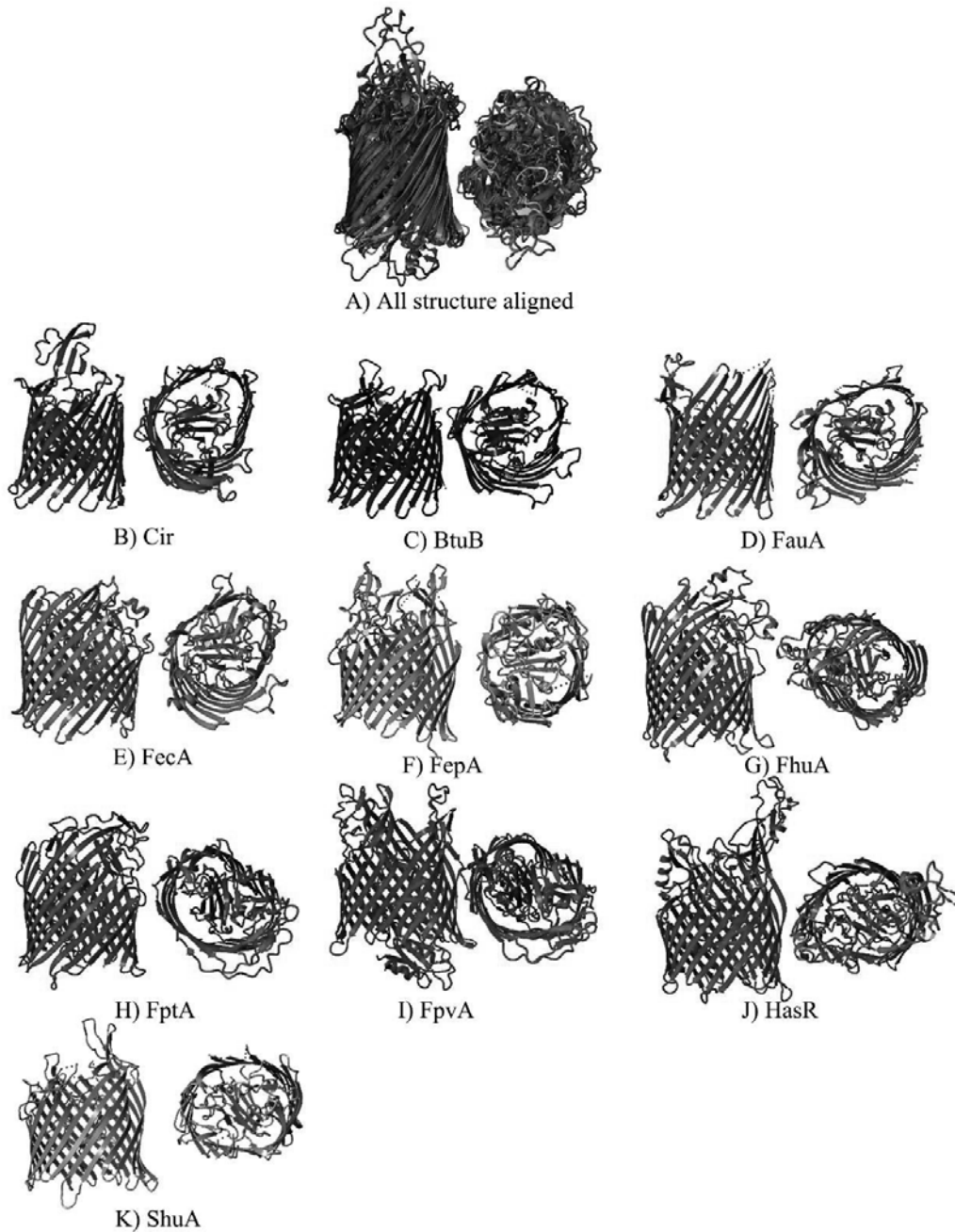
Appendix Figure 10: Siderophores. *Structural alignment of the TonB-dependent transporters. Cir colored purple, BtuB colored blue, FauA colored turquoise, FecA colored light green, FepA colored yellow, FhuA colored orange, FptA colored red, FpvA colored dark pink, HasR colored light pink, and ShuA colored grey (PDB ID: 2HDI, 2GUF, 3EFM, 1KMO, 1FEP, 1QFG, 1XKW, 2W16, 3CSC and 3FHH).*
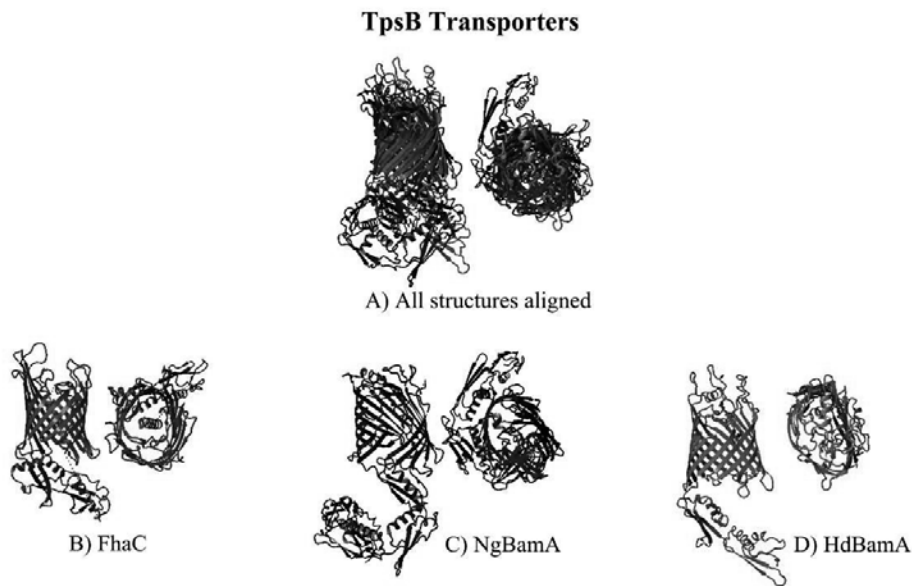
3.1.4 TpsB transporters

TbpA has the POTRA domain, but is a transferrin receptor with the largest size and therefore not included in the alignment with the 16 β-barrel structures (see Appendix Figure 11). The structure alignments of those with the same size do show quite large differences (with an average of 2.29 Å RMSD with 18.63% sequence identity for the nearly 20% aligned structure; see Appendix Table 7 and Appendix Figure 12). Looking at the structures, the large discrepancies are likely due to the periplasmic and plug domain.

TpsB transporter size differences

A) FhaC   B) TbpA

Appendix Figure 11: Size differences in TpsB transporters. FhaC colored purple and TbpA colored red. This figure illustrates the size difference found in this subfamily of TpsB transporters.

Appendix Table 7: TpsB transporters. Mustang structure alignment results for 16 β-stranded TpsB transporters, using same parameters as described for Appendix Table 1. See Appendix Table 8 for a complete list of protein details, including PDB IDs.

Object 1	Object 2	RMSD (Å)	Aligned (%)	Sequence identity (%)
FhaC	NgBamA	1.86	22.09%	12.28%
FhaC	HdBamA	2.35	29.46%	19.08%
NgBamA	HdBamA	2.66	9.96%	24.53%
Average		*2.29*	*20.50%*	*18.63%*

TpsB Transporters

A) All structures aligned

B) FhaC

C) NgBamA

D) HdBamA

Appendix Figure 12: TpsB transporters. *Structure alignment of the three 16 β-stranded TpsB transporters. FhaC colored purple, NgBamA colored blue and HdBamA colored green.* See Appendix Table 8 for PDB IDs.

Ushers

The solved structures, FimD and PapC, are both isolated from *E. coli*. The overall structure is conserved, yet quite high sequence variation between the structures is observed. The structural alignment between PapC and FimD had 1.65 Å RMSD with 32.74% sequence identity 90,37% of the protein PapC aligned (see Appendix Figure 13).

Ushers

A) Side-view

A) Top-view
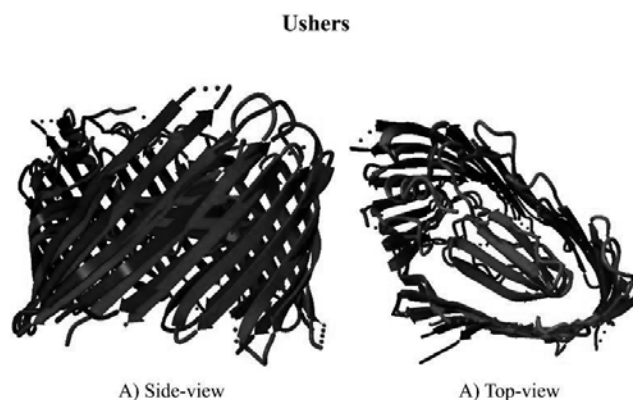
Appendix Figure 13: Ushers. *Structure alignment of the Usher structures. These large 24 stranded barrels contain a plug domain in the core of the structure. See Appendix Table 8 for PDB IDs.*

Adherence/Virulence Factors

Many OMPs have virulence and adherence properties (as discussed above), but this family are usually rather small-sized with long loops needed to adherence properties. Despite the long loops found in OmpA, we included this protein in the porin subfamily due to recent discovery of pore activity (see porin section 3.2 for more details). Intimin/invasin, Ail and PorB are found in this class, see Appendix Figure 14. The above structural alignment between Intimin and Invasin are very conserved (0.69 Å RMSD over 234 aligned residues with 51.28% sequence identity of the 96.3% aligned Invasin structure).

Appendix Figure 14:The Adherence/virulence Factors. This subfamily includes from left to right: Ail colored red, Intimin colored purple, Invasin colored blue and PorB (PDB ID: 3QRA, 4EIT/4FEIS, and 3VY8). Intimin and Invasin are structural aligned using Mustang aligner.

4   Protein structure details

Appendix Table 8: Structure details. This table holds structure information from all OMP used in this thesis. The structures highlighted in red was collected after the structure analyzes in 2016. Structures can be downloaded from the RCSB PDB website [www.rcsb.org].

Name	OMP subfamily	Bacteria (source organism)	Size (# of β-strands)	PDB ID	Resolution (Å)
CusC	Efflux pump	*Escherichia coli*	12	4K7R	2.09
TolC	Efflux pump	*Escherichia coli*	12	1EK9	2.10
OprM	Efflux pump	*Pseudomonas aeruginosa*	12	3D5K	2.40
VceC	Efflux pump	*Vibrio Cholerae*	12	1YC9	1.80
OprN	Efflux pumps	*Pseudomonas aeruginosa*	4x3	5AZO	2.70
CmeC	Efflux pumps	*Campylobacter jejuni*	4x3	4MT4	2.37
PagP	Enzyme	*Escherichia coli*	8	3GP6	1.40
PagL	Enzyme	*Pseudomonas aeruginosa*	8	2ERV	2.00
OmpT	Enzyme	*Escherichia coli*	10	1I78	2.60
Pla	Enzyme	*Yersinia pestis*	10	2X55	1.85
OMPLA	Enzyme	*Escherichia coli*	12	1QD6	2.10
LpxR	Enzyme	*Salmonella typhimurium*	12	3FID	1.90
RPA1785	Other	*Rhodopseudomonas palustris*	8	3C5O	2.20
TtoA	Other	*Thermus thermophilus*	8	3DZM	2.80
UPF0311	Other	*Clostridium acetobutylicum*	10	3G7G	1.99
OmpG	Porin (Class 3)	*Escherichia coli*	14	2IWV	2.3
PflBenF	Porin (Class 8)	*Pseudomonas fluorescens pf-5*	18	3JTY	2.58
OccD1/OprD	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SY7	2.15
OccD2/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SY9	2.8
OccD3/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SYB	2.7
OccK1/OpdK	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SYS	1.65
OccK2/OprF	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SZD	2.31
OccK3/OpdK	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3SZV	1.45
OccK4/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3T0S	2.2
OccK5/OpdK	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3T20	2.6
OccK6/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	3T24	2.4
OccK7/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	4FRT	3.17
OccK8/OprE	Porin (Class 8)	*Pseudomonas aeruginosa*	18	4FRX	1.9
OccK9/OprG	Porin (Class 8)	*Pseudomonas aeruginosa*	18	4FT6	2.6
OccK10/OpdC	Porin (Class 8)	*Pseudomonas aeruginosa*	18	4FSO	2.75
OccK11/OpdR	Porin (Class 8)	*Pseudomonas aeruginosa*	18	4FSP	2.32
EcOmpA	Porin (Class 1)	*Escherichia coli*	8	1QJP	1.65
KpOmpA	Porin (Class 1)	*Klebsiella pneumoniae*	8	2K0L	NA
KdgM	Porin (Class 2)	*Dickeya dadantii*	12	4FQE	1.93
NanC	Porin (Class 2)	*Escherichia coli*	12	2WJR	1.8

Omp32	Porin (Class 4)	*Delftia acidovorans*	16	2FGQ	1.45
EcOmpC	Porin (Class 4)	*Escherichia coli*	16	2J1N	2.00
EcOmpF	Porin (Class 4)	*Escherichia coli*	16	4GCS	1.87
PhoE	Porin (Class 4)	*Escherichia coli*	16	1PHO	3.00
OmpK36	Porin (Class 4)	*Klebsiella pneumoniae*	16	1OSM	3.20
Porin	Porin (Class 4)	*Rhodobacter capsulatus*	16	2POR	1.80
StOmpC	Porin (Class 4)	*Salmonella typhimurium*	16	3UPG	3.20
StOmpF	Porin (Class 4)	*Salmonella typhimurium*	16	3NSG	2.79
OprB	Porin (Class 5)	*Pseudomonas putida*	16	4GF4	3.1
OprP	Porin (Class 6)	*Pseudomonas aeruginosa*	16	2O4V	1.94
EcMaltoporin	Porin (Class7)	*Escherichia coli*	18	1AF6	2.4
StMaltoporin	Porin (Class7)	*Salmonella typhimurium*	18	2MPR	2.4
ScrY	Porin (Class7)	*Salmonella Typhimurium*	18	1OH2	2.4
CarO2	Porin	*Acinetobacter baumannii*	8	4RLB	2.70
OmpE36	Porin	*Enterobacter cloacae*	14	5FVN	1.45
CymA	Porin	*Klebsiella oxytoca*	14	4D51	2.30
PgaA Secretin	Porin	*Escherichia coli*	16	4Y25	2.82
OprO	Porin	*Pseudomonas aeruginosa*	16	4RJX	1.54
OccAB1	Porin	*Acinetobacter baumannii*	18	5DL5	2.05
OccAB2	Porin	*Acinetobacter baumannii*	18	5DL6	2.90
OCCAB3	Porin	*Acinetobacter baumannii*	18	5DL7	1.75
OCCAB4	Porin	*Acinetobacter baumannii*	18	5DL8	2.20
CsgG	Porin	*Escherichia coli*	36	3X2R	2.90
HpuA	Siderophore	*Neisseria gonorrhoeae*	8	5EE2	1.95
AbPirA	Siderophore	*Acinetobacter baumannii*	22	5FR8	2.83
ZnuD	Siderophore	*Neisseria meningitidis*	22	4RDR	2.47
FusA	Siderophore	*Pectobacterium atrosepticum*	22	4ZGV	3.20
PiuA	Siderophore	*Pseudomonas aeruginosa*	22	5FOK	1.90
PirA	Siderophore	*Pseudomonas aeruginosa*	22	5FP2	2.97
EcBamA	TpsB transporter	*Escherichia coli*	16	5LJO	4.90
OmpW	Transporter	*Escherichia coli*	8	2F1V	2.70
OprG	Transporter	*Pseudomonas aeruginosa*	8	2X27	2.40
Hia	Transporter	*Escherichia coli*	12	2GR8	2.00
EspP	Transporter	*Escherichia coli*	12	2QOM	2.66
Hbp	Transporter	*Escherichia coli*	12	3AEH	2.00
Tsx	Transporter	*Escherichia coli*	12	1TLY	3.01
NalP	Transporter	*Neisseria meningitidis*	12	1UYN	2.60
EstA	Transporter	*Pseudomonas aeruginosa*	12	3KVN	2.50
EcFadL	Transporter	*Escherichia coli*	14	1T16	2.60
PaFadL	Transporter	*Pseudomonas aeruginosa*	14	3DWO	2.20
TodX	Transporter	*Pseudomonas putida*	14	3BS0	2.60
TbuX	Transporter	*Ralstonia pickettii*	14	3BRY	3.20
FhaC	Transporter	*Bordetella pertussis*	16	2QDZ	3.15
HdBamA	Transporter	*Haemophilus ducreyi*	16	4K3C	2.91

NgBamA	Transporter	*Neisseria gonorrhoeae*	16	4K3B	3.20
AlgE	Transporter	*Pseudomonas aeruginosa*	18	4AFK	NA
FauA	Transporter	*Bordetella pertussis*	22	3EFM	2.33
Cir	Transporter	*Escherichia coli*	22	2HDI	2.50
BtuB	Transporter	*Escherichia coli*	22	2GUF	1.95
FecA	Transporter	*Escherichia coli*	22	1KMO	2.00
FhuA	Transporter	*Escherichia coli*	22	1QFG	2.50
FepA	Transporter	*Escherichia coli*	22	1FEP	2.40
TbpA	Transporter	*Neisseria meningitidis*	22	3V8X	2.60
FpvA	Transporter	*Pseudomonas aeruginosa*	22	2W16	2.71
FptA	Transporter	*Pseudomonas aeruginosa*	22	1XKW	2.00
HasR	Transporter	*Serratia marcescens*	22	3CSL	2.70
ShuA	Transporter	*Shigella dystentreiae*	22	3FHH	2.60
COG4313	Transporter	*Pseudomonas putida*	12	4RL8	2.30
SusC	Transporter	*Bacteroides thetaiotaomicron*	22	5FQ6	2.80
FimD	Usher	*Escherichia coli*	24	3RFZ	2.80
PapC	Usher	*Escherichia coli*	24	3FIP	3.15
OmpX	Virulence Factor	*Escherichia coli*	8	1QJ8	1.90
NspA	Virulence Factor	*Neisseria meningitidis*	8	1P4T	2.55
Ail	Virulence Factor	*Yersinia pestis*	8	3QRA	1.80
OpcA	Virulence Factor	*Neisseria meningitidis*	10	2VDF	1.95
Invasin	Virulence Factor	*Escherichia coli*	12	4E1S	1.86
Intimin	Virulence Factor	*Yersinia pseudotuberculosis*	12	4E1T	2.26
Wzi	Virulence Factor	*Escherichia coli*	18	2YNK	2.64
PorB	Virulence Factor / Transporter	*Neisseria meningitidis*	16	3VY8	2.30
Opa60	Virulence/ Niche Factor	*Neisseria gonorrhoeae*	8	2MAF	NA
Tp0751	Virulence/ Niche Factor	*Treponema pallidum*	8	5JK2	2.15
TamA	Virulence/ Niche Factor	*Escherichia coli*	16	4N74	2.90

5 References

1.   Vriend, G. Cmbi courses. http://swift.cmbi.ru.nl/teach/courses/index.html
2.   Branden, C.; Tooze, J. *Introduction to protein structure*. Garland Pub: New York, USA, 1991.
3.   Krieger, E.; Koraimann, G.; Vriend, G. Increasing the precision of comparative models with yasara nova--a self-parameterizing force field. *Proteins* **2002**, *47*, 393-402.
4.   Bishop, R.E. Structural biology of membrane-intrinsic beta-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim Biophys Acta* **2008**, *1778*, 1881-1896.
5.   Schulz, G.E. The structure of bacterial outer membrane proteins. *Biochim Biophys Acta* **2002**, *1565*, 308-317.
6.   Li, X.; Gu, Y.; Dong, H.; Wang, W.; Dong, C. Trapped lipopolysaccharide and lptd intermediates reveal lipopolysaccharide translocation steps across the escherichia coli outer membrane. *Scientific reports* **2015**, *5*, 11883.
7.   Schalk, I.J.; Mislin, G.L.; Brillet, K. Structure, function and binding selectivity and stereoselectivity of siderophore-iron outer membrane transporters. *Current topics in membranes* **2012**, *69*, 37-66.