REDUCTIONIST AND INTEGRATIVE APPROACHES TO EXPLORE THE *H.PYLORI* GENOME

Salvatore Loguercio

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Biology
Padova University

January 2008

Abstract

L'approccio riduzionistico teso a decomporre gli enti biologici in una collezione di parti ed elementi ha dominato la biologia molecolare per almeno cinquant'anni. Poiche' gli organismi si compongono soltanto di atomi e molecole senza l'intervento di forze esterne, si e' assunto che fosse possibile spiegare i sistemi biologici in termini delle proprieta' chimico-fisiche dei rispettivi elementi costituenti, fino al livello atomico. Tuttavia, nonostante il notevole successo riscosso dal metodo riduzionista nell' analisi dei singoli componenti cellulari, e' oggi ampiamente accettata l'idea che il comportamento dei sistemi biologici complessi non possa essere desunto soltanto dallo studio delle singole parti costituenti. Per gestire la complessita' inerente alla comprensione di reti estese di biomolecole interagenti, l'approccio integrativo enfatizza metodi derivati dalla cibernetica e dalla teoria dei sistemi, usando una combinazione di matematica, metodi computazionali ed osservazione empirica. Tale approccio comincia ad essere applicabile nei sistemi procariotici, per i quali e' disponibile un informazione completa del genoma e del trascrittoma parallelamente ad un'immagine ragionevolmente estesa del proteoma.

I batteri patogeni rientrano indubbiamente tra i procarioti piu' attivamente studiati. Un esempio e' costituito dal patogeno *H.pylori*, agente riconosciuto di molteplici e gravi patologie gastroduodenali, la cui infezione coinvolge all'incirca meta' della popolazione mondiale.

Obbiettivo del presente lavoro di tesi e' stato lo studio di molteplici aspetti connessi alla fisiologia molecolare di *H.pylori*, utilizzando entrambe gli approcci sopra enunciati- riduzionistico ed integrativo.

Nella Sezione I, abbiamo utilizzato una prospettiva riduzionistica nello studio dello stato di ossidazione delle Cisteine e nell'attribuzione dei ponti disolfuro di una peculiare chaperonina espressa da *H.pylori*, Heat Shock protein A (HspA) appartenente alla classe delle proteine GroES. Tale proteina presenta un elevato numero di Cisteine, e' in grado di legare il Nickel ed e' presente sia nel citoplasma che sulla superficie cellulare del batterio. Nel contesto del presente lavoro, noi abbiamo prodotto e caratterizzato la proteina

ricombinante HspA nonche' i mutanti Cys94Ala e C94A/C111A. La topologia dei ponti disolfuro e' stata determinata integrando metodologie biochimiche con analisi di spettrometria di massa. Ne e' emerso che, nella forma ossidata della proteina, tutte le Cys sono coinvolte in ponti disolfuro. In particolare, nel, i legami disolfuro costringono l'estremita' C-terminale ad assumere una peculiare struttura chiusa, adatta ad ospitare ioni Nickel. Tale nuovo arrangiamento strutturale per il legame del Nickel puo' essere messo in relazione al processo di cattura e rilascio del metallo nei riguardi dell' ureasi extracellulare, essenziale alla sopravvivenza del batterio nella mucosa gastrica.

Nella Sezione II, abbiamo utilizzato differenti metodi computazionali per studiare la rete di interazioni biomolecolari associata a *H.pylori.* L'obiettivo principale e' stato quello di individuare nuovi potenziali obiettivi per lo sviluppo di farmaci contro la proliferazione del batterio (Capitoli 4 e 5).

Il Capitolo 4 esplora la 'robusta ma fragile' natura della cellula di *H.pylori*, intesa come un sistema complesso nel quale la robustezza nei confronti di una data perturbazione e' inevitabilmente associata alla fragilita' nei confronti di altre perturbazioni. In considerazione di cio', abbiamo sviluppato una strategia generale allo scopo di identificare i nodi di controllo nella rete metabolica di un batterio, potendo tali nodi costituire potenziali obiettivi per lo sviluppo di nuovi farmaci. Tale metodologia e' implementata su Helicobacter pylori 26695.

L'intera rete metabolica del patogeno e' analizzata al fine di individuare punti critici da un punto di vista biochimico, ovvero enzimi in grado di produrre o consumare un unico metabolita. Una volta identificata, la lista degli enzimi critici e' filtrata allo scopo di isolare potenziali candidati non omologhi ad enzimi umani. Al termine, l'essenzialita' degli enzimi selezionati e' sottoposta ad una ulteriore validazione attraverso studi di delezione *in silico* utilizzando una tecnica di simulazione metabolica nota come Analisi del Bilancio di Flusso (Flux-Balance Analysis, FBA). La simulazione e' condotta su un recente modello metabolico siu scala genomica di *H.pylori*. Seguendo tale approccio, abbiamo identificato alcuni enzimi di interesse per studi di inibizione relativi all'infezione da *H.pylori*.

Lo studio riportato nel Capitolo 5 amplia l'approccio appena descritto alla luce di recenti studi teorici sulle reti biologiche. Tali studi suggeriscono che attacchi multipli di lieve entita' diretti ad obiettivi selezionati sono inevitabilmente piu'efficaci della soppressione di un singolo obiettivo. Tale osservazione e' consistente con la recente efficacia riscontrata nell'utilizzo dei farmaci multicomponente. Alla luce di tali considerazioni, abbiamo esplorato la robustezza metabolica di *H.pylori* simulando multiple inattivazioni parziali di enzimi selezionati, al fine di individuare insiemi-obiettivo per lo sviluppo di terapie combinatoriali.

I dati metabolici nonche' di interazione tra proteine disponibili nelle banche dati sono stati utilizzati per la costruzione di un network biomolecolare integrato del batterio patogeno, che e' stato successivamente analizzato allo scopo di identificare elementi centrali per l'architettura e la comunicazione del network stesso. Gli enzimi selezionati sono stati di seguito classificati sulla base dei dati biologici disponibili in un tentativo di predire eventuali combinazioni letali per il batterio. Allo scopo di valutare l'effetto di rete' innescato dalla parziale inattivazione degli enzimi obiettivo, un'analisi di robustezza e' stata effettuata su piccoli gruppi di targets avvalendosi nuovamente di simulazioni metaboliche basate sull' Analisi del Bilancio di Flusso (FBA). Le simulazioni sono state condotte su una recente ricostruzione metabolica di *Helicobacter pylori* su scala genomica ed hanno consentito di predire il fenotipo di crescita associato ad ogni particolare schema di inattivazione parziale. I risultati preliminari riportati in questo contesto possono essere utilizzati per ridurre l'insieme iniziale degli enzimi-obiettivo per la messa a punto di farmaci/terapie multicomponente contro l'infezione da *H.pylori*. In ultimo, tali dati forniscono alcune indicazioni concrete su come attacchi multipli di lieve entita' possono indurre perturbazioni significative in sistemi biologici complessi.

Abstract

The reductionist approach of decomposing biological systems into their constituent parts has dominated molecular biology for half a century. Since organisms are composed solely of atoms and molecules without the participation of extraneous forces, it has been assumed that it should be possible to explain biological systems on the basis of the physico–chemical properties of their individual components, down to the atomic level. However, despite the remarkable success of methodological reductionism in analyzing individual cellular components, it is now generally accepted that the behavior of complex biological systems cannot be understood by studying their individual parts in isolation. To tackle the complexity inherent in understanding large networks of interacting biomolecules, the integrative viewpoint emphasizes cybernetic and systems theoretical methods, using a combination of mathematics, computation and empirical observation. Such an approach is beginning to become feasible in prokaryotes, combining an almost complete view of the genome and transcriptome with a reasonably extensive picture of the proteome.

Pathogenic bacteria are undoubtedly the most investigated subjects among prokaryotes. A paradigmatic example is the the human pathogen *H.pylori*, a causative agent of severe gastroduodenal disorders that infects almost half of the world population.

In this thesis, we investigated various aspects of Helicobacter pylori molecular physiology using both reductionist and integrative approaches.

In Section I, we have employed a reductionist, bottom-up perspective in studying the Cysteine oxidised/reduced state and the disulphide bridge pattern of an unusual GroES homolog expressed by H.pylori, Heat Shock protein A (HspA). This protein possesses a high Cys content, is involved in nickel binding and exhibits an extended subcellular localization, ranging from cytoplasm to cell surface. We have produced and characterized a recombinant HspA and mutants Cys94Ala and C94A/C111A. The disulphide bridge pattern has been assigned by integrating biochemical methodologies with mass spectrometry. All Cys are engaged in disulphide bonds that force the C-term domain to

assume a peculiar closed loop structure, prone to host nickel ions. This novel Ni binding structural arrangement can be related to the Ni uptake/delivery to the extracellular urease, essential for the bacterium survival.

In Section II, we combined different computational methods with two main goals:

1) Analyze the *H.pylori* biomolecular interaction network in an attempt to select new molecular targets against H.pylori infection (Chapters 4 & 5);

2) Model and simulate the signaling perturbations induced by invading H.pylori proteins in the host ephitelial cells (Chapter 6).

Chapter 4 explores the 'robust yet fragile' feature of the *H.pylori* cell, viewed as a complex system in which robustness in response to certain perturbation is inevitably associated with fragility in response to other perturbations. With this in mind, we developed a general strategy aimed at identify control points in bacterial metabolic networks, which could be targets for novel drugs. The methodology is implemented on Helicobacter pylori 26695.

The entire metabolic network of the pathogen is analyzed to find biochemically critical points, e.g. enzymes which uniquely consume and/or produce a certain metabolite. Once identified, the list of critical enzymes is filtered in order to find candidate targets wich are non-homologous with the human enzymes. Finally, the essentiality of the identified targets is cross-validated by in silico deletion studies using flux-balance analysis (FBA) on a recent genome-scale metabolic model of H. pylori. Following this approach, we identified some enzymes which could be interesting targets for inhibition studies of *H.pylori* infection.

The study reported in Chapter 5 extends the previously described approach in light of recent theoretical studies on biological networks. These studies suggested that multiple weak attacks on selected targets are inevitably more efficient than the knockout of a single target, thus providing a conceptual framework for the recent success of multi-target drugs. We used this concept to exploit *H.pylori* metabolic robustness through multiple weak attacks on selected enzymes, therefore directing us toward target-sets discovery for combinatorial therapies.

We used the known metabolic and protein interaction data to build an integrated biomolecular network of the pathogen. The network was subsequently screened to find central elements of network communication, e.g. hubs, bridges with high betweenness centrality and overlaps of network communities. The selected enzymes were then classified on the basis of available data about cellular function and essentiality in an attempt to predict successful target-combinations. In order to evaluate the network effect triggered by the partial inactivation of candidate targets, robustness analysis was performed on small groups of selected enzymes using flux balance analysis (FBA) on a recent genome-scale metabolic model of H.pylori. In particular, the FBA simulation framework allowed to predict the growth phenotype associated to every partial inactivation set.

The preliminary results obtained so far may help to restrict the initial target-pool in search of target-sets for novel combinatorial drugs against *H.pylori* persistence. However, our long-term goal is to better understand the indirect network effects that lie at the heart of multi-target drug action and, ultimately, how multiple weak hits can perturb complex biological systems.

H.pylori produces various a cytotoxic protein, CagA, that interfere with a very important host signaling pathway, i.e. the epidermal growth factor receptor (EGFR) signaling network. EGFR signaling is one of the most extensively studied areas of signal transduction, since it regulates growth, survival, proliferation and differentiation in mammalian cells. In Chapter 6, we attempted to build an executable model of the EGFR-signaling core process using a process algebra approach. In the EGFR network, the core process is the heart of its underlying hour-glass architecture[1], as it plays a central role in downstream signaling cascades to gene expression through activation of multiple transcription factors. It consists in a dense array of molecules and interactions wich are tightly coupled to each other.

[1] The overall architecture of the EGFR network is a bow-tie (or hourglass) structure, a hallmark of robust evolvable systems that emerges at various scales in complex biological phenomena like metabolism, signal transduction, transcription and translation. A typical bow-tie architecture comprises a variety of redundant inputs and outputs that are connected by a conserved core process.

In order to build the executable model, a small set of EGFR core molecules and their interactions is tentatively translated in a BetaWB model. BetaWB is a framework for modelling and simulating biological processes based on Beta-binders language and its stochastic extension.

Once obtained, the computational model of the EGFR core process can be used to test and compare hypotheses regarding the principles of operation of the signaling network, i.e. how the EGFR network generates different responses for each set of combinatorial stimuli. In particular, probabilistic model checking can be used to explore the states and possible state changes of the computational model, whereas stochastic simulation (corresponding to the execution of the BetaWB model) may give quantitative insights into the dynamic behaviour of the system in response to different stimuli. Information from the above tecniques allows model validation through comparison within the experimental data available in the literature.

The inherent compositionality of the process algebra modeling approach enables further expansion of the EGFR core model, as well as the study of its behavior under specific perturbations, such as invading *H.pylori* proteins. This latter aspect might be of great value for *H.pylori* pathogenesis research, as signaling through the EGF receptors is intricately involved in gastric cancer and in many other gastroduodenal diseases.

Contents

Chapter 1

Introduction

1.1 Biology in the XXI Century

Biological and biomedical research is undergoing revolutionary developments that have an immense and lasting impact on society. Chemistry, biology, and physics have revealed an immense amount of information on molecular structure and function, and now we are poised to make use of it for atomic-level engineering. New discoveries are being made every day, and clever people are pressing these discoveries into service in every imaginable (and unimaginable) way.

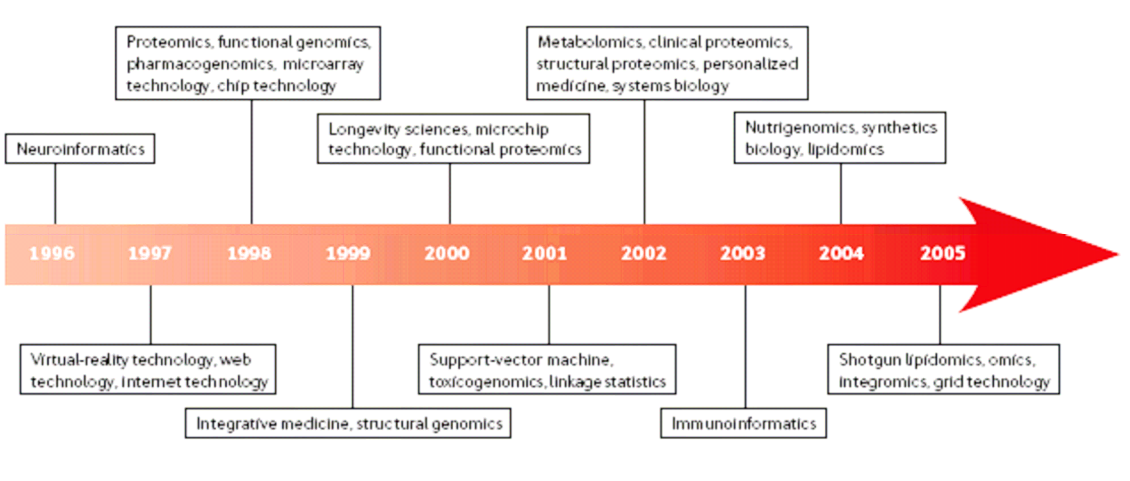


Figure 1.1: New areas and technologies in post-genomic research (Jensen et al., 2006)

The completion and publication of the *Haemophilus influenzae* genome sequence in 1995 (Fleischmann et al., 1995) marked a significant phase transition in the history of biological research. The advent of whole-genome sequencing and other high-throughput experimental technologies transformed biological research from a relatively data poor discipline into one that is data rich (*Figure* 1.1).

Biological information, which can be measured in terms of the numbers of articles and journals that are published, is increasing at a considerable rate, so that it is no longer possible for a researcher to keep up-todate with all the relevant literature manually, even on specialized topics (Jensen et al., 2006).

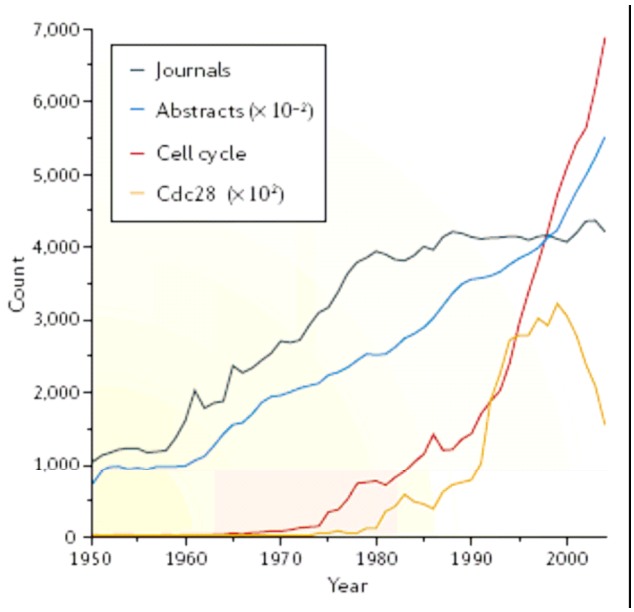


Figure 1.2: Growth of Medline.
The numbers of journals, papers (as represented by Medline abstracts), papers on the cell cycle and papers on Cdc28 that were published each year from 1950 to 2005 are shown. An average for 3 years was calculated for the Cdc28 curve because of much lower numbers. The number of new papers that were published each year continues to increase, especially on certain topics such as the cell cycle, for which it is no longer possible to read all new papers that are published. By contrast, specific proteins that are 'hot' at one point in time tend to lose their popularity later, as exemplified by Cdc28. (Jensen et al., 2006).

An important challenge that is faced by investigators today lies in interpreting large-scale 'omics' data sets and thereby deriving fundamental and applied biological information about whole systems.

The challenge that comes with information-rich environments is not unique to post-genomics biological research. Modern data sets in many disciplines are frequently immense in size. For example, National Oceanic and Atmospheric Administration (NOAA) satellites, which are used to monitor global climate change, generate approximately one terabyte of data per day (Ehrenman, 2005). Furthermore, the international retail giant Wal-Mart maintains a database of nearly 460 terabytes that contains product information and details on the customers who buy them (Hays, 2004). This onslaught of available information has driven the development of important data-mining techniques that are devoted to uncovering details that are of practical value for various applications (Hand et al., 2000).

Although their size is often on the same, or a smaller, scale compared with other modern data sets, significant challenges are unique to modern post-genomics data sets. For example, many technological platforms, both hardware and software, are available for several ‘omics’ data types, but some of these are prone to introducing technical artefacts (Kluger et al., 2003). This can bias the data, which can falsely expose sample differences in the absence of a biological cause. In addition, uniform, standardized data representations are not always adopted (Quackenbush, 2004), which complicates cross-experiment comparisons. Data-quality, context and lab-to-lab variations represent another important hurdle that must be overcome in genome-scale science (Bader & Hogue, 2002). Despite these challenges, however, investigators are making progress in identifying, extracting and interpreting biological insights from omics data sets.

Contemporary biology continues to obtain massive amounts of information about whole biological systems, via high-throughput experiments that provide relatively shallow and noisy data. The Human Genome Project is a prototypical example: the knowledge it accumulated is highly valuable, and was obtained in an automated and relatively efficient way, but is just the beginning of understanding the human genome. Similar effort are now underway in genomics (finding the collection of all genes, for many genomes), in

transcriptomics (the collection of all actively transcribed genes), in proteomics (the collection of all proteins), and in metabolomics (the collection of all metabolites). Bioinformatics is the rapidly growing discipline tasked with collecting and analyzing such omics data (*Figure* 1.3).

The main ambition of present day biology is to build, with such data, a science of the principles of operation of biological systems, based on the interactions between components.

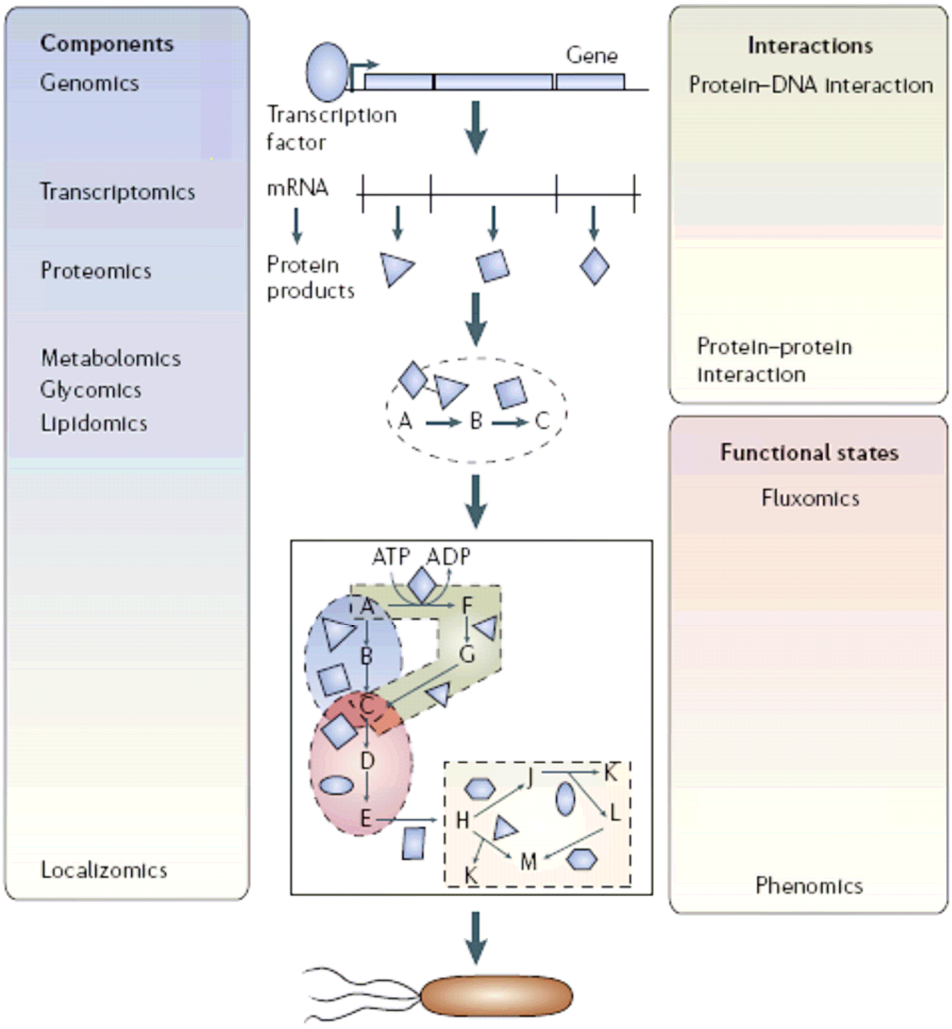


Figure 1.3: 'Omics' data are providing comprehensive descriptions of nearly all components and interactions within the cell.

Biological systems are obviously well-engineered: they are very complex and yet highly structured and robust. They have only one major engineering defect: they have not been designed, in any standard sense, and so are not laid out as to be easily understood. It is not clear that any of the engineering principles of operations we are currently familiar with are fully applicable. Understanding such principles will require an interdisciplinary effort, using ideas from physics, mathematics, and computing. These, then, are the promises of contemporary biology: it will teach us new principles of operation, likely applicable to other sciences, and it will leverage other sciences to teach us how cells work in an actionable way.

1.2 Complexity of living systems

Classically, an ideal scientific explanation reduces the apparent complexity of nature to a smooth plane of immutable natural laws (Popper, 2000). The living organisms, however, resist smooth explanations (Efroni & Cohen, 2002; Fleck, 1979). They obey the laws of physics and chemistry, but these basic laws do not explain their behaviour; each component part of a complex system participates in many different interactions and these interactions generate unforeseeable, emergent properties. For example, microscopic interactions between nonliving molecules, at the macroscopic level, produce a living cell.
Biology has done well in its plan to carve living systems into their component parts, but the reductionist program has not uncovered the fundamental laws from which we can deduce how a particular living system actually works. A cell cannot be reduced to a few formative principles, in the way, for example, the behaviour of the Solar System can be reduced to the laws of gravity and motion (Cohen, 2000).
Viewing the cell as a big integrated network of various molecular components offers a viable strategy for addressing the complexity of living systems. This new paradigm is summarized by the Life's Complexity pyramid, where the elementary building blocks of the cell —genes, RNAs, proteins, and metabolites— organize themselves into small recurrent patterns, called pathways in metabolism and motifs in genetic-regulatory networks. In turn, motifs and pathways are seamlessly integrated to form functional

modules —groups of nodes (for example, proteins and metabolites) that are responsible for discrete cellular functions (Hartwell et al., 1999). These modules are nested in a hierarchical fashion and define the cell's large-scale functional organization (Ravasz et al., 2002). (see the *Figure* 1.4).

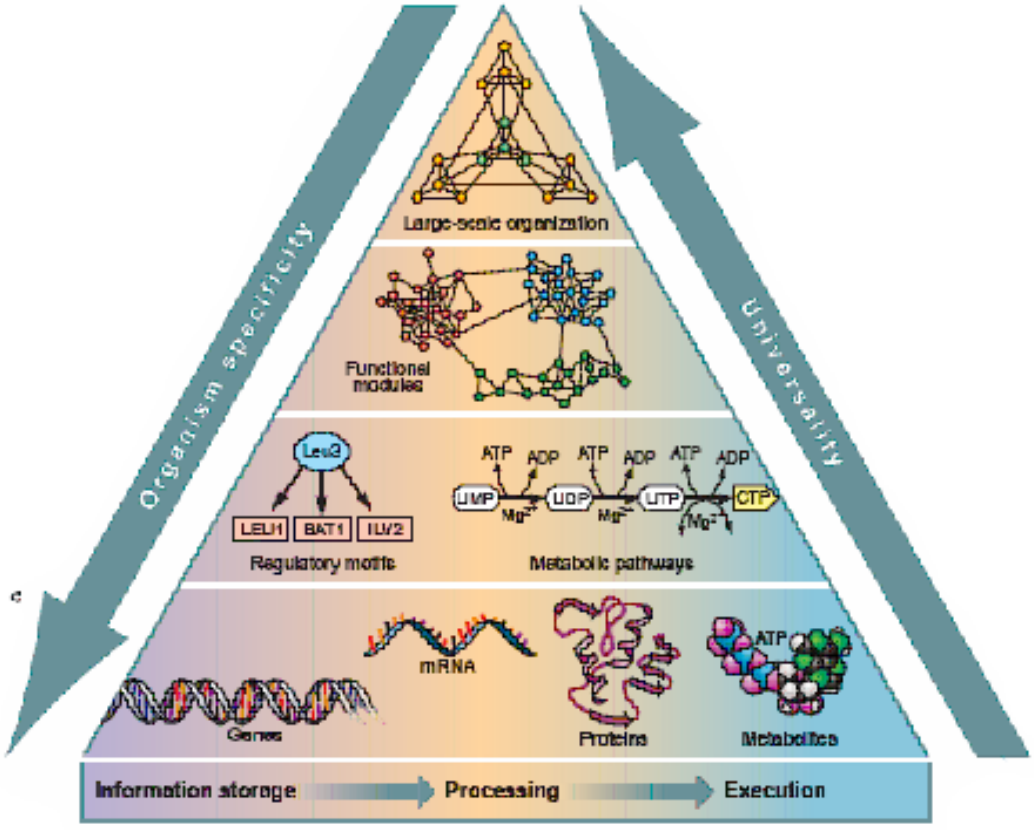


Figure 1.4: Life's Complexity pyramid

1.3 Reductionist and integrative approaches

Since the advent of the `molecular revolution' in biology, there has been an ongoing argument about the interrelation of *integrationism* and *reductionism* (*Figure* 1.5):

Integrationism represents the effort to comprehend biological function in its greater context, notably via recognition of the dynamic interactions between a system's components.

Reductionism, conversely, focuses on understanding a system's capacity via the detailed description of its constituent parts.

This leaves us with the question of whether or not the two directions are irreconcilable, or, to put it more positively, whether or not we may resolve the apparent controversy by developing a new framework for life sciences, a new concept, a new logic.

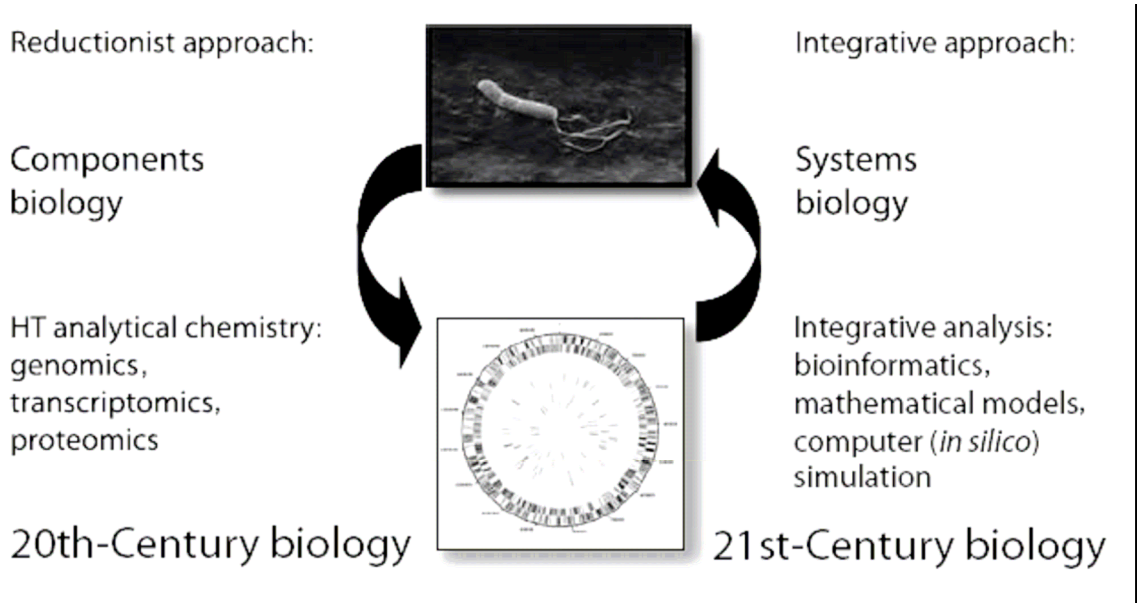


Figure 1.5: Reductionist and integrative approaches in biological research.

The logic of life will neither be recognized without precise understanding of the manifold of components that give rise to biological function, nor without a clear conception of the dynamic interactions between individual components on every level of functional integration. Likewise, the logic of life lies exclusively neither in the most incredible detail, nor in the most sweeping synopsis. Neither of the two - *integrationism* and *reductionism* - is self-sufficient, and both are obligatory to our quest for knowledge.

1.4 Motivation and purpose - Organization of the thesis

In this thesis we studied the Gram-negative bacterium Helicobacter pylori - a common human pathogen - using both reductionist (Section I) and integrative approaches (Section II).

In Section I, we pointed on the Cysteine oxidised/reduced state of an unusual GroES homolog expressed by H.pylori, Heat Shock protein A (HspA). This protein possesses a high Cys content, is involved in nickel binding and exhibits an extended subcellular localization, ranging from cytoplasm to cell surface. We have produced and characterized a recombinant HspA and mutants Cys94Ala and C94A/C111A. The disulphide bridge pattern has been assigned by integrating biochemical methodologies with mass spectrometry. All Cys are engaged in disulphide bonds that force the C-term domain to assume a peculiar closed loop structure, prone to host nickel ions. This novel Ni binding structural arrangement can be related to the Ni uptake/delivery to the extracellular urease, essential for the bacterium survival.

In Section II, we considered some applications of Systems Biology to antibacterial drug discovery, in an attempt to select new molecular targets against H.pylori infection.

These projects include:

- Critical point discovery in bacterial metabolic networks;
- Target-sets discovery for combinatorial therapies, using robustness analysis.

Chapter 2

Background: Helicobacter Pylori

2.1 Overview

In 1983, Warren and Marshall were the first to report the successful cultivation of the human pathogen *Helicobacter pylori* from gastric biopsy samples (Marshall & Warren, 1984; Warren JR, 1983). By self-ingestion experiments, they showed that this bacterium indeed caused gastroduodenal disorders, thereby fulfilling Koch's postulates (Marshall et al., 1985; Morris & Nicholson, 1987). This important discovery, rewarded with the 2005 Nobel prize for physiology and medicine, has changed peptic ulcer disease from a chronic, relapsing disease of uncertain cause into a curable infectious disease. Today, *H pylori* is accepted as the causative agent of acute and chronic gastritis, and a major predisposing factor for peptic ulcer disease, gastric carcinoma, and gastric lymphoma (Kusters et al., 2006; Suerbaum & Michetti, 2002) (*Figure* 2.1).

The human stomach is a unique ecological niche characterized by very acidic pH - a condition lethal for most microbes. *H. pylori* is so well adapted to this unfriendly environment that, after the first infection, which usually occurs early in life, it establishes a life-long chronic infection . The selection of a niche with no competition and the ability to establish a chronic infection make *H. pylori* one of the most successful human

bacterial parasites, which colonizes more than half of the human population (Montecucco & Rappuoli, 2001).

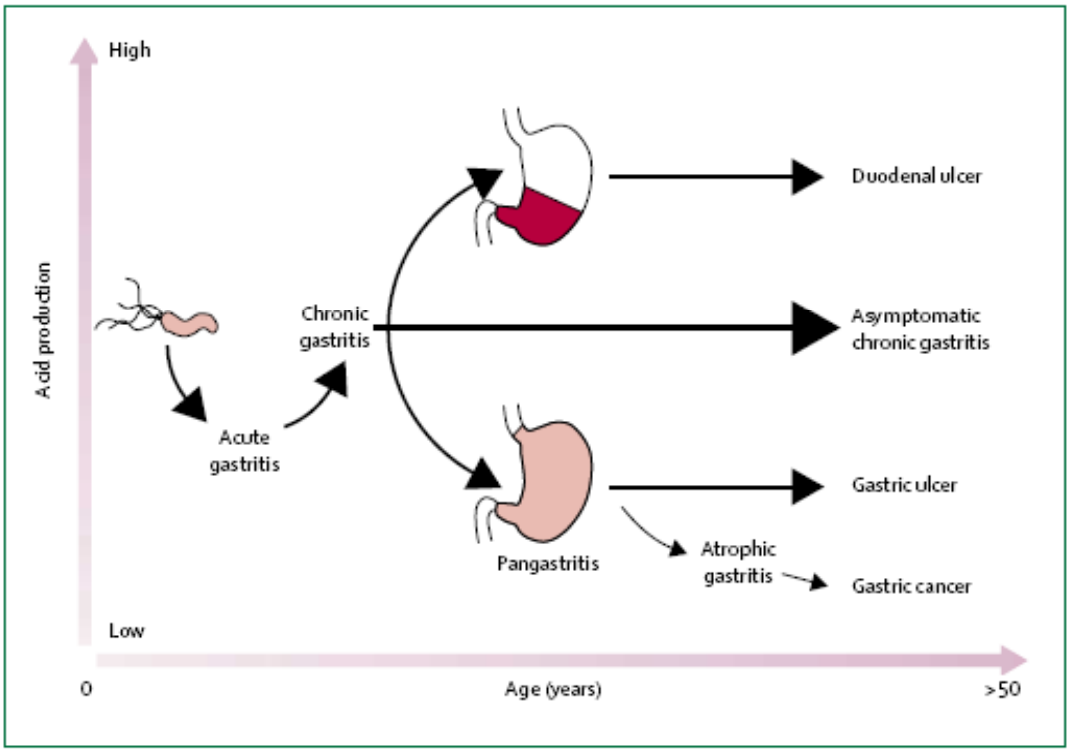


Figure 2.1: Schematic representation of the *H.pylori* infection.
Acquisition of *H.pylori* usually occurs during childhood. Once acquired and left untreated, the infection persists for life. After the acute phase, most *H.pylor*-positive patients develop a chronic gastritis without symptoms. In some patients, more severe manifestations will develop later in life. A normal or high acid secretion predisposes to duodenal ulcers, whereas a low acid secretion predisposes to gastric ulcers and gastric cancer.

The successful life-lasting colonization of the human stomach by *H. pylori* is achieved through a combination of factors, which address the different challenges presented by the harsh environment (Figure 2.2).

H. pylori synthesizes a urease to buffer the pH of its immediate surroundings within the stomach. Its helicoidal shape and the action of flagella allow it to cross the thick layer of mucus lining the stomach. *H. pylori* then binds to lewis antigens present on host gastric cells, and it secretes factors that attract and stimulate inflammatory cells, as well as the

multifunctional toxin VacA. Last, the presence of the *cag* pathogenicity island, a 40-kb DNA that encodes a type IV secretion system, seems to be necessary for optimal fitness of the bacterium and the appearance of pathogenic traits.

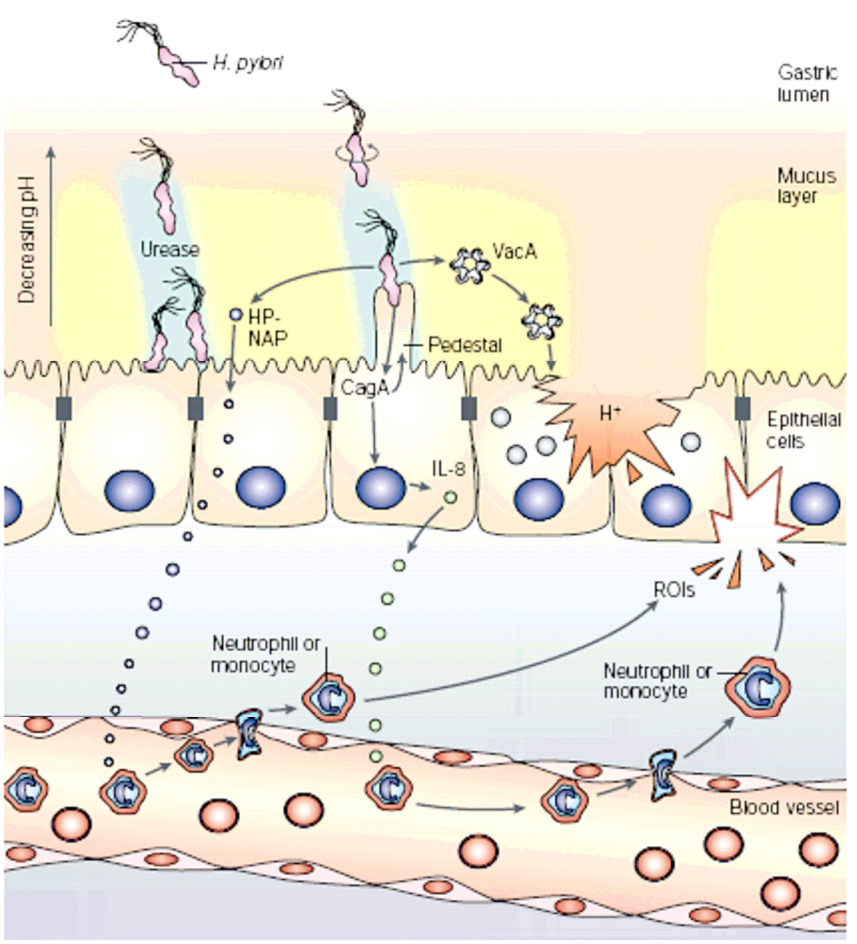


Figure 2.2: Schematic representation of the stomach mucosa colonized by *Helicobacter pylori,* showing the main virulence factors involved in colonization and disease.

Disease outcome is complex, because it depends on many factors, including bacterial genotype, host physiology, genotype and dietary habits. *H. pylori*-induced gastritis, during the decades that follow initial infection, can remain silent or evolve into more-severe diseases, such as atrophic gastritis, peptic ulcer, or lymphoma of the mucosa-associated lymphoid tissue (Romano et al., 2006). Epidemiological studies first indicated that colonization by *H. pylori* increased the risk of developing distal (noncardia) gastric

cancer; the earlier the age of acquisition of *H. pylori* infection, the higher the risk of developing cancer. *H. pylori*-infected subjects have at least a twofold increase in the risk of gastric cancer when compared with uninfected subjects (Huang et al., 1998).

The strong association between *H. pylori* infection and gastric cancer led the World Health Organization to classify *H. pylori* as a class 1 carcinogen (IARC, 1994).

Infection with *H pylori* occurs worldwide, but there are substantial differences in the prevalence of the infection both within and between countries. In industrialised countries, the overall carriage rate of *H pylori* infection in middle-aged adults is 20–50%, compared with 80% or more in many developing countries (Frenck & Clemens, 2003; Rothenbacher & Brenner, 2003).

Acquisition of *H. pylori* predominantly occurs during childhood, and once acquired, the infection persists throughout life unless specifically treated. In developing countries the carriage rate of *H. pylori* remains relatively stable, but in the industrialised world these values have substantially decreased over recent decades, probably as a result of improved hygiene and sanitation, especially during childhood, and active elimination of carriership via antimicrobial treatment (Gerrits et al., 2006).

H pylori-associated disorders usually regress or heal completely after successful treatment of *H pylori* infection with antimicrobials. However, the available antimicrobial therapies for *H pylori* infection have many shortcomings, e.g. side-effects, the need for combination therapy, and limited efficacy, in particular because of the development of antimicrobial resistance. The continuous increase in the prevalence of antimicrobial resistance in *H.pylori*, together with the lack of forthcoming novel treatment options, already negatively affects eradication of *H pylori* infection, and is predicted to lead to serious problems for treatment of *H pylori*-associated disorders in the near future.

2.2 Bacteriology

Helicobacter pylori is a micro-aerophilic, Gram-negative, slow-growing, spiral-shaped and flagellated organism. It is currently classified in the Proteobacteria, a large, diverse division of Gram-negative bacteria which includes, among others, *H. influenzae* and *E.*

coli. The bacterium is a member of a rapidly growing genus. New species are being isolated at a fast rate from many vertebrate hosts. Also, other *Helicobacter* species are being isolated from nongastric sites in humans and may be implicated in diseases that previously had no assigned etiologic agent. *H. pylori* is motile via a tuft of polar-sheathed flagella; these structures also carry a terminal bulb, which perhaps makes it more adapted to swimming through mucus. Also, on the surface, the lipopolysaccharide has unique biological properties and the genes that control addition of the O-side chains can display phase-dependent variation, a mechanism for avoidance of host responses. In addition, *H. pylori* has a unique peptidoglycan structure that differs from other gram-negative bacteria (Mobley, 2001).

2.3 Unique features of *H. pylori* genomic organisation

A milestone in microbial genomics was set when *H. pylori* became the first bacterial species to have its genome sequenced and compared from two independent isolates (Alm et al., 1999). The subsequent comparison has provided the first detailed look at the physical chromosomal organization and has begun to identify a minimal set of common genes that can be used as candidates for therapeutic strategies.

There have been 1590 putative open reading frames (ORFs) identified in *H. pylori* 26695 circular chromosome (~1.7 Mbp). Almost 60% of the ORFs have a predicted function, whereas 24% are conserved in other bacterial species but are of unknown function and 17% are *H. pylori* specific with no known homologs in the current databases. These ORFans may encode proteins unique to *H. pylori* and provide selective targets for antibiotic therapy (Tomb et al., 1997).

The sequenced genome revealed a profile of an organism that was fine-tuned for its niche in the gastric mucosa, lacking many of the regulatory features found in the larger *E. coli* genome. (By contrast to the *E. coli* sequence, 10-fold less regulatory sequences have been identified in *H. pylori*) (Dorrell & Wren, 1998). Tight regulation of gene expression is imperative for enteropathogenic bacteria whether they are continually responding to the harsh acidic environment of the stomach or bile salts in the intestine. The remarkable

economy of regulatory elements in *H. pylori* may reflect the very limited range of environments in the which it survives, suggesting a highly evolved inter-relationship between man and microbe.

One intriguing feature of the genomic organization is that ~1% of the genome of *H.pylori* encodes a family of 32 outer-membrane proteins that are well conserved between the J99 and 26695 strains (Ge & Taylor, 1999). The OMPs identified provide a tractable subset of the total genome, comprising most of the proteins known to be involved in virulence (for example, those required for adherence to gastric epithelial cells and evasion of the immune system). Of particular interest is the presence of tandem repeat sequences upstream of some OMPs. In other mucosal pathogens increasing or decreasing the numbers of repeats by slipped-strand misparing and recombination affects transcription of the downstream genes. The extensive use of slipped strand mispairing is a clever form of regulation that allows the organism to present many faces to the host in terms of expression of outer membrane proteins and other surface structures. In this way minor reversible mutations rapidly change the antigen profile of the pathogen, leading to evasion of the host immune system (High et al., 1993).

Numerous restriction-modification systems are present in this species, but they differ between the two genomes analyzed. There are more than twenty homologs associated with DNA restriction and modification systems identified in other bacteria, including type I, type II, and type III systems (Alm et al., 1999; Tomb et al., 1997). The role of these enzymes is unclear at present. It has been suggested that the enzymes are involved in the breakdown of intracellular and/or intercellular DNA or that they are necessary for stimulating the formation of recombinants by DNA fragmentation (Berg et al., 1997).

The presence of the *cag* pathogenicity island was identified prior to the sequencing of the whole genome, revealing a 40-kb stretch of DNA whose presence correlates with more virulent isolates (Censini et al., 1996).

Most of the traditional protein secretion systems are encoded in the *H. pylori* genome, including ABC transporters, *sec*-dependent (leader peptide) transport, flagellar assembly (a prototype of the type III secretion system), type IV homologs in the pathogenicity island, and autotransporters such as VacA (Mobley, 2001).

H. pylori populations are extremely diverse at the genomic level. Moreover, a single host can carry several *H. pylori* strains, and isolates within an individual can change over time, as endogenous mutations and/or chromosomal rearrangements or recombination between strains occur (Cooke et al., 2005; Peek & Blaser, 2002). The comparison of *H. pylori* isolated from patients of different ethnical origin and geographical locations indicates that their nucleotide sequences segregate similarly to those of humans.
This suggests that this bacterium was already present in the stomach of humans when they left Africa to colonize the world, and co-evolved with them since then (Covacci et al., 1999).

2.4 A survey of *H. pylori* proteome

2.4.1 Proteins involved in the adaptation of *H. pylori* to the gastric environment

To survive in the extremely acidic gastric lumen, *H. pylori* utilizes unique mechanisms, the most efficient of which is ammonia production by the potent nickel-containing urease (Cussac et al., 1992b). Urease hydrolyzes urea imported by an acid-activated channel, UreI (Bury-Mone et al., 2001; Skouloubris et al., 1998)), to produce large amounts of buffering ammonia. Its activation through incorporation of nickel at the active site necessitates the action of the accessory proteins UreE, F, G, and H (Cussac et al., 1992a).
Urease is regulated by an intricate interplay of different environmental signals such as the concentration of urea and metal ions, and/or the pH. The activity of urease critically depends on the availability of nickel ions, as a functional urease complex requires 24 Ni^{2+} (Hawtin et al., 1991; Mobley et al., 1988). Hpn, Hpn-like, HspA and NixA proteins are major contributors to nickel bioavailability either through sequestering or transporting nickel ions when required (Ferrero et al., 1995; Mobley et al., 1999; Seshadri et al., 2007). Their synthesis in cells depends on the Nickel responsive transcriptional regulator NikR (Dian et al., 2006), wich together with the Iron dependent regulator Fur was found to be necessary for efficient colonization of the mouse stomach by *H. pylori* (Baldwin et al., 2007).

Two other ammonia-producing enzymes, AmiE and AmiF amidases, together with the arginase RocF, which generates urea from arginine, have been implicated in adaptive mechanisms to the low pH environment of the gastric lumen (Skouloubris et al., 1997; Skouloubris et al., 2001).

Finally, the bacterium must deal with reactive oxygen species that are generated by phagocytic cells of the host immune response. In *H. pylori*, genes encoding superoxide dismutase (SOD), catalase, and several putative peroxidases have been identified (Odenbreit et al., 1996; Pesci & Pickett, 1994; Tomb et al., 1997). Persistent colonization of *H. pylori* in a mouse infection model appears to require the function of the *sodB*, *katA* and *kapA* gene products(Harris et al., 2003).

2.4.2 Main proteins involved in *H.pylori* pathogenicity mechanisms

Numerous studies have focused on the elucidation of *H.pylori* pathogenicity mechanisms. The *cag*-PAthogenicity Island (PAI), consisting of a group of 31 genes involved in the biogenesis of a type IV secretion system (T4SS) and the translocation of the immunodominant CagA antigen, has been shown to contribute to the induction of the pro-inflammatory IL-8 cytokine by host epithelial cells through the activation of the nuclear factor κB (NFκB) pathway, and thus has been recognized as one of the major pro-inflammatory players (Odenbreit & Haas, 2002). T4SSs are ubiquitous secretion machineries formed by at least 12 proteins named VirB1-B11 and VirD4, three of them (VirB4, VirB11, and VirD4) are ATPases believed to power the assembly of the T4SS and drive substrates through it.

The VacA vacuolating cytotoxin is also one of the most studied virulence factor of *H.pylori*. The cytotoxin is thought to interfere with intracellular vesicles trafficking and its role in the persistence of the bacterial infection has been recently proposed (Gebert et al., 2003).

Adhesion of *H.pylori* to gastric epithelial cells is another important early step in the infection process. One of the most studied adhesins is the blood-group antigen binding adhesin named BabA, encoded by the *bab*A2 gene (Ilver et al., 1998). More recently,

other adhesins have been described such as AlpA-AlpB, SabA or HopZ (Mahdavi et al., 2002; Odenbreit et al., 2002)).

2.5 Surface-associated proteins of Helicobacter pylori

The surface of *H. pylori* is the first point of contact between the pathogen and the host. The biology of this microbe-host interface is sophisticated and complex, and it is made by molecular constituents that account for many of the distinguishing properties of *H.pylori*.

Surface proteins of *H. pylori* mediate important pathogen-host interactions that are essential for colonization, adherence, survival, and virulence of this pathogen. Moreover, surface proteins are particularly accessible to host immune responses and to drug therapy and thus could be used to control this important human pathogen.

2.5.1 Cell envelope, outer membrane, and LPS.

The *H. pylori* cell has been shown to have an overall relatively hydrophilic and negatively charged surface *in vitro* (Smith et al., 1990). The outer membrane composition is unique in its protein content and lipopolysaccharide structure (Aspinall & Monteiro, 1996; Aspinall et al., 1996; Monteiro et al., 1998; Moran et al., 1992). However, the physical organization of the cell envelope is similar to that of other gram-negative bacteria. It consists of an inner (cytoplasmic) membrane, periplasm with peptidoglycan, and an outer membrane. The outer membrane consists of phospholipids and LPS. The *H. pylori* outer membrane phospholipid moiety contains cholesterol glucosides The *H. pylori* outer membrane phospholipid moiety contains cholesterol glucosides (Bukholm et al., 1997; Haque et al., 1996; Tannaes & Bukholm, 2005; Tannaes et al., 2005), which is very rare in bacteria. LPS usually consists of lipid A, core oligosaccharide, and an O side chain. The lipid A moiety of *H. pylori* LPS has low biological activity compared to lipid A from other bacteria (Muotiala et al., 1992). Clinical isolates of *H.pylori* produce high-

molecular-weight (smooth) LPS with an O antigen, but during *in vitro* culturing the bacteria may convert to rough LPS variants, which lack the O side chain (Moran et al., 2002; Walsh & Moran, 1997). The O side chain of *H.pylori* can be fucosylated and mimics Lewis blood group antigens (Lewis x [Le^x] and Le^y), aiding molecular mimicry of host antigens and associated immune evasion (Appelmelk et al., 1997). The O antigen can also mimic other blood group antigens (Monteiro et al., 1998). *H. pylori* LPS displays phase variation through length variation of poly(C) tracts in the genes encoding α-1,3-fucosyltransferases (Appelmelk et al., 1999) and a poly(C) tract and poly(TAA) repeats in the gene encoding the α-1,2-fucosyltransferase (Wang et al., 1999). This LPS phase variation contributes to population heterogeneity and may allow adaptation of *H. pylori* to changing conditions in the gastric mucosa (Moran et al., 2002; Tannaes et al., 2005; Tannaes et al., 2001). The *H. pylori* genome encodes a large array of outer membrane proteins, which have been grouped into five paralogous families (Alm et al., 2000; Doig & Trust, 1994). The largest gene family consists of 21 Hop (Hsp70/Hsp90 Organizing Protein) and 12 Hor (Hop-related) outer membrane proteins, and this family includes the known (putative) adhesins of *H. pylori* (Evans & Evans, 2000). Other families include porins (Doig et al., 1995), iron transporters, flagellum-associated proteins, and proteins of unknown function.

2.5.2 Surface Localization of Cytoplasmic Proteins

The cell surface of *H. pylori* has the unusual property of being able to incorporate proteins such as urease, catalase, HspA, HspB, and superoxide dismutase (SOD), which are found virtually exclusively within the cytoplasm in other bacteria (Bode et al., 1993; Dunn et al., 1992; Hawtin et al., 1990; Radcliff et al., 1997; Spiegelhalder et al., 1993). Cryo-immunolocalization techniques have demonstrated that urease, catalase, and HspB are located strictly within the cytoplasm of freshly subcultured, early log-phase *H. pylori*. However, at the end of the log-phase, these proteins are also surface associated or extracellular (Phadnis et al., 1996). Significant fractions of urease and HspB are also surface associated *in vivo* (Dunn et al., 1997); see *Figure* 2.3). Indirect gold

immunostaining of *H. pylori* SOD with a polyclonal antibody directed against the iron-containing SOD of *E. coli* showed a surface localization of the enzyme (Spiegelhalder et al., 1993). Gold particles were distributed on the cell envelope and on the sheath of the flagella. SOD was located on the outer surface of a limited number of bacteria, but the enzyme was cytoplasmic in most cells.

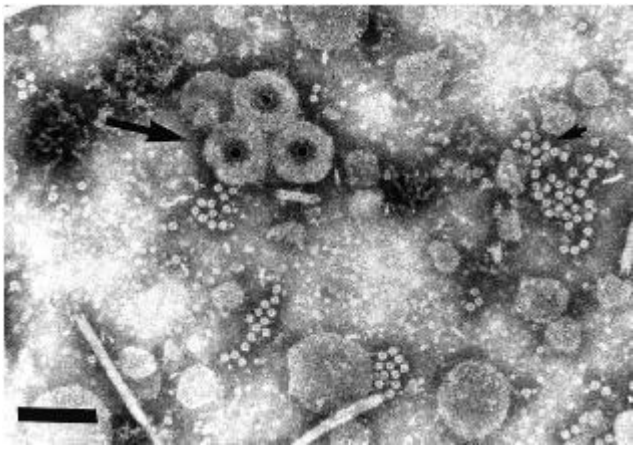


Figure 2.3: Surface structures of *H.pylori.* Negative stain of bacteria showing urease as represented by the small donut-like structures (black arrow).

2.5.3 Mechanism of Surface Localization of Cytoplasmic Proteins

In contrast to other *H. pylori* proteins with specific secretion pathways (Cao et al., 1998; Telford et al., 1994), the mechanism whereby urease, HspA, HspB, catalase, and SOD become associated with the other membrane of *H. pylori* is controversial. Urease, HspA and HspB are large oligomeric proteins (Austin et al., 1992) that are typically found exclusively in the cytoplasm of bacteria and would not be expected to cross the bacterial outer membrane. *H. pylori* urease, HspA and SOD, and catalase have been shown genetically to lack leader peptides (Cussac et al., 1992b; Ferrero et al., 1992; Odenbreit et al., 1996; Spiegelhalder et al., 1993; Suerbaum et al., 1994). Alternative transport mechanisms must therefore exist. Some very elegant work has shown that although *H.*

2.5.4 Significance of Surface Localization of Cytoplasmic Proteins

The ability of released cytoplasmic proteins of *H. pylori* to bind to the cell surface is considered to be biologically important. For example, the surface locations of catalase and SOD could enable these enzymes to function more effectively as a defense mechanism against phagocytic attack. In the case of urease, one study indicated that free or extracellular urease is irreversibly inhibited at pH of <4.5. The authors concluded that external urease is ineffective as an acid-protective device at the lower pH values in the stomach (Scott et al., 1998). However, in another study, bacteria with only cytoplasmic urease showed significantly reduced survival when exposed to acid in the presence of 5 mM urea when compared to the survival of bacteria with both cytoplasmic and extracellular urease (Krishnamurthy et al., 1998). This indicates that cytoplasmic urease activity alone is not enough to enable *H. pylori* to survive in acid. The authors of this study speculate that possibly the association of urease with the outer membrane of *H. pylori* protects urease from inactivation by acid. Development of mutants in which urease is located strictly within the cytoplasm would be helpful to resolve the discrepancies between the two studies. In addition to mediating acid resistance, other roles have been postulated for surface-associated urease that include binding to mucin at acidic pH (Icatlo et al., 1998).

The observation that urease, HspA, HspB, and catalase are surface associated helps explain how they can serve as vaccine components in animal trials (Ferrero et al., 1994; Ferrero et al., 1995; Michetti et al., 1994; Radcliff et al., 1997). However, it is likely that

in vivo a subpopulation of *H. pylori* cells do not carry these antigens on the cell surface, and such a subpopulation may therefore evade the protective effect of the vaccine.

2.6 Conclusions

Characterization of the cell envelope of *H. pylori* has identified a number of important features that distinguish it from other bacterial pathogens. These are the simple structure of its peptidoglycan, its unusual cellular fatty acid and lipid profile, molecular mimicry of Lewis antigens by LPS, and the presence of cytoplasmic proteins such as urease, Hsps, and SOD on the cell surface. In addition to these characteristics, polar flagella and a unique repertoire of outer membrane proteins are all likely to contribute to the ability of the organism to colonize the stomach and cause disease. The significance of the outer membrane components, many of which are likely to function as adhesins, is highlighted by the fact that *H. pylori* devotes a significantly high proportion of its coding capacity to them. Many of the vaccine candidates for *H. pylori* are proteins found on the cell surface, underlining the importance of further characterization of these proteins and of elucidating the precise mechanism of interaction of the *H. pylori* cell surface with gastric mucosa.

To address the different challenges presented by the human stomach - a unique ecological niche characterized by very acidic pH - *H. pylori* has developed a unparalleled combination of virulence factors. The study of those factors disclosed some interesting points. First, some proteins wich are found exclusively within the cytoplasm in other bacteria are often surface-associated in HP. In addition, proteins with known functional roles in other organisms can display additional domains thereby coupling different functions on a single molecular device.

Section I

Reductionist Thinking

Reductionism is the explanatory arrow that always points downward.

Stephen Weinberg

Chapter 3

Unique properties of heat shock protein A from *Helicobacter pylori*

3.1 Introduction

Helicobacter pylori synthetizes an unusual GroES homolog, Heat Shock Protein A (HspA). Besides its classical co-chaperone activity, HspA plays additional roles being involved in nickel binding and urease activity. Furthermore, it exhibits an extended subcellular localization, ranging from cytoplasm to bacterial cell surface. This study pointed on another unique characteristic of HspA among all GroES proteins: an high content of cysteine residues. Cysteine is the less represented residue in all GroES proteins examined so far. In this context we have produced and characterized a recombinant HspA. In particular the study has been addressed on the Cys oxidised/reduced state; the disulphide bridge pattern has been assigned by integrating classical biochemical methodologies with mass spectrometry. These results can be related to the different redox environments that the protein experience inside and outside the bacterial cell.

3.1.1 GroEL and GroES chaperonins

Heat shock proteins (HSPs) of the GroEL/GroES class are a highly conserved group of proteins found in all domains of life (Craig et al., 1993; Gupta, 1995; Langer & Neupert, 1991). Despite their designation, they are expressed at a low level and play a vital role at all temperatures in maintaining normal cell function (Fayet et al., 1989). Increased synthesis of these proteins occurs in response to many environmental stresses, including temperature changes (hence their name), oxygen limitation, glucose starvation and iron deprivation for bacteria, but also, as far as eucaryotic cells are concerned, inflammation, irradiation, malignant transformation and T-lymphocyte activation with mitogens or lymphokines (Neidhardt & VanBogelen, 1987).

These proteins are involved in the intracellular folding and assembly of various other polypeptide chains into oligomeric complexes, without becoming part of of the mature products (Ellis & van der Vies, 1991). The Hsp60 chaperonin (GroEL homologue) act together with the Hsp10 protein (GroES homologue). Both form homo-oligomers of two stacked rings, each ring having a seven-fold simmetry (Craig et al., 1993; Jaenicke & Creighton, 1993). The present view of the mechanism by wich they act suggests that the Hsp60 alone interacts with the unfolded polypeptide, subsequently the Hsp10 binds to the Hsp60-polypeptide complex and plays a role in the release of the polypeptide by modulating the ATPase activity of the Hsp60 moiety (*Figure* 3.1).

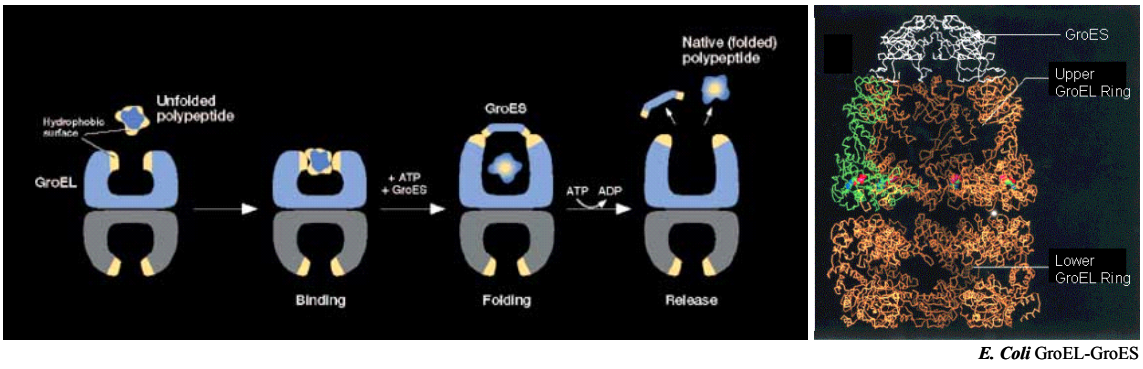


Figure 3.1: The GroEL-GroES fold recovery machine.

The Hsp10 and Hsp60 proteins are commonly encoded by a single bicistronic operon that leads to the coordinated expression of the two genes. The thermo-inducible nature of the heat-shock genes depends upon the bacteria. Two major mechanisms of regulation at the transcriptional level have been identified:

I. In most bacteria the genes are induced by activating transcription from promoters specifically recognized by an RNA polymerase containing the Sigma 32 factor (σ32). Therefore, the level of expression of the *hsp* genes is directly proportional to the amount of the σ32 factor present in the cell, encoded by the *rpoH* gene (Cowing et al., 1985).
II. A second mechanism involves the presence of a conserved motif forming a hairpin-loop structure, TTAGCACTC-N9-GAGTGCTAA, located between a vegetative promoter sequence and the start site of the structural gene (Zuber & Schumann, 1994). This motif is known as CIRCE (controlling inverted repeats of chaperone expression).

Interestingly, none of the two regulation systems mentioned above have been identified in the *H.pylori hsp* gene cluster. It seems that the Hsp10 and Hsp60 proteins have vital functions for *H.pylori,* since the attempts to construct *H.pylori* mutants in wich either the hsp60 or the hsp10 gene has been disrupted have been unsuccessful (Kansau & Labigne, 1996).

3.1.2 *H.pylori* Heat Shock protein A

The study of the proteins involved in the HP adaption to the gastric mucosa outlined some interesting points. First, some proteins wich are found exclusively within the cytoplasm in other bacteria are often surface-associated in HP. In addition, proteins with known functional roles in other organisms can display additional domains thereby coupling different functions on a single molecular device.

In this context, the GroES-like Heat Shock protein A of *Helicobacter pylori* can be regarded as a paradigm for enhanced protein localisation and functionality. Whilst *H.pylori* GroEL-like (Heat-Shock protein B, HspB) was shown to be very similar to other bacterial GroEL-homologues, *H.pylori* appeared to be unique in structure. Indeed, *H.pylori* HspA consists of 118 amino-acids divided in two domains: an N-terminal domain (domain A, residues 1-90), that is homologous with other GroES bacterial proteins, and a C-terminal domain (domain B, residues 91-118), wich other GroES-like proteins lack (Suerbaum et al., 1994). This striking C-terminal domain of HspA was found only in *H.pylori* 26695, *H.pylori* J99 and *H. acinonychis* genomes and is histidine- and cysteine-rich, with eight histidine and four cysteine residues among 27 amino acids. HspA primary structure is shown in *Figure* 3.2.

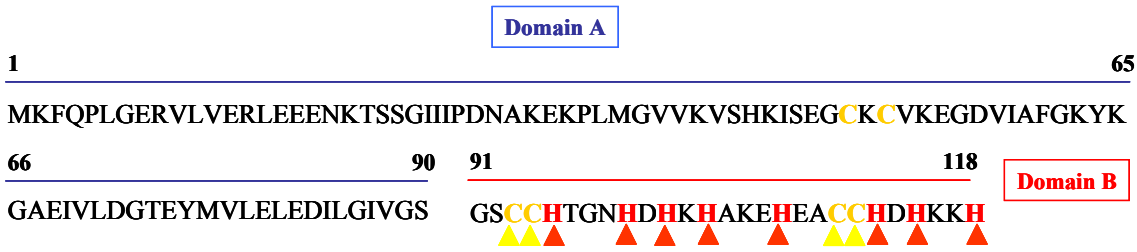


Figure 3.2: Heat Shock protein A primary structure.

The B domain is highly charged and it exhibits a high and specific affinity for nickel ions in comparison with its affinity for other divalent cations (copper, zinc, cobalt). It has two distinct Ni^{2+}-binding sites: a high-affinity site with Kd ~ 2.8 μM and a lower-affinity site, which binds Ni2+ at concentrations above 30 μM (Kansau et al., 1996). A recent study found that HspA binds Bi3+ with high affinity via coordination to cysteine residues. Bismuth is a commonly recommended metallodrug for the treatment of *H.pylori* infections (Ge et al., 2007).

The presence of a C-terminal histidine-rich domain associated with nickel binding has already been described for UreE, a nickel carrier protein that is expressed by most of the

urease gene clusters as one of the accessory proteins required for urease metallocentre assembly (i.e. activation of the apoenzyme) (Benoit & Maier, 2003). Urease is a Ni-dependent enzyme that produce ammonia and then acts as a virulence factor in Hp, by protecting the bacterium from the acidic environment (Ha et al., 2001). Curiously, none of the H.pylori accessory proteins - including the H.pylori UreE homologue - exhibits such a nickel-binding motif that, instead, occurs in HspA (Maier et al., 2007). Therefore, HspA has been suggested to play a role in the assembly and/or stabilization of the urease/Ni complex through its histidine rich C-term domain. So far, however, all attempts to visualize a complex consisting of the polypeptides UreA, UreB, HspA and HspB have been unsuccessful. Interestingly, evidence for a possible interaction of HspA with the H.pylori urease holo-enzyme was provided by a genetic approach. In fact, the co-expression in E.coli of the urease gene cluster and of the HspA gene leads to a fourfold increase of urease activity when compared to the expression of the urease gene cluster alone (Kansau et al., 1996).

HspA is an abundant protein, known to be immunogenic. The immune response to this protein in *H. pylori*-positive adults increases with age (Eamranond et al., 2004). Among many clinical isolates tested, the A domain is highly conserved, whereas the B domain encompasses two variant type sequences that differ from each other by the simultaneous substitution of seven amino acids . Interestingly, none of the substitutions affects the hystidine and cysteine residues. The different domains of HspA elicit distinct host immunological responses, and the A domain was found to be the immunodominant domain (Kansau et al., 1996).

In 1995, Ferrero *et al.* demonstrated that both HspA and HspB confer protective immunity against mucosal infection in mice (Ferrero et al., 1995). Since then, the development of HspA-based vaccines against *H. pylori* became an active research field (Jiang et al., 2003; Todoroki et al., 2000; Zhang et al., 2006) . *H. pylori* HspA is particularly appealing as a vaccine component because it has a unique structure, with the C terminal domain wich is absent from other known heat shock homologues, including those of eukaryotic organisms. Moreover, the capacity of this domain to bind nickel ions facilitates large-scale purification of the polypeptide. Lastly, this antigen is also

expressed by all the isolates and is higly conserved at the amino acid level (Kansau & Labigne, 1996).

HspA exhibits an extended subcellular localization, ranging from cytoplasm to bacterial cell surface (Phadnis et al., 1996). It is worth noting that HspA is the only nickel chaperon wich was found to be surface-associated among the several nickel-binding and accessory proteins facilitating Ni-enzyme maturation in H.pylori. Several mechanisms have been proposed for the HspA extracellular release, including specific secretion pathways, autolysis, and membrane vesicle formation, but a conclusive explanation is still missing (Phadnis et al., 1996; Vanet & Labigne, 1998). Therefore, the protein experience different pH and redox environments when it moves from the bacterial cell to the extracellular space.

There is another unique characteristic of HspA among all GroES proteins: an high content of cysteine residues. The protein sequence contains 6 cysteines (two in the A domain and four in the B domain) whereas cysteine is the less represented residue in all GroES proteins examined so far. It can be deduced that cysteine is not an essential residue for chaperonins, and hence HspA has to play additional roles mediated by its unusually high number of Cys residues.

In this context we have produced and characterized a recombinant wild-tipe HspA, together with a single mutant (Cys94Ala) and a double mutant (C94A/C111A). In particular the study has been focused on the Cys oxidised/reduced state in order to deeper understand the HspA enhanced functionality. The disulphide bridge pattern has been assigned by integrating classical biochemical methodologies with mass spectrometry. These results can be related to the different redox environments that the protein experience inside and outside the bacterial cell.

3.2 Results

3.2.1 Bioinformatics

Sequence analysis

A statistical survey of cysteine distribution in GroES proteins was carried out using the available sequences in the non redundant database RefSeq (Pruitt et al., 2007). Following the selection criteria defined in the Method section, and excluding HspA and its homologues, we identified 306 GroES sequences. Notably, only 13 out of them contain Cys residues. The Cys frequency across GroES proteins is 14/30165 (0.05%) revealing that cysteine is the less represented aminoacid within the selected sequence subset. As a result, it can be deduced that Cys is not an essential residue for the function of GroES chaperonins.

With this data at hand, the high Cys content (6/118) peculiar of Helicobacter-related HspAs suggests that they might play additional and still unknown roles, peculiar of GroES-like proteins from bacteria that attach the gastric mucosa.

3.2.2 Chemico-physical characterisation

SDS-PAGE experiments

The first experimental information about disulphide bridges in HspA can be obtained from SDS-PAGE experiments. As displayed in *Figure* 3.3, protein samples made according to the original Laemmli's protocol exhibit a characteristic two-band pattern when loaded onto SDS-polyacrilamide gels. Besides the expected monomer band at 13KDa, there is an additional band at 26KDa, thereby suggesting the presence of HspA in a dimeric form. The 26KDa form is unusually resistant to the denaturing effects of SDS, as well to the action of strong reducing agents like DTT, BME or TCEPP. SDS-PAGE of

the protein in absence of reducing agents gives a third band at 40KDa, corresponding to a trimeric form.

Addition of DTT, or BME, removes the trimer but has little or no effect on the dimer (*Figure* 3.3).

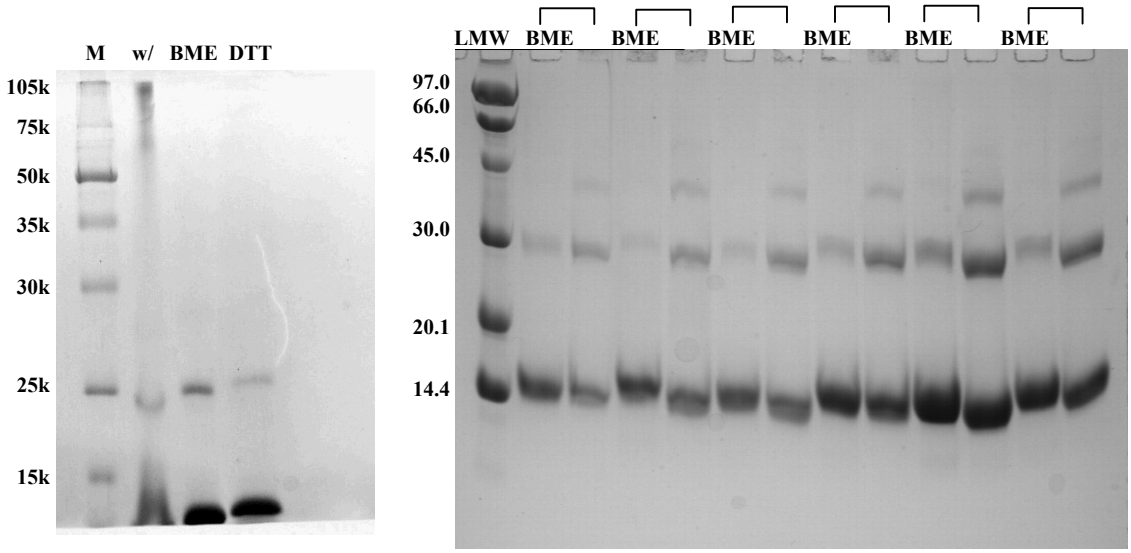


Figure 3.3: Typical SDS-PAGE pattern for HspA, with and without reducing agents.

Although the 26KDa band persists after incubation with BME or DTT, it seems that the dimeric form is covalently bounded by disulphide bridges. There are at least two experiments that support this hypothesis:

1) "Strong" denaturation step, e.g. incubation with an high concentration of SDS (3%) before electrophoresis, has no effect on the monomer-dimer pattern;

2) After alkylation of HspA Cys residues, the 26KDa dimeric band disappear from SDS-PAGE (*Figure* 3.4).

Hence, HspA can exist in dimeric forms and these multimeric forms appeared to be resistant to the denaturing effects of SDS.

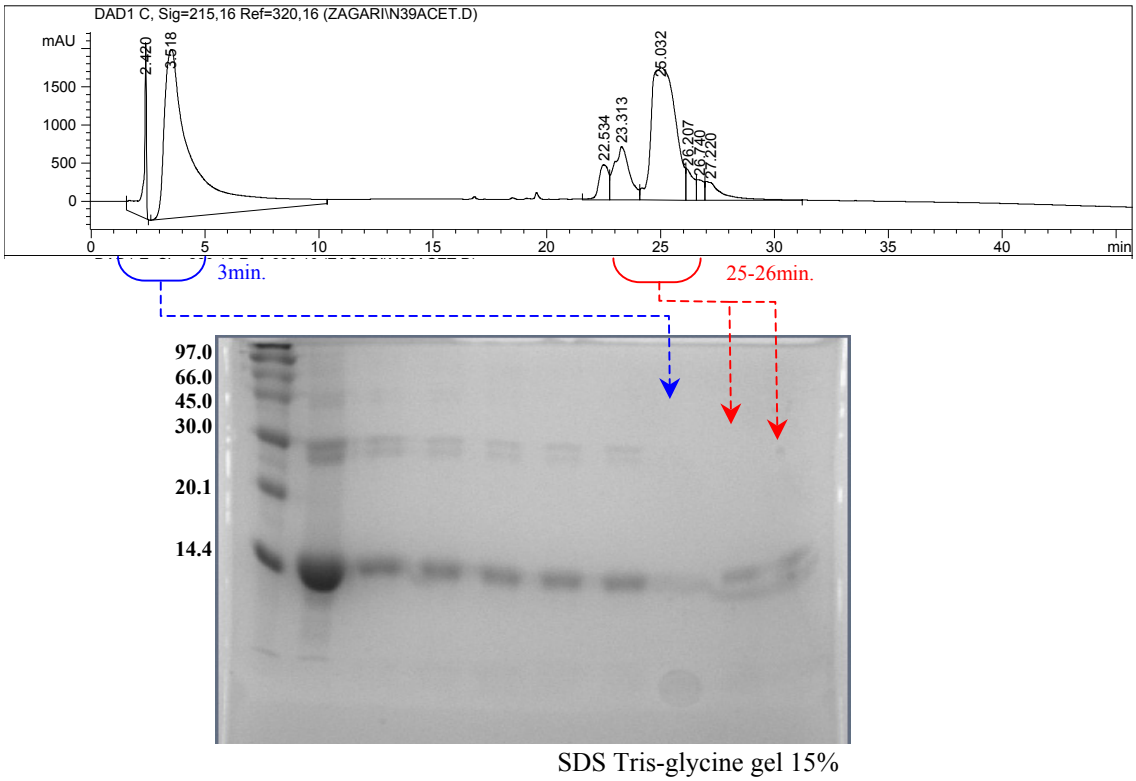


Figure 3.4: Alkylation of HspA Cys residues and corresponding SDS-PAGE pattern.

Secondary structure content and thermal stability

A typical CD spectrum for HspA is shown in *Figure* 3.5. As can be noted, the CD profile is qualitatively similar to the CD profile for E.coli GroES (Boudker et al., 1997). The main difference between the two spectra is about the magnitude of the molar ellipticity, HspA value being approximately six times less than the value reported for E.coli GroES. This suggests that HspA is still folded and contains mostly beta structures. The lower magnitude observed for HspA can be related to the fact that the protein has a lower secondary structure content than E.coli GroES.

Thermal denaturation experiments on the same sample allowed to draw the thermal denaturation curve and to evaluate the associated melting point (80C), as shown in *Figure* 3.5.

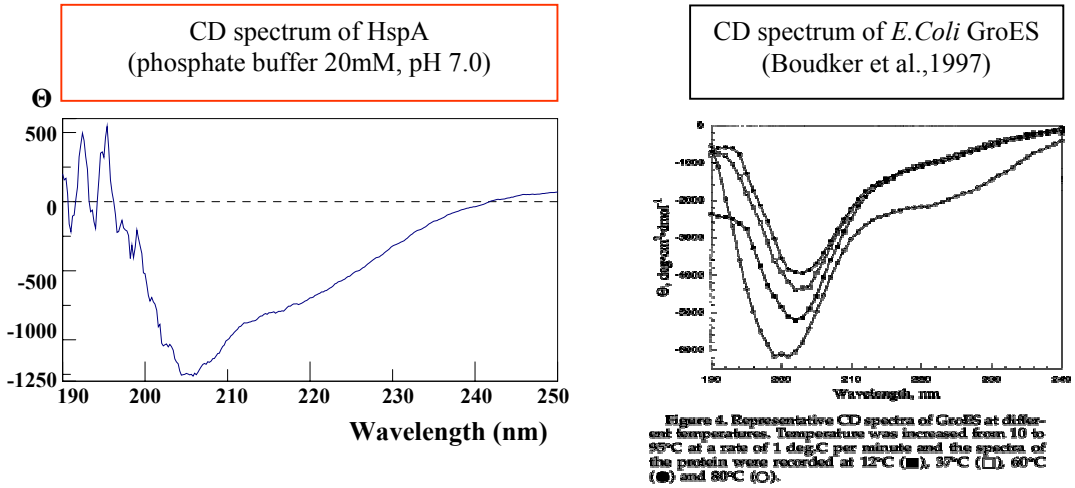


Figure 3.5: CD spectrum of HspA; CD spectrum of *E.Coli* GroES

The same experiments have been repeated on a reduced sample, e.g. HspA incubated with DTT. The CD spectrum showed a lower molar ellipticity value, while holding the same overall profile. Thermal denaturation revealed a lower melting point -around 65C- for the reduced form (*Figure* 3.6).

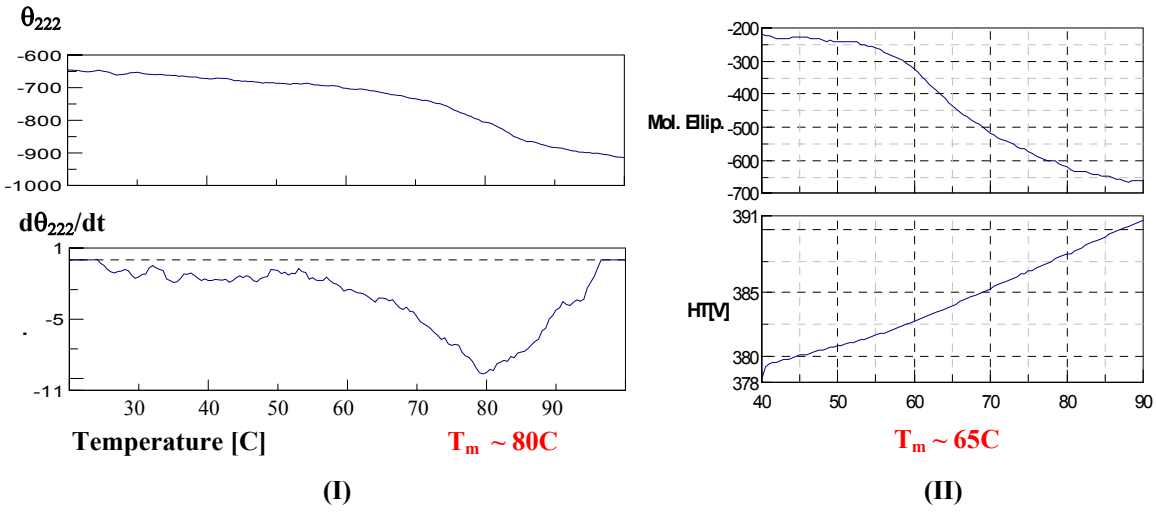


Figure 3.6: Thermal denaturation profiles for oxidized (I) and reduced HspA (II)

Insights into oligomer/multimer equilibrium

After gel-filtration, the native protein (13KDa) is eluted mainly as a 90KDa oligomer -probably an heptamer- comparing the elution profile with the calibration curve. A second peak at smaller elution volumes is also eluted, containing multimeric aggregates whose molecular weights are higher than 90KDa (*Figure* 3.7).

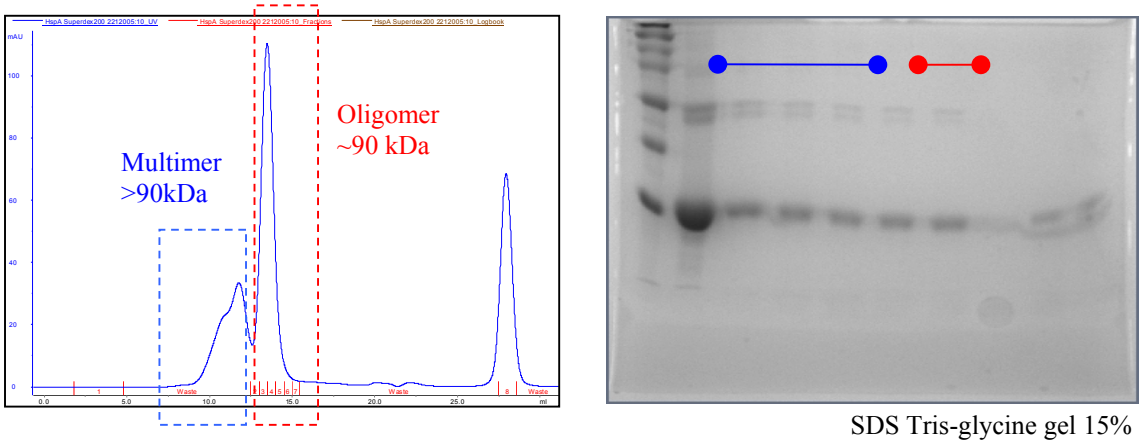


Figure 3.7: HspA gel filtration profile followed by SDS-PAGE

If the gel filtration is repeated few days after the initial purification on a sample stored at 4C, the chromatograms show a peculiar time-course profile. The multimeric peak grows whereas the peak associated to the oligomer decreases, until an equilibrium is reached. In the equilibrium chromatogram, the multimer peak is asymmetric and it contains several aggregated forms. Moreover, the ratio between the two peaks is constant. If the oligomer peak is re-injected, it gives essentially the same profile (*Figure* 3.8).

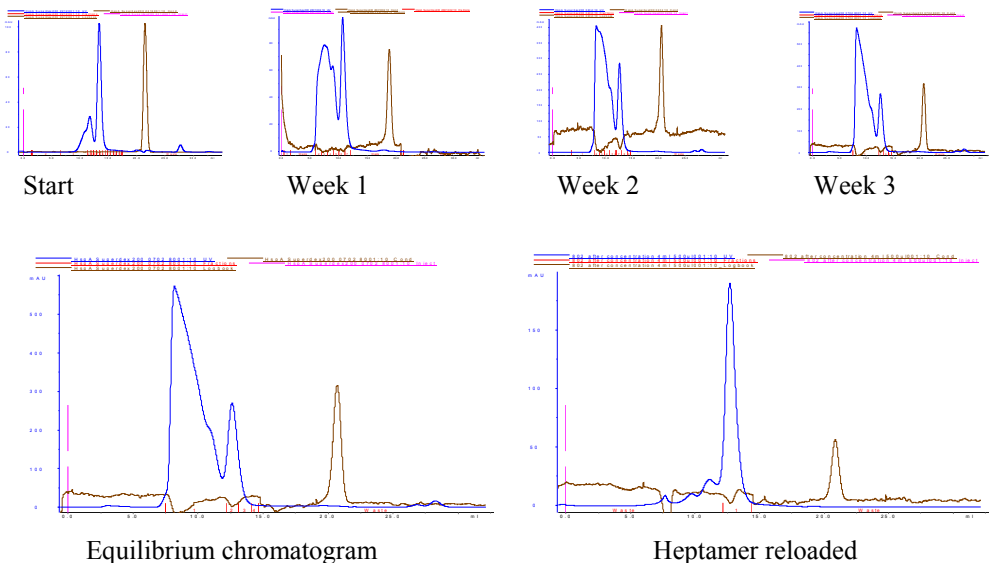


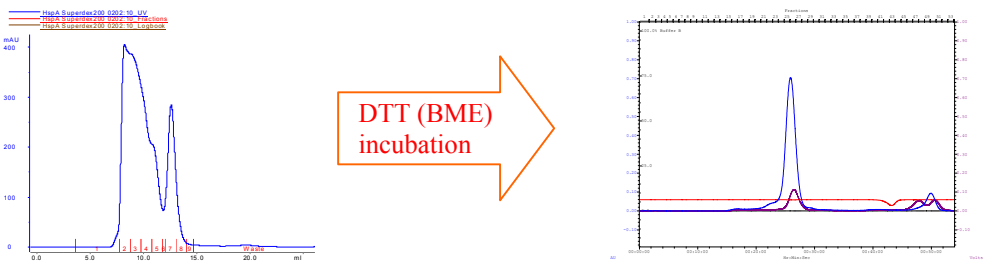
Figure 3.8: Time - course experiments on HspA using gel filtration

Incubation with a reducing agent like DTT, followed by gel filtration, converts all multimeric forms into the oligomeric one (*Figure* 3.9a). This evidence comes also from non-denaturing PAGE experiments (*Figure* 3.9b). This strongly suggests that aggregation is due to oxidation, e.g link of several oligomers together by disulphide bridges.

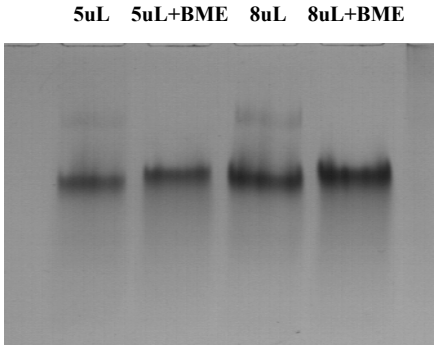
3.2.3 Assignment of the disulphide bridge pattern via mass spectrometry

Assessment of the oxidation state of HspA Cys residues

The oxidation state of the cysteine residues occurring in HspA and its mutants was assessed by determination of the accurate molecular mass of the proteins by electrospray mass spectrometry (ESMS) following a procedure previously developed (Amoresano et al., 2003). Aliquots of the HspA samples were directly analysed by ESMS producing the spectra shown in *Figure* 3.10. Native HspA exhibited a molecular mass of 12984.0±0.8



(a)



Native gel 7.5%

(b)

Figure 3.9: Reduction converts all HspA multimeric forms into the oligomeric one

Da, about 6 Da lower than expected for the fully reduced protein (theoretical mass value 12990.9) thus suggesting that the 6 Cys residues were involved in 3 disulphide bridges. Accordingly, the molecular mass of HspA shifted to 12989.9±0.4 Da following reduction with DTT, exhibiting the expected increase of about 6 Da, whereas no differences in the mass value were observed following direct alkylation of the protein with an excess of iodoacetamide (IAM), confirming the absence of free Cys residues (data not shown). Finally, when the native protein was first reduced with DTT and then carboxyamidomethylated with IAM, the corresponding ESMS spectrum exhibited a molecular mass of 13331.5±0.8, as shown in *Figure* 3.10b, with an increase of 340.6 Da, corresponding to the alkylation of 6 Cys residues.

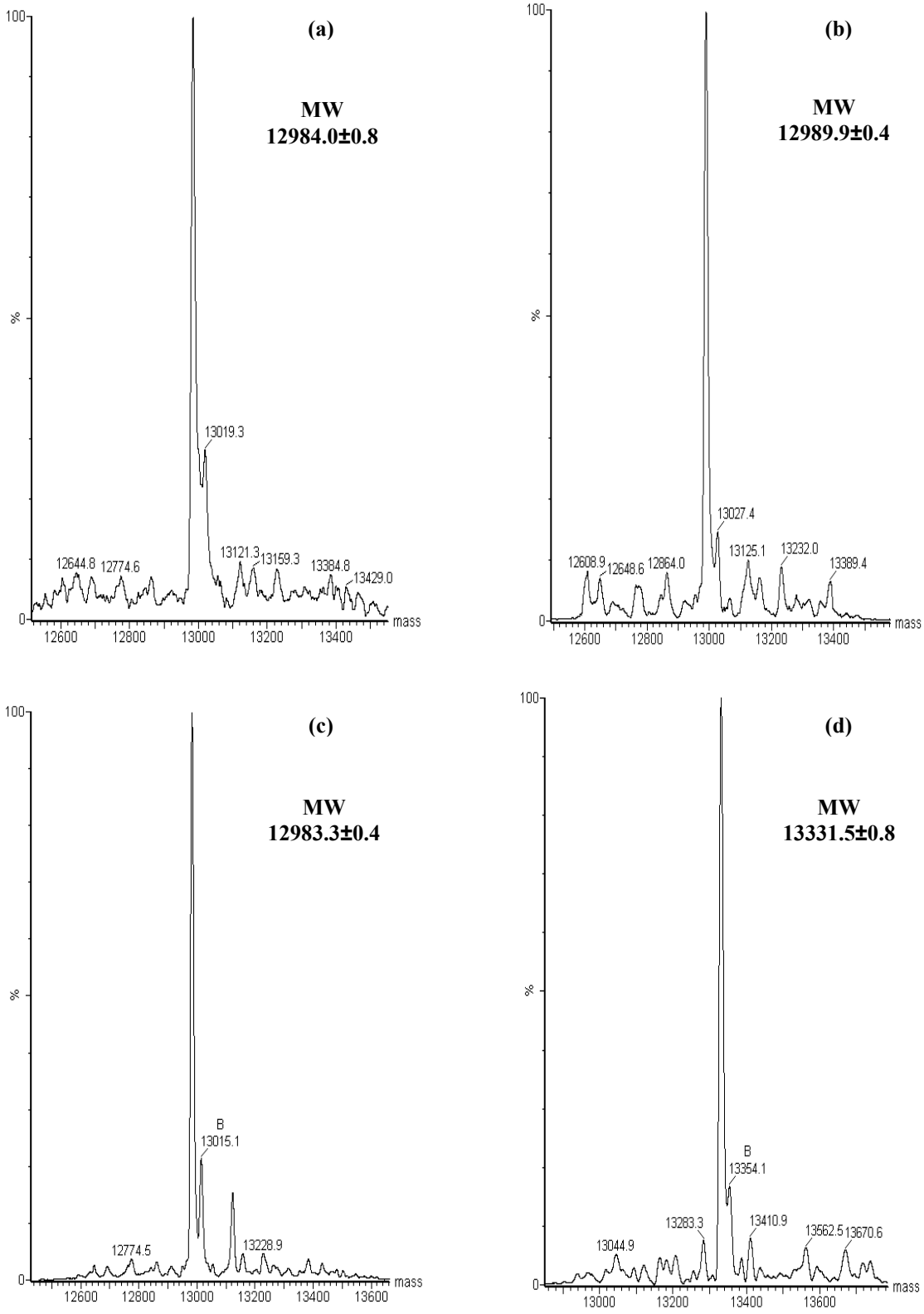


Figure 3.10: ESMS spectra. a)WT HspA; b) HspA carboxymethylated at position 51, 53, 94, 95, 111, 112, after reduction with DDT; c) C94A/C111A double mutant; d) dimeric C94A mutant.

ESMS spectrum of the Cys94Ala/Cys111Ala double mutant exhibiting a molecular mass of 12921.0±0.4 Da, about 4 Da lower than expected for the fully reduced protein, suggesting the occurrence of two S-S bridges. This result was confirmed by the reduction and reduction/alkylation experiments analogous to those described above for native HspA, indicating that very likely the two mutated cystein residues, Cys94 and Cys111, were joined by a disulphide bond in the native HspA molecule. Finally, the ESMS analysis of the Cys94Ala single mutant yielded a mass value of 25903.8±1.0 Da, as shown in *Figure* 3.10d, corresponding to a dimeric species stabilised by an intermolecular disulphide bridge very likely involving the free Cys111 residue of each HspA molecule.

Disulphide bridge pattern

Assignment of the disulphide bridges pattern in both native HspA and the double mutant was essentially accomplished according to the well established mass mapping strategy (Amoresano et al., 2001; Amoresano et al., 2003; Galliano et al., 2003; Morris & Pucci, 1985). Native HspA was digested with trypsin in 50mM ammonium acetate buffer at pH 6.0 to avoid scrambling of S-S bridges and the resulting peptide mixture was directly analysed by MALDI-MS. *Figure* 3.11 shows the high mass region of the resulting MALDI spectrum recorded in linear mode where a series of mass signals was attributed to S-S bridged peptides. The peaks at m/z 4234.8 and 3969.6 were interpreted as arising from the peptide 78-103, originated by aspecific cleavage at Tyr77, joined to the fragments 107-118 and 107-116, respectively by two disulphide bridge involving the four Cys residues at positions 94, 95, 111 and 112. Both mass signals were accompanied by satellite peaks at about 335 Da higher (m/z 4571.8 and 4304.5) corresponding to the same peptide 78-103 linked to the fragments 104-116 and 104-118 respectively via the same two S-S bridges. Finally the mass signals at m/z 3891.1 and 3626.3 were attributed to the peptide 81-103 bridged to the fragments 104-116 and 107-116 by two disulphides involving the above mentioned four cysteines. All these signals disappeared following incubation with dithiothreitol.

Although the assignment of the individual S-S bridges could not be defined on the basis of these results, the mass spectral investigation of HspA provided the overall scheme of cysteine pairings with Cys94, 95, 111 and 112 originating a network of two disulphide bridges and Cys51 and Cys53 forming the third S-S bridge. This interpretation was supported by the weak mass signal at m/z 2404.3 observed in the MALDI reflectron spectrum. This peak occurred 2 Da lower than the peptide 42-64, suggesting the presence of an intermolecular disulphide bridge joining Cys51 and Cys53.

The correct pairings of the Cys residues in HspA was obtained by the MALDI mass spectral analysis of the tryptic digest of the HspA double mutant. As shown in *Figure* 3.11b, mutation of the two Cys residues at positions 94 and 111 seemed not to affect the C-terminal pattern of disulphide bridges. The mass signals at m/z 5344.7, 5053.7 (peptides 65-103 and 67-103 joined to 107-116 respectively) and 3907.7, 3564.6 and 3322.4 (peptides 78-103, 81-103 and 86-103 linked to the fragment 107-116) essentially

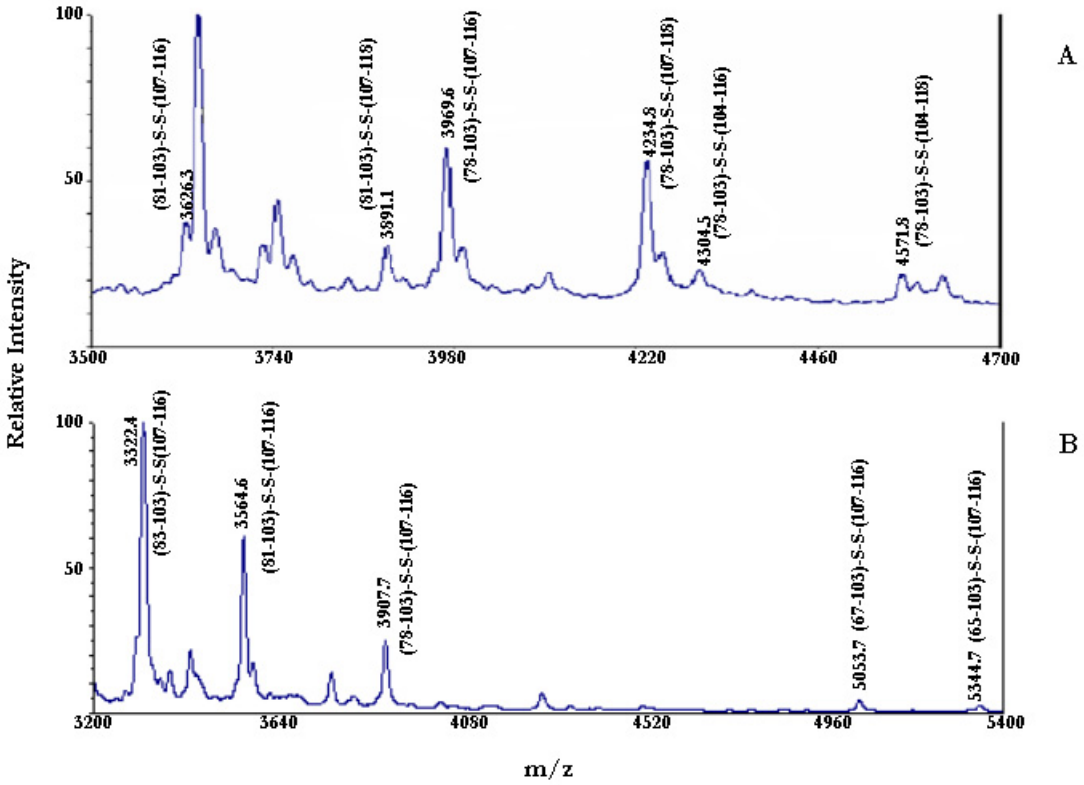


Figure 3.11: MALDI-MS spectra after tryptic digestion of : a) WT Hspa;
b) C94A/C111A double mutant.

corresponded to the previously observed S-S bridged peptides. This could only be explained by considering that the two mutated cysteines are held together by an S-S bridge in the native protein. The C-terminal fragments are then still joined by the second disulphide bond involving Cys95 and Cys112. The third S-S bridge involving Cys51 and Cys53 was confirmed by the mass signal at m/z 964.5 (peptide 47-55 containing an intramolecular disulphide bridge) leading to the unambiguous assignment of the S-S bridge pattern in HspA.

3.3 Materials and methods

3.3.1 Cys distribution in GroES proteins

The search was carried out against all sequences in the non redundant database RefSeq (www.ncbi.nlm.nih.gov). The database was accessed in February, 2007. The keyword *GroES* in the Gene Name field restricted to master sequences (master of set of segmented sequences) resulted in 313 hits; 4 sequences consisting of a residue number less than 50 or greater than 220 were discarded; the 3 HspA sequences from HELPY, HELPJ, HELAC were also discarded. The working subset of GroES sequences contained 306 sequences consisting of 30145 residues in total. The search for Cys residue revealed only 14 cysteines in 13 sequences. For comparison, the frequency of all 20 naturally occurring aminoacids was computed. the most abundant aminoacid was valin with 3603 residues (12.0%) whereas the less represented – apart from Cys – was tryptophan, with 66 residues (0.2%), as expected.

A query in the PDB database for the 3D-structure of any of the 13 Cys-containing sequences resulted in 0 hits.

3.3.2 Expression and purification of HspA from H. Pylori

The gene encoding HspA (*hp0011)* was inserted into the pET11a expression vector (Novagen) to give pILL948 as described (Kansau et al., 1996). Single (Cys94Ala) and double (Cys94Ala, Cys111Ala) HspA mutants were constructed using the QuickChange site directed mutagenesis protocol (Stratagene). The vector was introduced into the *E. coli* BL21star(DE3)pLysS expression strain (Invitrogen) using heat-shock: 1ul of vector (50 ng/ul) was added to 100ul of cells. The mixture was incubated for 30 min. at 4°C, then at 42°C for 45 sec. and finally at 4°C for 5 min. The transformed cells were transferred on LB-Agar plates (15 g/L Agar, 10g/L NaCl, 10g/L Tryptone, 5g/L Yeast extract) supplemented with antibiotics (100ug/mL Ampicillin, 50ug/mL Cloramphenicol) and incubated overnight at 310K. The pre-inoculum was made in 10mL of LB (Luria-Bertani growth medium: 10g/L NaCl, 10g/L Tryptone, 5g/L Yeast extract) supplemented with antibiotics (100ug/mL Ampicillin, 50ug/mL Cloramphenicol), using selected single colonies from LB-Agar plates; the coltures were then incubated overnight at 310K. For protein preparation, 1l of LB medium supplemented with 100 mg/l ampicillin and 50 mg/l chloramphenicol was inoculated with 10mL of pre-inoculum described above and the resulting culture was grown at 310K. When the OD_{600} reached 0.4, expression of recombinant protein was induced by the addition of isopropyl-*d*-thiogalactopyranoside (IPTG) to 1 mM final concentration and cells were incubated for further 4h at 310K. The cells were harvested by centrifugation at 5000g for 30min and resuspended in lysis buffer (20 mM Na_2HPO_4 pH 7.4, 500 mM NaCl and 1 mM $MgCl_2$); protease inhibitors (Roche), DNase I (Sigma-Aldrich) and lysozyme (Roche) were added and the cells were disrupted by french press (three runs at 10000psi) or sonication (5min. total time, 20sec pulse + 20sec pause). The recombinant protein was then isolated from the cell debris by centrifugation at 21000g for 30 min.

The resulting protein extract was loaded onto a HiTrap HP Chelating Sepharose column (5 ml) charged with 100 mM $NiSO_4$ and equilibrated with buffer A (20 mM Na_2HPO_4 pH 7.4, 500 mM NaCl). The protein was eluted from the column using a gradient of 0-

100% buffer B (20 mM Na_2HPO_4 pH 7.4, 500 mM NaCl and 500 mM imidazole) (Fig.1). Fractions were analyzed by SDS-PAGE (Fig.2) and those that showed a prominent band corresponding to a molecular weight of 13 kDa were pooled and dialysed overnight into buffer C (20 mM Tris pH 7.4, 200 mM NaCl) before size exclusion chromatography (Superdex 200 (10/30) column; Amersham Biosciences) in the same buffer. The protein was probably eluted as an heptamer, comparing the elution profile with the calibration curve (Fig.3). The pooled fractions of HspA were concentrated to 20 mg/ml using an Amicon-5k concentrator (Millipore) and stored at 277K in Buffer C. The purity and homogeneity was tested by SDS-PAGE and by mass spectrometry.

3.3.3 Circular dicroism experiments

All CD spectra were recorded with a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system (Model PTC-423-S). The spectropolarimeter was calibrated with an aqueous solution of 1S-(+)-10-camphorsulfonic acid at 290 nm. Molar ellipticity per mean residue, [θ] in degrees cm^2 $dmol^{-1}$, was calculated from the equation: $[\theta] = [\theta]_{obs}mrw(10lC)^{-1}$, where $[\theta]_{obs}$ is the ellipticity measured in degrees, mrw is the mean residue molecular mass (111.5 Da), *C* is the protein concentration in g L^{-1}, and *l* is the optical path length of the cell in centimeters. Far-UV measurements (183-250 nm) were carried out at 20 °C using a 0.01 cm optical path-length cell and a protein concentration of 3.0 mg mL^{-1}. Cuvettes with path lengths of 0.2 and 0.5 cm and protein concentrations of 0.2 and 0.8 mg/mL were used in the far-UV and near-UV regions, respectively. CD spectra were recorded with a time constant of 4 s, a 2 nm bandwidth, and a scan rate of 5 nm/min, and were signal-averaged over at least three scans, and baseline corrected by subtracting a buffer spectrum. Thermal unfolding curves were recorded in the temperature mode, over the range of 5-110 °C, with a scan rate of 1.0 °C/min. Samples were rapidly cooled after the first heating run and scanned for a second time to estimate the reversibility of the unfolding transition.

3.3.4 Chemical modification of cysteine residues

Reduction of Cys residues in HspA was carried out in 0.25 M Tris-HCl, 1.25 mM EDTA, pH 8.5 containing 6 M guanidinium chloride by incubation with a 10:1 molar excess of dithiothreitol (DTT) over the total -SH groups at 37°C for two hours under nitrogen atmosphere. Alkylation of native HspA was performed by incubation of the protein in the same buffer with a 10-fold molar excess of iodoacetamide over the total -SH groups for 10 min at room temperature in the dark. For reduction and carboxyamidomethylation of the Cys residues, HspA was incubated with a 10:1 molar excess of DTT over the total -SH groups at 37°C for two hours under nitrogen atmosphere followed by alkylation with a 10-fold molar excess of iodoacetamide over the total -SH groups for 10 min at room temperature in the dark. In all cases, the protein samples were desalted by reverse phase HPLC on a C4 column using a HP 1100 chromatograph (Agilent Technologies).

3.3.5 Enzymatic hydrolysis

Native HspA was digested with trypsin in 50 mM ammonium acetate, pH 6.0 at 37°C overnight, using an enzyme/substrate ratio of 1:50 (by mass). The mixtures of tryptic peptides were either directly analyzed by mass spectrometry.

3.3.6 Mass spectral analyses

HPLC-desalted HspA samples were directly analysed by ES/MS using a Quattro Micro triple quadrupole mass spectrometer (Waters). Samples were injected into the ion source by a Harward syringe pump at a flow rate of 10 µl/min. Data were acquired and processed by the Mass Lynx software provided by the manufacturer. The instrument was calibrated by a separate injection of horse heart myoglobin (average molecular mass 16,951.5 Da); all masses are reported as average mass.

Mass mapping experiments were performed by MALDI mass spectrometry analysis using a Voyager DE PRO instrument (Applied Biosystem) operating both in linear and reflectron mode. Typically, 1 μl of analyte solution was mixed with 1 μl of a-cyano-4-hydroxycinammic acid 10 mg/ml in acetonitrile/50 mM ammonium citrate, 70:30 v/v, containing 250 fmol of bovine insulin. The mixture was applied onto the metallic sample plate and air dried. Mass calibration was performed using the quasi-molecular ions of insulin at m/z 5734.5 and a matrix peak at m/z 379.1 in linear mode and by multipoint calibration in reflectron mode using a peptide calibration mixture provided by the manufacturer. All mass values in linear mode are reported as average masses, whereas those measured in reflectron mode are reported as monoisotopic values.

3.4 Discussion

Cysteine repeats are an unique feature of *Helicobacter* HspA, that is not present in any other of the GroES homologues so far characterized.

In this study we found that, under non reducing conditions, three intramolecular disulfide bonds can be formed in HspA. The first bond involves the Cys 51 - 53 of the GroES-like domain A, whereas the remaining two bonds link together the two Cys-tandems of the Nickel-binding domain B (Cys 94-Cys111 and Cys95-Cys112). The bonds are stable under acidic conditions, since they can scramble at pH values above 6.5.

While the first disulfide bond might play a role in stabilizing the core domain A, the remaining two disulfide bridges alter dramatically the topology of the C-terminal domain, thereby generating two rings: a small, rigid 16-atoms membered ring built up from two pairs of consecutive Cys and a large flexible loop structure formed by enclosing 19 residues of the B domain. The large loop, here first identified, embodies five His leaving out of the ring the neighbouring three C-term His (*Figure* 3.11).

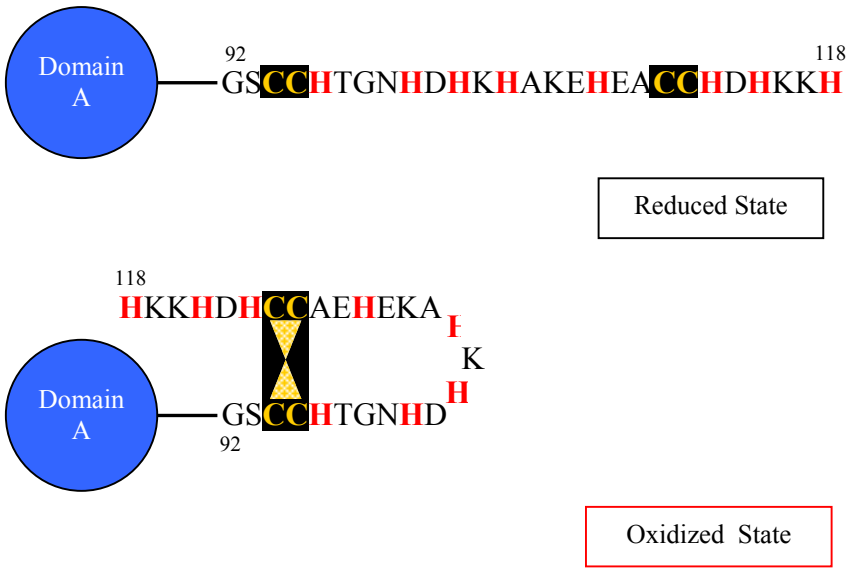


Figure 3.11: Schematic representation of the B domain in Wt HspA. Disulphide bridges are highlighted in yellow.

This sequence/structure motif is peculiar for *H.pylori* HspA, as it has never observed before; the resulting conformation would be optimal for His residues to coordinate up to 2 Ni(II) ions.

Hence, the redox state of HspA is linked to the conformational arrangement of its C-terminus, and this in turn affects the overall Ni-binding affinity of the protein.

It has been reported that, under non-reducing conditions, MBP-fused HspA, binds 2 Ni(II)/molecule with an apparent dissociation constant of 1.8 uM, whereas free HspA binds only 0.6 Ni/molecule with a lower dissociation constant (0.6 uM) (Kansau et al., 1996). So the C-terminal domain displays at least two Nickel binding states: a high-affinity binding state, in wich the oxidized cysteines force the domain B in a loop structure, and a low-affinity binding state with reduced cysteines and an open structure. It can be argued that the high-affinity binding state, with its relatively rigid loop, is involved in Ni storage, whereas the low-affinity state, open and flexible, can deliver the Ni ions to other molecular systems when required. The protein can therefore cycle between the storage/delivery states dependent on the redox and pH conditions of the environment.

Intriguingly, HspA is the only nickel chaperon wich was found to be surface-associated among the several nickel-binding and accessory proteins facilitating Ni-enzyme maturation in *H.pylori*. Recent experiments, performed through 2D-PAGE, confirmed the *H.pylori* proteome to be highly dynamic (Backert et al., 2005). In particular, HspA occurs at approximately 50% either in the soluble protein fraction or in the structure bound protein fraction, hence, the subcellular localization might vary, depending on the environmental conditions. Accordingly, our findings suggest the out of membrane HspA to possess 3 S-S bridges; whereas the cytoplasmatic protein likely exhibits a diverse Cys redox-state.

Although the two subcellular localizations that HspA experience - the cytoplasm and the cell surface- display different redox features, their unifying characteristic is a pH below neutrality (4-5 for the gastric mucosa and 6 for the HP cytoplasm). It seems that an high pH is unusual for this peculiar pathogen-host system, so a local raise in pH can be a key signal to modulate the Ni-binding affinity of the domain B. This holds particularly for the

cell surface, where the protein is embedded in an oxidizing environment at very low pH, deprived of canonical redox-balancing systems such as the thioredoxin/glutaredoxin pair. Under these conditions, HspA is thought to be mostly in its oxidized/Nickel storage state and a local increase in pH can effectively trigger the Ni release from the protein.

Surface-associated HspA could participate in the extracellular urease machinery, essential for the bacterium survival, since up to 30% of total urease associates also with the surface of intact cells upon lysis of neighboring bacteria (Ha et al., 2001). In HspA, the large loop plasticity is reduced upon Ni binding and hence may trigger a conformational switch for molecular recognition of urease complex.

Helicobacter pylori adaptation to the gastric mucosa relies on Ni (II) and hence has developed intricate *ad hoc* mechanisms related to Nickel transport, storage and delivery. The recent crystal structure determination of apo-and metallated structures of HpNikR (Dian 2006, Fauquant 2006) has revealed a complex landscape for Ni binding in *H.pylori*. Accordingly, our data provide evidence for a novel Ni binding structural arrangement in HspA, determined by a double consecutive disulphide bridges. This confirms that *H.pylori* proteins are regulated by an intricate interplay of signals depending on the structure and environment.

Section II

Integrative Approaches

The problem of biology is not to stand aghast at the complexity but to conquer it.
Sydney Brenner

Chapter 4

Critical points in pathogen bacterial networks and their implication for drug discovery

4.1 Introduction

Complex systems such as living cells have a 'robust yet fragile' feature, in which robustness in response to certain perturbation is inevitably associated with fragility in response to other perturbations (Kitano, 2007a). Finding the points of fragility of the pathogenic microbial cells is a key issue in antibacterial drug discovery. Here we describe a general approach to identify the above points of fragility in a microbial metabolic network. The local and global aspects of metabolic network analyses allow us to identify enzymes or reactions that are crucial for the survival of the organism(s), therefore directing us towards the discovery of potential drug targets.

Our strategy is currently implemented on the human gastric pathogen Helicobacter pylori 26695. The entire metabolic network of the pathogen is analyzed to find biochemically critical points, e.g. enzymes which uniquely consume and/or produce a certain metabolite. Once identified, the list of critical enzymes is filtered in order to find candidate targets wich are non-homologous with the human enzymes. Finally, the

essentiality of the identified targets is cross-validated by in silico deletion studies using flux-balance analysis (FBA) on a recent genome-scale metabolic model of H. pylori. Following this approach, we identified some enzymes which could be interesting targets for inhibition studies of *H.pylori* infection.

4.2 Points of fragility in a metabolic network

A typical metabolic network consists of reactions, metabolites and enzymes, which can be modelled using graph theory (Girvan & Newman, 2002; Jeong et al., 2000; Ma et al., 2004; Oltvai & Barabasi, 2002; Schuster et al., 2000).
The local and global connectivity of a metabolic network can be evaluated in order to identify enzymes or reactions that are crucial for the survival of the corresponding organism. In particular, we consider here the concepts of 'load points' and 'choke points' (Rahman & Schomburg, 2006):

- Load points help determine the importance of enzymes and metabolites in the biochemical network. They are defined as hot spots in the metabolic network (enzymes/metabolites), based on the ratio of number of k-shortest paths passing through a metabolite/enzyme (in/out) and number of nearest neighbour links (in/out) attached to it, compared to the average load value in the network.

- Choke points are defined as bio-chemically essential points in the network, e.g. enzymes that uniquely consume and/or produce a certain metabolite. They are ranked by the number of k-shortest paths (in/out) passing through it and the load point (in/out) on it.

Choke points are critical points in metabolic networks. Inactivation of choke points may lead to an organism's failure to produce or consume particular metabolites which could cause serious problems for fitness or survival of the organism (Yeh et al., 2004). We expect the inactivation of an enzyme that consumes a unique substrate to result in the

accumulation of the unique substrate (potentially toxic to the cell), whereas the inhibition of an enzyme that produces a unique product to result in the starvation of the unique product (potentially crippling essential cell functions). Thus, critical points may be crucial for fitness or survival of the organism and therefore can be viewed as potential drug targets.

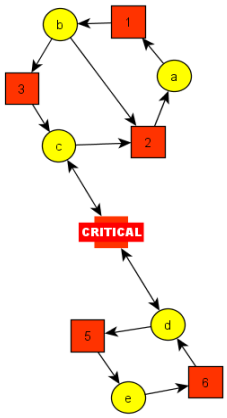


Figure 4.1: Petri-Net view of a metabolic network. The red central node is a critical point (enzyme).

For example, in Plasmodium falciparum—a parasite causing malaria in humans—a host cell enzyme 4.2.1.24 (d-aminolevulinate dehydratase; ALAD) involved in heme biosynthesis was suggested as an antimalarial target (Bonday et al., 2000). This enzyme is also a choke point enzyme and identifying such potential targets in pathogens can accelerate drug discovery. Also all three clinically validated drug targets for malaria are chokepoint enzymes. A total of 87.5% of proposed drug targets with biological evidence in the literature are chokepoint reactions (Yeh et al., 2004).

Herein, the concept of choke points was used to extract potential drug targets from the metabolic network of *Helicobacter pylori* 26695.

4.3 Results

4.3.1 Critical point analysis of the network

Probable critical points in the network have been ranked according to the number of k-shortest paths (N_{SP}) passing through it. Assuming that N_{SP} is normally distributed, candidate enzymes with N_{SP} values above the following treshold T_{SP} are selected:

$$T_{SP} = <N_{SP}> + \sigma$$

Where $<N_{SP}>$ is the mean value and σ the standard deviation.

Using these selection criteria, we screened the entire *Helicobacter pylori* 26695 metabolic reconstruction (see Methods) and subsequently identified the top 30 critical enzymes in the network.

4.3.2 Comparative evaluation of gene essentiality

Unfortunately, there are no systematic gene inactivation data available for *Helicobacter pylori* - genome-scale knockouts have been reported hitherto only for *S.cerevisiae, E.coli* and *B.subtilis*. Hence, biomedical literature mining retrieved single knockout data for 72 genes only. In order to expand and integrate information about gene essentiality in *H.pylori,* a database of essential genes was built by gathering data from single knockout experiments, global transposon mutagenesis and *in silico* metabolic studies. We merged the following datasets:

- Essential genes for HP growth in vitro (Salama et al., 2004)
  Method: Global transposon mutagenesis (Experimental)
  - 344 genes

- HP genes that contribute to gastric colonization (Salama et al., 2004)

  Method: Global transposon mutagenesis in a mouse infection model (Experimental)

  - 223 genes

- HP metabolic core genes (Almaas et al., 2005)

  Method: Flux-balance analysis (FBA) on a *H.pylori* reconstructed metabolic network in 30000 diverse simulated environments (Computational)

  - 110 genes

- HP single-knockout mutants from biomedical literature

  Method: Single deletion mutagenesis (Experimental)

  - 72 genes

Figure 4.2 provides an overall view of the above datasets, together with their overlaps. Notably, there are very few genes in common across the different lists, since they have been obtained using diverse methodologies and/or different experimental conditions.

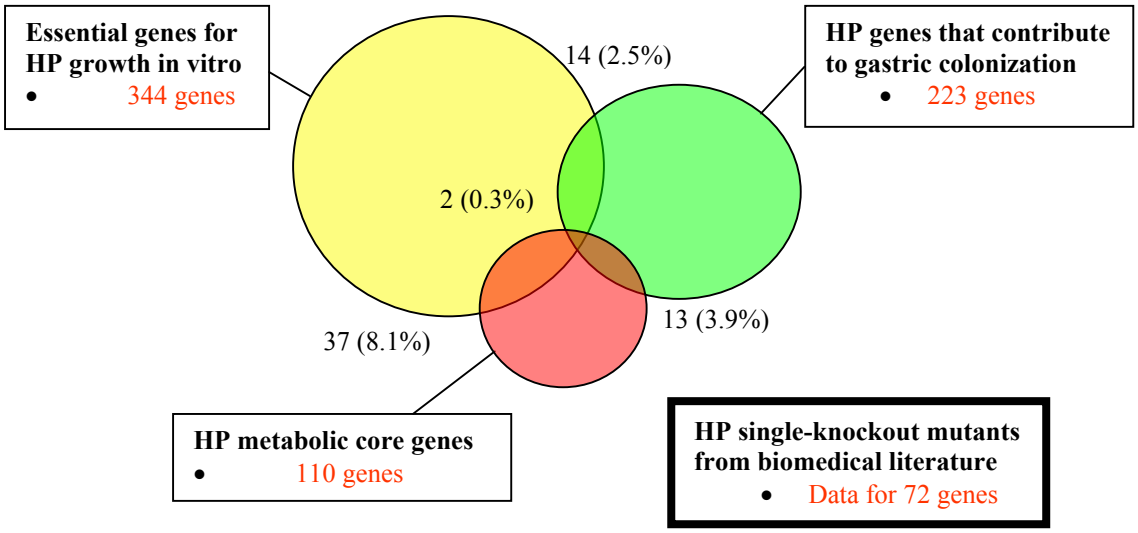


Figure 4.2: Venn diagram illustrating essential gene sets currently available for H.pylori 26695.

Probable critical enzymes are tagged and evaluated according to the essentiality data discussed here, in an attempt to extract consensus targets for the bacterium`s survival and pathogenicity (*Table* 4.1).

4.3.3 Selection of potential target enzymes

In order to confer biological meaning to the graph-based approach of finding critical points, a comparative study of such points with the human metabolic network is essential to identify possible interference of the drugs with the human metabolism which might lead to side effects. It has to be kept in mind though that presently a large number of genes have unidentified functions which could affect erroneous prediction of critical points.

the list of potential targets was filtered in two steps. First, a network based comparative study was performed between the human metabolic network and pathogen critical points to discriminate common critical enzymes. In *Table* 4.1, '+' implies that a particular enzyme acts as a critical point in the human biochemical network as well as in the pathogen whereas '-' indicates that this enzyme is only a critical point in the pathogen and not in the human biochemical network.

A homology search was then performed against human genome to find non-homologus potential drug targets from the pathogen critical points. Enzymes with a closest homologue with BLAST *e*-values $<1.0^{-5}$ were removed (*Figure* 4.3).

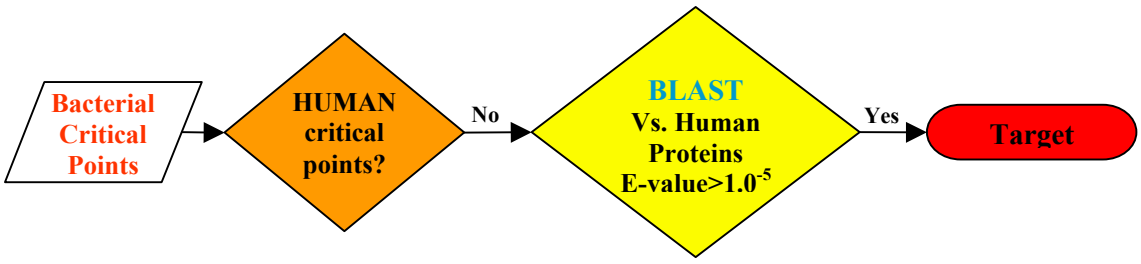


Figure 4.3: Selection flowchart for potential target enzymes.

Table 4.1 - Top 30 Critical Points in *H.Pylori* Metabolic Network. Target enzymes are highlighted in red.

Enzyme name	Gene Ids	Human Critical Point	Top BLAST hit (*e*-value > 1.0e-05)	Essential Gene Set[a]
Adenine phosphoribosyltransferase	HP0572	+	2E-24	B
Acetate-CoA ligase	HP1045	+	1E-161	
Cystathionine gamma-synthase	HP0106	-	9E-97	A
Transketolase	HP1088	+	8E-49	A
Aspartate transaminase	HP0672	+	5E-38	A
Adenylosuccinate synthase	HP0255	-	9E-90	A
Succinate dehydrogenase, iron-sulfur subunit	HP0191	-	1E-24	
Succinate dehydrogenase, flavoprotein subunit	HP0192	-	1E-74	
Succinate dehydrogenase, cyt. b subunit	**HP0193**	**-**	**9E-01**	
Succinyldiaminopimelate transaminase	HP0624	-	3E-14	A
Adenylosuccinate lyase	HP1112	-	2E-65	A
Nucleoside-diphosphate kinase	HP0198	-	2E-29	A,D
Aconitate hydratase	**HP0779**	**-**	**3E-05**	
GTP cyclohydrolase I	HP0928	+	2E-44	A,B,D
Isocitrate dehydrogenase (NADP+)	HP0027	-	3E-42	
Purine-nucleoside phosphorylase	HP1178	+	6E-01	
Serine O-acetyltransferase	**HP1210**	**-**	**8E+00**	**A**
Glutamate dehydrogenase (NADP+)	HP0380	-	9E-33	
GMP synthase (glutamine-hydrolysing)	HP0409	-	2E-78	A,B
Phospho-N-acetylmuramoyl-pentapeptide-tr.	**HP0493**	**-**	**9E-03**	**A,D**
L-serine ammonia-lyase	HP0132	+	9E-01	
Branched-chain-amino-acid transaminase	HP1468	-	2E-40	A,D
IMP dehydrogenase	HP0829	-	5E-86	A
Cysteine synthase	HP0107	-	7E-46	A
5'-nucleotidase	HP0930	+	1E+00	
Undecaprenyldiphospho-muramoylp-peptide	**HP1155**	**-**	**2E-02**	**A,D**
1-pyrroline-5-carboxylate dehydrogenase	HP0056	+	1E-37	
Succinyl-diaminopimelate desuccinylase	**HP0212**	**-**	**3E-05**	**A,C,D**
2,3,4,5-tetrahydropyridine-N-succinyltr.	**HP0626**	**-**	**3E+00**	**A,D**
Dihydrodipicolinate reductase	**HP0510**	**-**	**2E+00**	**A,B,D**

[a] A, exp. single deletion mutant; B, essential gene for HP growth *in vitro*; C, essential gene for gastric colonisation; D, metabolic core gene

4.3.4 Cross-validation of target essentiality using Flux Balance Analysis (FBA)

A simulation of growth phenotypes resulting from single and double deletion of target enzymes was carried out on a recent genome-scale metabolic reconstruction of *H.pylori*, in order to cross-validate the essentiality of selected candidates (*Table* 4.2). *In silico* deletion studies were performed using costraint-based reconstruction and analysis (COBRA tecnique) on the H. pylori iIT341 GSM/GPR *in silico* metabolic model.

Table 2 –Target enzymes validated by in silico deletion experiments.

Enzyme name	Gene Ids	Essential Gene Set	COBRA Single Deletion Phenotype	COBRA Double Deletion Partner
Succinate dehydrogenase, cytochrome b subunit	HP0193		Growth	HP0724: Lethal
Aconitate hydratase	HP0779		Growth	-
Serine O-acetyltransferase	**HP1210**	**A**	**Lethal**	-
Phospho-N-acetylmuramoyl-pentapeptide-transferase	**HP0493**	**A,D**	**Lethal**	-
Undecaprenyldiphospho-muramoylpentapeptide	**HP1155**	**A,D**	**Lethal**	-
Succinyl-diaminopimelate desuccinylase	**HP0212**	**A,C,D**	**Lethal**	-
2,3,4,5-tetrahydropyridine- N-succinyltransferase	**HP0626**	**A,D**	**Lethal**	-
Dihydrodipicolinate reductase	**HP0510**	**A,B,D**	**Lethal**	-

[a] A, exp. single deletion mutant; B, essential gene for HP growth *in vitro*; C, essential gene for gastric colonisation; D, metabolic core gene

4.4 Data and methods

We consider a reconstructed metabolic network (enzymes/metabolites) of Helicobacter Pylori 26695, built using the LIGAND database from KEGG and the BRENDA enzyme database. We used the Pathway Hunter algorithm (Rahman & Schomburg, 2006) to estimate and rank the bio-chemically critical points in the network. A database of essential genes was built using data from single knockout experiments and from global transposon mutagenesis (Baldwin et al., 2007; Salama et al., 2004). The homology search between the human and *H.pylori* enzymes was executed using BLAST and enzymes with a closest homologue with e-values < 1.0e-02 were removed. In silico deletion studies were performed using costraint-based reconstruction and analysis (COBRA technique) (Becker et al., 2007) on the H.pylori iIT341 GSM/GPR *in silico* metabolic model (Thiele et al., 2005).

4.5 Discussion

Drug target identification based on 'omics' networks (di Bernardo et al., 2005; Giaever et al., 2004; Holzhutter & Holzhutter, 2004; Yeh et al., 2004) is a very promising approach that has only recently become possible. The concept of choke points (Dawson & Elliott, 1980) in a given network contributes effectively in identification of the lethality/bottleneck (here potential drug targets) in a network.

Often drug targets are identified by a unique pathogen-specific metabolic activity, as in the case of reverse transcriptase in the case of HIV (Imamichi, 2004). However, the screening of the entire metabolic network of the pathogen to find choke point-based potential drug candidates followed by a comparative study with human metabolic network provides additional targets. Examples are the anti-malarial drugs (Sixsmith et al., 1984) pyrimethamine and cycloguanil targeting a choke point enzyme dihydrofolate

reductase (1.5.1.3) (also a human homologue) in P.falciparum with some side effects on humans but lethal to the parasite.

In this study, we proposed a number of potential target enzymes, based on the analysis of the top 30 critical points in *H.pylori* 26695 metabolic network. Cross-validation, through experimental data and in silico single-gene deletion studies, of the eight candidate enzymes (*Table* 4.1) has allowed us the identification of six essential enzymes. Indeed, removal of these six enzymes results in lethal growth phenotypes (*Table* 4.2). Interestingly, the enzyme Dihydrodipicolinate reductase (EC: 1.3.1.26) has a homologue of known 3D structure: DapB from E.Coli (PDB code 1ARZ, 40% identity). This makes rational drug design on this enzyme feasible.

Double-deletion studies, performed for both nonessential enzymes (*Table* 4.2), have identified a lethal growth phenotype for the mutant lacking both HP0193 and HP0724. It is worth noting that HP0724 (anaerobic C4-dicarboxylate transport protein) is non homologus to human proteins, with a BLAST e-value > 1.0-5, i.e. a good candidate for drug discovery.

We here identified seven enzymes which are ideal targets for inhibition studies of *H.pylori* infection. Indeed, (i) none of these enzymes are critical points in the human metabolic network and (ii) they share a low identity with the human genome (*Table* 4.1). This means that blocking these enzymes might selectively affect the pathogen, without interfering with human metabolism.

Chapter 5

Target-sets discovery for multi-target drugs – A robustness- based approach

5.1 Introduction

5.1.1 Multiple weak attacks confuse complex systems

Recent studies on biological networks suggested that multiple weak attacks on selected targets are inevitably more efficient than the knockout of a single target. This might provide theoretical accounts for the recent success of multi-target drugs (Agoston et al., 2005; Csermely et al., 2005; Korcsmaros et al., 2007).

Distributed attacks are not only better because they affect the network at more sites, they can, particularly if scattered over the entire network, perturb complex systems more than concentrated attacks even if the number of targeted interactions is the same (*Figure* 5.1). This may help the design of new antimicrobial drugs, where network damage corresponds well to the desired drug action.

Here we describe a possible approach to find potential target-sets for multi-target drugs. The strategy is implemented on the human gastric pathogen Helicobacter pylori 26695 (Beswick et al., 2006). We used the known metabolic and protein interaction data to build an integrated biomolecular network of the pathogen. The network was subsequently

screened to find central points of network communication, e.g. hubs, bridges with high betweenness centrality and overlaps of network communities.

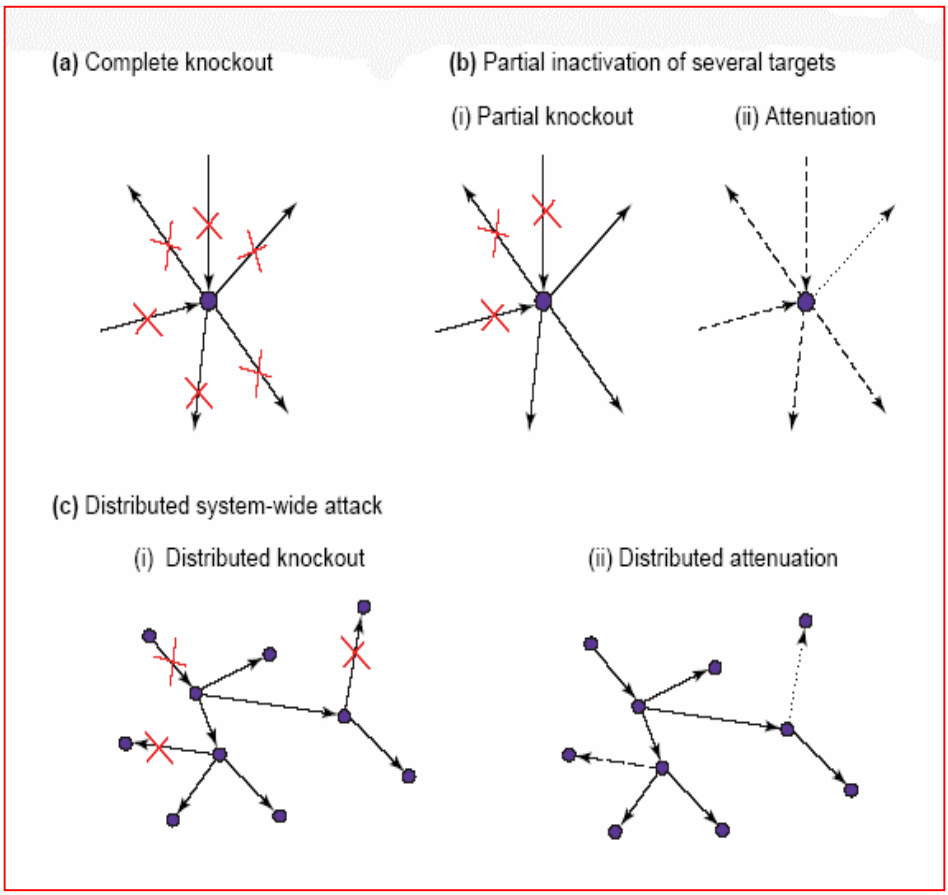


Figure 5.1: Partial target inhibition strategies.

Once identified, the selected enzymes were classified on the basis of available data about cellular function and essentiality. In order to predict the growth phenotype associated to the partial inactivation of candidate targets, robustness analysis was performed on small groups of selected enzymes using flux balance analysis (FBA) on a recent genome-scale metabolic model of H.pylori (Thiele et al., 2005). The initial results may help to restrict the initial target-pool in search of target-sets for multi-target drugs.

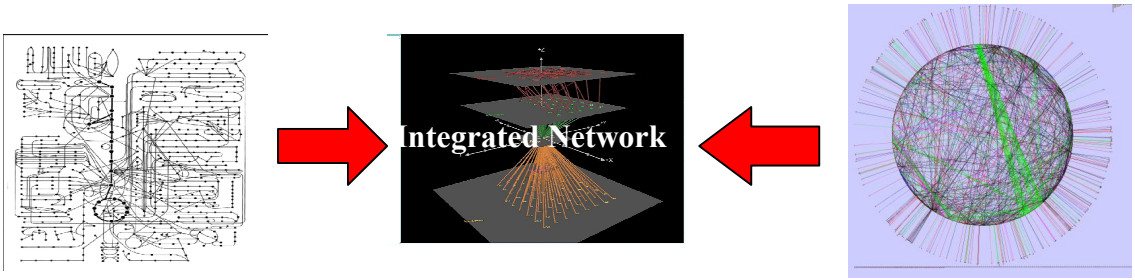
5.1.2 Robustness Analysis

In general, robustness means the persistence of a system's characteristic behavior under perturbation or conditions of uncertainty. Robustness is, hence, defined for a specific system, which, however, may have arbitrary structural and behavorial features. The concept is closely related to stability in dynamical systems theory, but usually employed with respect to a broader class of phenomena (Carlson & Doyle, 2002; Kitano, 2007b). In metabolic modelling (Flux-balance analysis), robustness outlines the effect of reducing flux through a single reaction on growth. This is a network property of great interest, since it allows the computation of how an objective of interest (e.g., growth rate) changes as the flux through one or more specific reactions of interest varies in magnitude.

5.2 Results

5.2.1 Integrated network building

The available metabolic and protein interaction data are used to build an integrated biomolecular network of the pathogen (*Figure* 5.2).



Reconstructed Metabolic Network
H.Pylori 26695
- Model: *i*IT341 GSM/GPR[3]
  - 341 enzymes, 554 reactions, 485 metabolites

Protein Interaction network
H.Pylori 26695
- 1539 nodes
- 2523 edges

Figure 5.2: Integrated network building.

5.2.2 Analysis of central elements for network communication

Local network topology: HUBS

Hubs are connection-rich elements, with a much greater number of neighbours than average. The Pathway Hunter Algorithm (Rahman & Schomburg, 2006) was used to catch enzyme hubs within the H.pylori metabolic network.

Global network topology: LOAD POINTS

Load points are nodes with high betweenness centrality. Betweenness centrality of a node refers to the number of shortest paths between any two elements of the network across the given node. Here, probable load points are ranked according to the number of metabolic shortest paths passing through them, yet using the Pathway Hunter Algorithm (Rahman & Schomburg, 2006).

Metabolic modules: CORRELATED REACTIONS SETS

Correlated reaction sets define modules of reactions in networks that have to be co-utilized in precise stoichiometric ratios within the metabolic network. Here, co-sets and corresponding enzymes mes are identified using the Flux Coupling Finder procedure (Burgard et al., 2004) on the H.pylori *i*IT341 metabolic model (Thiele et al., 2005).

Overall network architecture: OVERLAPS OF COMMUNITIES

The graph of communities is a network emerging above the level of individual nodes in the hierarchical organisation of a complex system. In this graph the nodes correspond to communities (highly interconnected subgraphs, also called modules or clusters), and the links refer to members shared by two communities (overlaps of communities). The interaction network was dissected to communities of overlapping modules using the CFinder algorithm (Adamcsek et al., 2006). Related enzymes were subsequently identified.

Table 1–Map of key enzymes for *H.pylori* integrated network communication. Enzyme features are marked in blue.

Enzyme name	Gene Ids	Hub	Load Point	Co-sets	Communities	Community Overlap	Essential Gene Set[a]
Phosphoribosylpyrophosphate synthetase	HP0742						C,D
6-Phosphogluconate dehydratase	HP1100						A
UDP-3-O-glycosamine acyltransferase	HP0196						B,D
Glutamyl-tRNA reductase	HP0239						
4-Aminobenzoate synthase	HP0587						B,C,D
LipidA_HP disaccaride synthase	HP0867						B,C,D
Kdo-Lipid IV (A) synthase	HP0957						B,D
UDP-3-O-acetylglucosamine deacetylase	HP1052						B,D
Quinolinate synthase	HP1356						B
UDP-N-acetylglucosamine acyltransferase	HP1375						B,D
Urea transmembrane channel	HP0071						C
Iron (II) transporter - ABC system	HP0687						A
Dihydroneopterin mono-P dephosphorylase	HP1228						D
Aconitase	HP0779						
Tryptophan synthase TrpA	HP1277						C,D
Tryptophan synthase TrpB	HP1278						C,D
Dihydrofolate synthase	HP1545						B,D
Phosphoglucosamine mutase	HP0075						A,B,D
2,6-diaminopimeloyl-D-alanine synthetase	HP0740						A,B,D
2-oxobutanoate hydroxymetyltransferase	HP1058						B,C,D
Pantothenate synthase	HP0006						B,D
UDP-N-acetylmuramoyl - synthase	HP0034						B,D
Aspartate-semialdehyde dehydrogenase	HP1189						

[a] A, exp. single deletion mutant; B, essential gene for HP growth *in vitro*; C, essential gene for gastric colonisation; D, *in silico* deletion.

5.2.4 Robustness analysis on selected enzymes

As a preliminary result, robustness analysis for HP1189 and HP0106 (an enzyme hub not reported in *Table* 5.1) is displayed in *Figure* 5.3.

Both enzymes are nonessential in single-deletion studies. The growth rate is sustained near the optimal value over a range of values for both HP0106 and HP1189 reactions, indicating network robustness with respect to flux changes in both reactions. However, when the two reaction fluxes decrease simultaneously the predicted growth rate goes rapidly to zero. A complete deletion of the two enzymes would be predicted to result in a lethal phenotype.

Similar results have been obtained so far for the nonessential enzyme couples HP0577-HP0183, HP1108-HP1180 and HP1179-HP0574.

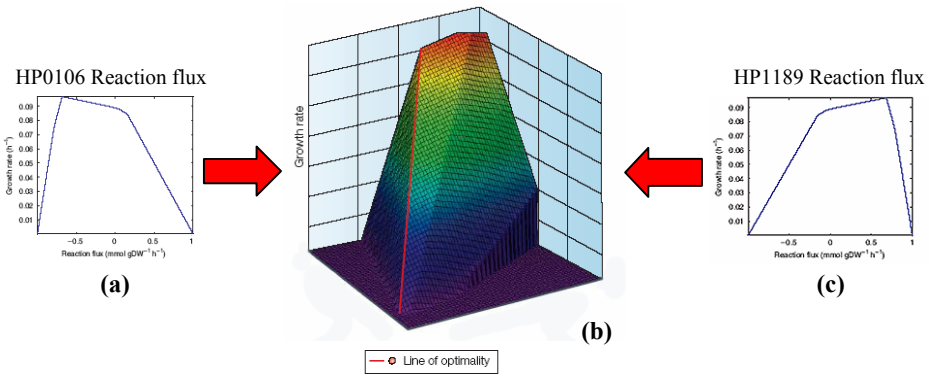


Figure 5.3: Robustness analysis. Predicted optimal growth rate of H.pylori in rich media conditions[3] as a function of the flux through the (**a**) HP0106 reaction, (**c**) HP1189 reaction, (**b**) both reactions simultaneously.

5.3 Data and Methods

We consider a genome-scale metabolic network of Helicobacter Pylori 26695 wich accounts for 341 enzymes and 485 metabolites (Thiele et al., 2005). Protein interaction data were integrated adding 849 yeast-two-hybrid protein-protein interactions (Rain et al., 2001) and 1674 functional associations from Prolinks database (Bowers et al., 2004). The Pathway Hunter algorithm (Rahman & Schomburg, 2006) was used to find enzyme hubs and bridges with high betweenness centrality (load points) in the metabolic network. Metabolic modules were identified according to correlated reaction sets (Burgard et al., 2004). The interaction network was dissected to communities of overlapping modules using the CFinder algorithm (Adamcsek et al., 2006). A database of essential genes was built collecting data from single knockout experiments, global transposon mutagenesis (Baldwin et al., 2007; Salama et al., 2004) and *in silico* deletion studies on H.pylori 26695 (Thiele et al., 2005).

Robustness analysis was performed using costraint-based reconstruction and analysis (COBRA technique) (Becker et al., 2007) on the H.pylori iIT341 GSM/GPR *in silico* metabolic model (Thiele et al., 2005).

5.4 Discussion

The development of effective multi-target, low affinity drugs wich affect the complex equilibrium of whole cellular networks is a fascinating concept. The search of target-combinations for a successful system attack is a daunting task, however.

Here we proposed a network strategy to reduce the pool of initial targets, and a robustness-based approach to evaluate *in silico* the phenotypic effect triggered by the partial inactivation of selected candidates. In this preliminary study we found that four H.pylori nonessential enzyme pairs can dramatically alter the predicted growth of the

pathogen when the two enzymes in a pair are partially inhibited at the same time. Ongoing research on this topic include: additional integration of biological data to further decrease the initial target-pool, and robustness analysis on bigger groups of selected enzymes. The aim is to better understand the indirect network effects that lie at the heart of multi-target drug action and, ultimately, how multiple weak hits can perturb complex biological systems.

Chapter 6

Modeling host-pathogen interactions: toward an executable model of the EGFR-signaling process core

6.1 Introduction

The epidermal growth factor receptor (EGFR) signaling network is one of the most extensively studied areas of signal transduction, since it regulates growth, survival, proliferation and differentiation in mammalian cells. The overall architecture of the EGFR network is a bow-tie (or hourglass) structure, a hallmark of robust evolvable systems that emerges at various scales in complex biological phenomena like metabolism, signal transduction, transcription and translation. A typical bow-tie architecture comprises a variety of redundant inputs and outputs that are connected by a conserved core process. In the EGFR network, the core process is the heart of the signaling system, as it plays a central role in downstream signaling cascades to gene expression through activation of multiple transcription factors. It consists in a dense array of molecules and interactions wich are tightly coupled to each other.

Herein, we attempt to build an executable model of the EGFR-signaling core process using a process algebra approach. To this end, a small set of EGFR core molecules and their interactions is tentatively translated in a BetaWB model. BetaWB is a framework for modelling and simulating biological precesses based on Beta-binders language and its stochastic extension.

Once obtained, the computational model of the EGFR core process can be used to test and compare hypotheses regarding the principles of operation of the signaling network, i.e. how the EGFR network generates different responses for each set of combinatorial stimuli. In particular, probabilistic model checking can be used to explore the states and possible state changes of the computational model, whereas stochastic simulation (corresponding to the execution of the BetaWB model) may give quantitative insights into the dynamic behaviour of the system in response to different stimuli. Information from the above tecniques allows model validation through comparison within the experimental data available in the literature.

The inherent compositionality of the process algebra modeling approach enables further expansion of the EGFR core model, as well as the study of its behavior under specific perturbations, such as drugs or environmental factors. This latter aspect might be of great value for biomedical research, as signaling through the EGF receptors is intricately involved in human cancer and in many other pathological conditions.

6.2 The bow-tie structure in biological systems

The notion of robustness has recently received considerable interest in diverse fields for which the existence of complex networks is characteristic. Examples include the internet, social networks, and biology (Strogatz, 2001; Stelling et al., 2004b; Kitano, 2004a).It is a fundamental and ubiquitously observed property of complex systems that cannot be understood by looking at the individual components in isolation. Kitano defined robustness as ' a property that allows a system to maintain its functions despite external and internal perturbations' (Kitano, 2004). Robustness is attained by several underlying principles that are universal to both biological organisms and sophisticated engineering

systems, such as system control, alternative (or fail-safe) mechanisms, modularity and decoupling. The architectural features of a modularized bow-tie structure are optimal for enhancing the robustness of the various aspects of a system.

The overall architecture that meets the above requirements is probably a modularized nested bow-tie, or hour-glass, structure. This organisational framework has many inputs and outputs that are connected through a conserved core and versatile weak linkage with the extensive system control governing the system's dynamics. Core processes and versatile interfaces overlap or merge in some cases (*Figure* 6.1). This bow-tie structure occurs in various aspects of biological systems, from global structure to specific functional levels such as metabolism, signal transduction, transcription and translation (*Figure* 6.2).

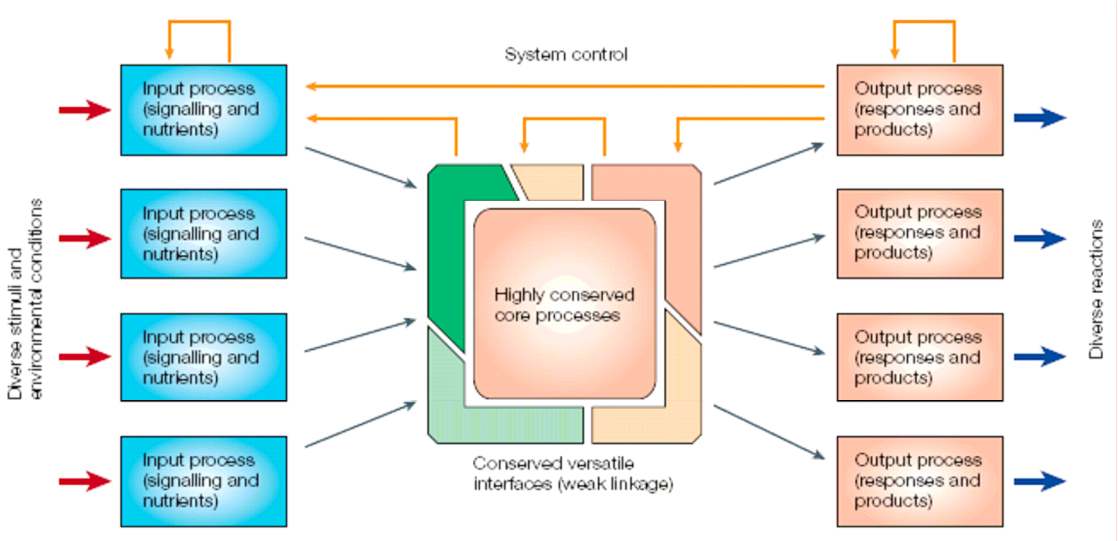


Figure 6.1: General features of bow tie structures.

In signal transduction, diverse stimuli are initially received by receptors, different isoforms of G-proteins are activated, but converge mainly to second messengers that have a limited variety and cause weak linkage. Then, modulations in second messengers influence core processes to trigger expression of different genes and, ultimately, different reactions. However, this process is not a simple flow of information, as extensive local and global feedback regulations are imposed at every step. During metabolism, diverse

nutrients are processed into precursors that core metabolic pathways then covert into basic cellular 'currencies' such as ATP and NADH, as well as activating biosynthesic pathways to produce amino acids, nucleotides, sugar, and so on. Transcription and translation also represent structures where common machineries are used to decode a wide range of genetic information and produce diverse proteins, but versatile mechanisms themselves make up the conserved core. Various processes are interfaced with core processes through versatile interfaces, so that novel processes can be added and removed easily without seriously affecting other parts of the system.

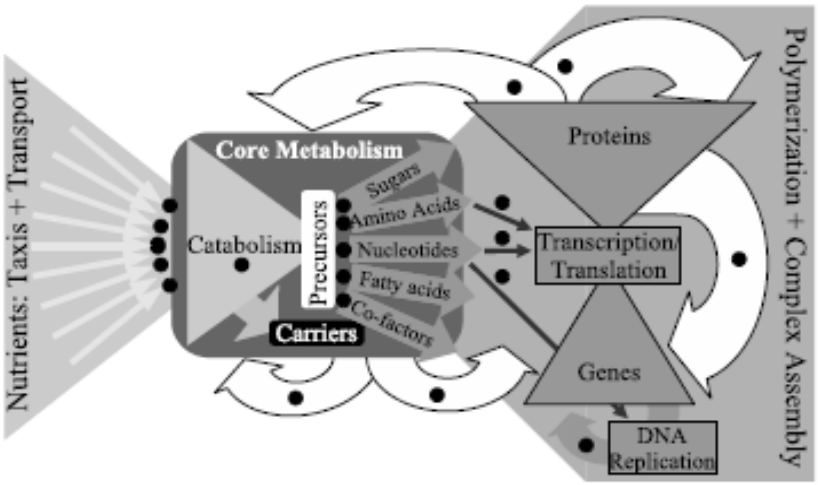


Figure 6.2: Bow tie abstraction of cellular organization.
Open arrows denote cellular regulation and control. Involvement of carriers such as ATP and NAD(P)H in individual processes is indicated by •.

References

(1994). Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum* 61, 1-241.

Adamcsek, B., Palla, G., Farkas, I. J., Derenyi, I. & Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021-3.

Agoston, V., Csermely, P. & Pongor, S. (2005). Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. *Phys Rev E Stat Nonlin Soft Matter Phys* 71, 051909.

Alm, R. A., Bina, J., Andrews, B. M., Doig, P., Hancock, R. E. & Trust, T. J. (2000). Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infect Immun* 68, 4155-68.

Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F. & Trust, T. J. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* 397, 176-80.

Almaas, E., Oltvai, Z. N. & Barabasi, A. L. (2005). The activity reaction core and plasticity of metabolic networks. *PLoS Comput Biol* 1, e68.

Amoresano, A., Orru, S., Siciliano, R. A., De Luca, E., Napoleoni, R., Sirna, A. & Pucci, P. (2001). Assignment of the complete disulphide bridge pattern in the human recombinant follitropin beta-chain. *Biol Chem* 382, 961-8.

Amoresano, A., Pucci, P., Duro, G., Colombo, P., Costa, M. A., Izzo, V., Lamba, D. & Geraci, D. (2003). Assignment of disulphide bridges in Par j 2.0101, a major allergen of Parietaria judaica pollen. *Biol Chem* 384, 1165-72.

Appelmelk, B. J., Martin, S. L., Monteiro, M. A., Clayton, C. A., McColm, A. A., Zheng, P., Verboom, T., Maaskant, J. J., van den Eijnden, D. H., Hokke, C. H., Perry, M. B., Vandenbroucke-Grauls, C. M. & Kusters, J. G. (1999). Phase variation in Helicobacter pylori lipopolysaccharide due to changes in the lengths of poly(C) tracts in alpha3-fucosyltransferase genes. *Infect Immun* 67, 5361-6.

Appelmelk, B. J., Negrini, R., Moran, A. P. & Kuipers, E. J. (1997). Molecular mimicry between Helicobacter pylori and the host. *Trends Microbiol* 5, 70-3.

Aspinall, G. O. & Monteiro, M. A. (1996). Lipopolysaccharides of Helicobacter pylori strains P466 and MO19: structures of the O antigen and core oligosaccharide regions. *Biochemistry* 35, 2498-504.

Aspinall, G. O., Monteiro, M. A., Pang, H., Walsh, E. J. & Moran, A. P. (1996). Lipopolysaccharide of the Helicobacter pylori type strain NCTC 11637 (ATCC 43504): structure of the O antigen chain and core oligosaccharide regions. *Biochemistry* 35, 2489-97.

Austin, J. W., Doig, P., Stewart, M. & Trust, T. J. (1992). Structural comparison of urease and a GroEL analog from Helicobacter pylori. *J Bacteriol* 174, 7470-3.

Backert, S., Kwok, T., Schmid, M., Selbach, M., Moese, S., Peek, R. M., Jr., Konig, W., Meyer, T. F. & Jungblut, P. R. (2005). Subproteomes of soluble and structure-bound Helicobacter pylori proteins analyzed by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* 5, 1331-45.

Bader, G. D. & Hogue, C. W. (2002). Analyzing yeast protein-protein interaction data obtained from different sources. *Nat Biotechnol* 20, 991-7.

Baldwin, D. N., Shepherd, B., Kraemer, P., Hall, M. K., Sycuro, L. K., Pinto-Santini, D. M. & Salama, N. R. (2007). Identification of Helicobacter pylori genes that contribute to stomach colonization. *Infect Immun* 75, 1005-16.

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O. & Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2, 727-38.

Benoit, S. & Maier, R. J. (2003). Dependence of Helicobacter pylori urease activity on the nickel-sequestering ability of the UreE accessory protein. *J Bacteriol* 185, 4787-95.

Berg, D. E., Hoffman, P. S., Appelmelk, B. J. & Kusters, J. G. (1997). The Helicobacter pylori genome sequence: genetic factors for long life in the gastric mucosa. *Trends Microbiol* 5, 468-74.

Beswick, E. J., Suarez, G. & Reyes, V. E. (2006). H pylori and host interactions that influence pathogenesis. *World J Gastroenterol* 12, 5599-605.

Bode, G., Malfertheiner, P., Lehnhardt, G., Nilius, M. & Ditschuneit, H. (1993). Ultrastructural localization of urease of Helicobacter pylori. *Med Microbiol Immunol* 182, 233-42.

Bonday, Z. Q., Dhanasekaran, S., Rangarajan, P. N. & Padmanaban, G. (2000). Import of host delta-aminolevulinate dehydratase into the malarial parasite: identification of a new drug target. *Nat Med* 6, 898-903.

Boudker, O., Todd, M. J. & Freire, E. (1997). The structural stability of the co-chaperonin GroES. *J Mol Biol* 272, 770-9.

Bowers, P. M., Pellegrini, M., Thompson, M. J., Fierro, J., Yeates, T. O. & Eisenberg, D. (2004). Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol* 5, R35.

Bukholm, G., Tannaes, T., Nedenskov, P., Esbensen, Y., Grav, H. J., Hovig, T., Ariansen, S. & Guldvog, I. (1997). Colony variation of Helicobacter pylori: pathogenic potential is correlated to cell wall lipid composition. *Scand J Gastroenterol* 32, 445-54.

Burgard, A. P., Nikolaev, E. V., Schilling, C. H. & Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 14, 301-12.

Bury-Mone, S., Skouloubris, S., Labigne, A. & De Reuse, H. (2001). The Helicobacter pylori UreI protein: role in adaptation to acidity and identification of residues essential for its activity. *Mol Microbiol* 42, 1021-1034.

Cao, P., McClain, M. S., Forsyth, M. H. & Cover, T. L. (1998). Extracellular release of antigenic proteins by Helicobacter pylori. *Infect Immun* 66, 2984-6.

Carlson, J. M. & Doyle, J. (2002). Complexity and robustness. *Proc Natl Acad Sci U S A* 99 Suppl 1, 2538-45.

Censini, S., Lange, C., Xiang, Z., Crabtree, J. E., Ghiara, P., Borodovsky, M., Rappuoli, R. & Covacci, A. (1996). cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A* 93, 14648-53.

Cohen, I. R. (2000). *Tending Adam's garden: evolving the cognitive immune self*, Academic Press, London, UK.

Cooke, C. L., Huff, J. L. & Solnick, J. V. (2005). The role of genome diversity and immune evasion in persistent infection with Helicobacter pylori. *FEMS Immunol Med Microbiol* 45, 11-23.

Covacci, A., Telford, J. L., Del Giudice, G., Parsonnet, J. & Rappuoli, R. (1999). Helicobacter pylori virulence and genetic geography. *Science* 284, 1328-33.

Cowing, D. W., Bardwell, J. C., Craig, E. A., Woolford, C., Hendrix, R. W. & Gross, C. A. (1985). Consensus sequence for Escherichia coli heat shock gene promoters. *Proc Natl Acad Sci U S A* 82, 2679-83.

Craig, E. A., Gambill, B. D. & Nelson, R. J. (1993). Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol Rev* 57, 402-14.

Csermely, P., Agoston, V. & Pongor, S. (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci* 26, 178-82.

Cussac, V., Ferrero, R. L. & Labigne, A. (1992a). Expression of Helicobacter pylori urease genes in Escherichia coli grown under nitrogen-limiting conditions. *J. Bacteriol.* 174, 2466-2473.

Cussac, V., Ferrero, R. L. & Labigne, A. (1992b). Expression of Helicobacter pylori urease genes in Escherichia coli grown under nitrogen-limiting conditions. *J Bacteriol* 174, 2466-73.

Dawson, S. V. & Elliott, E. A. (1980). Use of the choke point in the prediction of flow limitation in elastic tubes. *Fed Proc* 39, 2765-70.

di Bernardo, D., Thompson, M. J., Gardner, T. S., Chobot, S. E., Eastwood, E. L., Wojtovich, A. P., Elliott, S. J., Schaus, S. E. & Collins, J. J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat Biotechnol* 23, 377-83.

Dian, C., Schauer, K., Kapp, U., McSweeney, S. M., Labigne, A. & Terradot, L. (2006). Structural basis of the nickel response in Helicobacter pylori: crystal structures of HpNikR in Apo and nickel-bound states. *J Mol Biol* 361, 715-30.

Doig, P., Exner, M. M., Hancock, R. E. & Trust, T. J. (1995). Isolation and characterization of a conserved porin protein from Helicobacter pylori. *J Bacteriol* 177, 5447-52.

Doig, P. & Trust, T. J. (1994). Identification of surface-exposed outer membrane antigens of Helicobacter pylori. *Infect Immun* 62, 4526-33.

Dorrell, N. & Wren, B. W. (1998). From genes to genome biology: a new era in Helicobacter pylori research. *Gut* 42, 451-3.

Dunn, B. E., Roop, R. M., 2nd, Sung, C. C., Sharma, S. A., Perez-Perez, G. I. & Blaser, M. J. (1992). Identification and purification of a cpn60 heat shock protein homolog from Helicobacter pylori. *Infect Immun* 60, 1946-51.

Dunn, B. E., Vakil, N. B., Schneider, B. G., Miller, M. M., Zitzer, J. B., Peutz, T. & Phadnis, S. H. (1997). Localization of Helicobacter pylori urease and heat shock protein in human gastric biopsies. *Infect Immun* 65, 1181-8.

Eamranond, P. P., Torres, J., Munoz, O. & Perez-Perez, G. I. (2004). Age-specific immune response to HspA in Helicobacter pylori-positive persons in Mexico. *Clin Diagn Lab Immunol* 11, 983-5.

Efroni, S. & Cohen, I. R. (2002). Simplicity belies a complex system: a response to the minimal model of immunity of Langman and Cohn. *Cell Immunol* 216, 23-30.

Ehrenman, G. (2005). Mining what others miss: highlighting the subtleties in 10e12 bytes of data, technology tries to clear up its own complex mess. *Mechanical Engineering-CIME* 127, 26-31.

Ellis, R. J. & van der Vies, S. M. (1991). Molecular chaperones. *Annu Rev Biochem* 60, 321-47.

Evans, D. J., Jr. & Evans, D. G. (2000). Helicobacter pylori adhesins: review and perspectives. *Helicobacter* 5, 183-95.

Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. (1989). The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. *J Bacteriol* 171, 1379-85.

Ferrero, R. L., Cussac, V., Courcoux, P. & Labigne, A. (1992). Construction of isogenic urease-negative mutants of Helicobacter pylori by allelic exchange. *J Bacteriol* 174, 4212-7.

Ferrero, R. L., Thiberge, J. M., Huerre, M. & Labigne, A. (1994). Recombinant antigens prepared from the urease subunits of Helicobacter spp.: evidence of protection in a mouse model of gastric infection. *Infect Immun* 62, 4981-9.

Ferrero, R. L., Thiberge, J. M., Kansau, I., Wuscher, N., Huerre, M. & Labigne, A. (1995). The GroES homolog of Helicobacter pylori confers protective immunity against mucosal infection in mice. *Proc Natl Acad Sci U S A* 92, 6499-503.

Fleck, L. (1979). *Genesis and development of a scientific fact*, University of Chicago Press, Chicago, IL.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. & et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269, 496-512.

Frenck, R. W., Jr. & Clemens, J. (2003). Helicobacter in the developing world. *Microbes Infect* 5, 705-13.

Galliano, M., Minchiotti, L., Campagnoli, M., Sala, A., Visai, L., Amoresano, A., Pucci, P., Casbarra, A., Cauci, M., Perduca, M. & Monaco, H. L. (2003). Structural and biochemical characterization of a new type of lectin isolated from carp eggs. *Biochem J* 376, 433-40.

Ge, R., Sun, X., Gu, Q., Watt, R. M., Tanner, J. A., Wong, B. C., Xia, H. H., Huang, J. D., He, Q. Y. & Sun, H. (2007). A proteomic approach for the identification of bismuth-binding proteins in Helicobacter pylori. *J Biol Inorg Chem*.

Ge, Z. & Taylor, D. E. (1999). Contributions of genome sequencing to understanding the biology of Helicobacter pylori. *Annu Rev Microbiol* 53, 353-87.

Gebert, B., Fischer, W., Weiss, E., Hoffmann, R. & Haas, R. (2003). Helicobacter pylori vacuolating cytotoxin inhibits T lymphocyte activation. *Science* 301, 1099-1102.

Gerrits, M. M., van Vliet, A. H., Kuipers, E. J. & Kusters, J. G. (2006). Helicobacter pylori and antimicrobial resistance: molecular mechanisms and clinical implications. *Lancet Infect Dis* 6, 699-709.

Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D. F., Chu, A. M., Jordan, M. I., Arkin, A. P. & Davis, R. W. (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci U S A* 101, 793-8.

Girvan, M. & Newman, M. E. (2002). Community structure in social and biological networks. *Proc Natl Acad Sci U S A* 99, 7821-6.

Gupta, R. S. (1995). Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol Microbiol* 15, 1-11.

Ha, N. C., Oh, S. T., Sung, J. Y., Cha, K. A., Lee, M. H. & Oh, B. H. (2001). Supramolecular assembly and acid resistance of Helicobacter pylori urease. *Nat Struct Biol* 8, 505-9.

Hand, D. J., Blunt, G., Kelly, M. G. & Adams, N. M. (2000). Data mining for fun and profit. *Stat. Sci.* 15, 111-131.

Haque, M., Hirai, Y., Yokota, K., Mori, N., Jahan, I., Ito, H., Hotta, H., Yano, I., Kanemasa, Y. & Oguma, K. (1996). Lipid profile of Helicobacter spp.: presence of cholesteryl glucoside as a characteristic feature. *J Bacteriol* 178, 2065-70.

Harris, A. G., Wilson, J. E., Danon, S. J., Dixon, M. F., Donegan, K. & Hazell, S. L. (2003). Catalase (KatA) and KatA-associated protein (KapA) are essential to persistent colonization in the Helicobacter pylori SS1 mouse model. *Microbiology* 149, 665-72.

Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999). From molecular to modular cell biology. *Nature* 402, C47-52.

Hawtin, P. R., Delves, H. T. & Newell, D. G. (1991). The demonstration of nickel in the urease of Helicobacter pylori by atomic absorption spectroscopy. *FEMS Microbiol. Letters* 61, 51-54.

Hawtin, P. R., Stacey, A. R. & Newell, D. G. (1990). Investigation of the structure and localization of the urease of Helicobacter pylori using monoclonal antibodies. *J Gen Microbiol* 136, 1995-2000.

Hays, C. L. (2004). What Wal-Mart Knows About Customers' Habits. In *New York Times (14 Nov 2004)*.

High, N. J., Deadman, M. E. & Moxon, E. R. (1993). The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the Haemophilus influenzae lipopolysaccharide epitope alpha Gal(1-4)beta Gal. *Mol Microbiol* 9, 1275-82.

Holzhutter, S. & Holzhutter, H. G. (2004). Computational design of reduced metabolic networks. *Chembiochem* 5, 1401-22.

Huang, J. Q., Sridhar, S., Chen, Y. & Hunt, R. H. (1998). Meta-analysis of the relationship between Helicobacter pylori seropositivity and gastric cancer. *Gastroenterology* 114, 1169-79.

Icatlo, F. C., Jr., Kuroki, M., Kobayashi, C., Yokoyama, H., Ikemori, Y., Hashi, T. & Kodama, Y. (1998). Affinity purification of Helicobacter pylori urease. Relevance to gastric mucin adherence by urease protein. *J Biol Chem* 273, 18130-8.

Ilver, D., Arnqvist, A., Ogren, J., Frick, I. M., Kersulyte, D., Incecik, E. T., Berg, D. E., Covacci, A., Engstrand, L. & Boren, T. (1998). Helicobacter pylori adhesin

binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279, 373-7.

Imamichi, T. (2004). Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Curr Pharm Des* 10, 4039-53.

Jaenicke, R. & Creighton, T. E. (1993). Protein folding: junior chaperones. *Curr Biol* 3, 234-5.

Jensen, L. J., Saric, J. & Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat Rev Genet* 7, 119-29.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature* 407, 651-4.

Jiang, Z., Huang, A. L., Tao, X. H. & Wang, P. L. (2003). Construction and characterization of bivalent vaccine candidate expressing HspA and M(r)18,000 OMP from Helicobacter pylori. *World J Gastroenterol* 9, 1756-61.

Kansau, I., Guillain, F., Thiberge, J. M. & Labigne, A. (1996). Nickel binding and immunological properties of the C-terminal domain of the Helicobacter pylori GroES homologue (HspA). *Mol Microbiol* 22, 1013-23.

Kansau, I. & Labigne, A. (1996). Heat shock proteins of Helicobacter pylori. *Aliment Pharmacol Ther* 10 Suppl 1, 51-6.

Kitano, H. (2004). Biological robustness. *Nat Rev Genet* 5, 826-37.

Kitano, H. (2007a). A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 6, 202-10.

Kitano, H. (2007b). Towards a theory of biological robustness. *Mol Syst Biol* 3, 137.

Kluger, Y., Yu, H., Qian, J. & Gerstein, M. (2003). Relationship between gene co-expression and probe localization on microarray slides. *BMC Genomics* 4, 49.

Korcsmaros, T., Kovacs, I. A., Szalay, M. S. & Csermely, P. (2007). Molecular chaperones: the modular evolution of cellular networks. *J Biosci* 32, 441-6.

Krishnamurthy, P., Parlow, M., Zitzer, J. B., Vakil, N. B., Mobley, H. L., Levy, M., Phadnis, S. H. & Dunn, B. E. (1998). Helicobacter pylori containing only cytoplasmic urease is susceptible to acid. *Infect Immun* 66, 5060-6.

Kusters, J. G., van Vliet, A. H. & Kuipers, E. J. (2006). Pathogenesis of Helicobacter pylori infection. *Clin Microbiol Rev* 19, 449-90.

Langer, T. & Neupert, W. (1991). Heat shock proteins hsp60 and hsp70: their roles in folding, assembly and membrane translocation of proteins. *Curr Top Microbiol Immunol* 167, 3-30.

Ma, H. W., Kumar, B., Ditges, U., Gunzer, F., Buer, J. & Zeng, A. P. (2004). An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res* 32, 6643-9.

Mahdavi, J., Sonden, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K. A., Altraja, S., Wadstrom, T., Kersulyte, D., Berg, D. E., Dubois, A., Petersson, C., Magnusson, K. E., Norberg, T., Lindh, F., Lundskog, B. B., Arnqvist, A., Hammarstrom, L. & Boren, T. (2002). Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science* 297, 573-578.

Maier, R. J., Benoit, S. L. & Seshadri, S. (2007). Nickel-binding and accessory proteins facilitating Ni-enzyme maturation in Helicobacter pylori. *Biometals* 20, 655-64.

Marshall, B. J., Armstrong, J. A., McGechie, D. B. & Glancy, R. J. (1985). Attempt to fulfil Koch's postulates for pyloric Campylobacter. *Med J Aust* 142, 436-9.

Marshall, B. J. & Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1, 1311-5.

Michetti, P., Corthesy-Theulaz, I., Davin, C., Haas, R., Vaney, A. C., Heitz, M., Bille, J., Kraehenbuhl, J. P., Saraga, E. & Blum, A. L. (1994). Immunization of BALB/c mice against Helicobacter felis infection with Helicobacter pylori urease. *Gastroenterology* 107, 1002-11.

Mobley, H. L., Cortesia, M. J., Rosenthal, L. E. & Jones, B. D. (1988). Characterization of urease from Campylobacter pylori. *J Clin Microbiol* 26, 831-6.

Mobley, H. L., Garner, R. M., Chippendale, G. R., Gilbert, J. V., Kane, A. V. & Plaut, A. G. (1999). Role of Hpn and NixA of Helicobacter pylori in susceptibility and resistance to bismuth and other metal ions. *Helicobacter* 4, 162-9.

Mobley, H. L. T., Mendz, G. L. , Hazell, S. L. . (2001). *Helicobacter pylori: Physiology and Genetics*, ASM Press, Washington (DC).

Montecucco, C. & Rappuoli, R. (2001). Living dangerously: how Helicobacter pylori survives in the human stomach. *Nat Rev Mol Cell Biol* 2, 457-66.

Monteiro, M. A., Chan, K. H., Rasko, D. A., Taylor, D. E., Zheng, P. Y., Appelmelk, B. J., Wirth, H. P., Yang, M., Blaser, M. J., Hynes, S. O., Moran, A. P. & Perry, M. B. (1998). Simultaneous expression of type 1 and type 2 Lewis blood group antigens by Helicobacter pylori lipopolysaccharides. Molecular mimicry between h. pylori lipopolysaccharides and human gastric epithelial cell surface glycoforms. *J Biol Chem* 273, 11533-43.

Moran, A. P., Helander, I. M. & Kosunen, T. U. (1992). Compositional analysis of Helicobacter pylori rough-form lipopolysaccharides. *J Bacteriol* 174, 1370-7.

Moran, A. P., Knirel, Y. A., Senchenkova, S. N., Widmalm, G., Hynes, S. O. & Jansson, P. E. (2002). Phenotypic variation in molecular mimicry between Helicobacter pylori lipopolysaccharides and human gastric epithelial cell surface glycoforms. Acid-induced phase variation in Lewis(x) and Lewis(y) expression by H. Pylori lipopolysaccharides. *J Biol Chem* 277, 5785-95.

Morris, A. & Nicholson, G. (1987). Ingestion of Campylobacter pyloridis causes gastritis and raised fasting gastric pH. *Am J Gastroenterol* 82, 192-9.

Morris, H. R. & Pucci, P. (1985). A new method for rapid assignment of S-S bridges in proteins. *Biochem Biophys Res Commun* 126, 1122-8.

Muotiala, A., Helander, I. M., Pyhala, L., Kosunen, T. U. & Moran, A. P. (1992). Low biological activity of Helicobacter pylori lipopolysaccharide. *Infect Immun* 60, 1714-6.

Neidhardt, F. C. & VanBogelen, R. A. (1987). Heat shock response. In *Escherichia Coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., ed.), Vol. 2, pp. 1334-45. American Society for Microbiology, Washington, DC.

Odenbreit, S. & Haas, R. (2002). Helicobacter pylori: impact of gene transfer and the role of the cag pathogenicity island for host adaptation and virulence. *Curr Top Microbiol Immunol* 264, 1-22.

Odenbreit, S., Kavermann, H., Puls, J. & Haas, R. (2002). CagA tyrosine phosphorylation and interleukin-8 induction by Helicobacter pylori are independent from alpAB, HopZ and bab group outer membrane proteins. *Int J Med Microbiol* 292, 257-266.

Odenbreit, S., Wieland, B. & Haas, R. (1996). Cloning and genetic characterization of Helicobacter pylori catalase and construction of a catalase-deficient mutant strain. *J Bacteriol* 178, 6960-7.

Oltvai, Z. N. & Barabasi, A. L. (2002). Systems biology. Life's complexity pyramid. *Science* 298, 763-4.

Peek, R. M., Jr. & Blaser, M. J. (2002). Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* 2, 28-37.

Pesci, E. C. & Pickett, C. L. (1994). Genetic organization and enzymatic activity of a superoxide dismutase from the microaerophilic human pathogen, Helicobacter pylori. *Gene* 143, 111-6.

Phadnis, S. H., Parlow, M. H., Levy, M., Ilver, D., Caulkins, C. M., Connors, J. B. & Dunn, B. E. (1996). Surface localization of Helicobacter pylori urease and a heat shock protein homolog requires bacterial autolysis. *Infect Immun* 64, 905-12.

Popper, K. R. (2000). *The logic of scientific discovery*. Paperback edit. Classics, Routledge, London, UK.

Pruitt, K. D., Tatusova, T. & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35, D61-5.

Quackenbush, J. (2004). Data standards for 'omic' science. *Nat Biotechnol* 22, 613-4.

Radcliff, F. J., Hazell, S. L., Kolesnikow, T., Doidge, C. & Lee, A. (1997). Catalase, a novel antigen for Helicobacter pylori vaccination. *Infect Immun* 65, 4668-74.

Rahman, S. A. & Schomburg, D. (2006). Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. *Bioinformatics* 22, 1767-74.

Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A. & Legrain, P.

(2001). The protein-protein interaction map of Helicobacter pylori. *Nature* 409, 211-5.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551-5.

Romano, M., Ricci, V. & Zarrilli, R. (2006). Mechanisms of disease: Helicobacter pylori-related gastric carcinogenesis--implications for chemoprevention. *Nat Clin Pract Gastroenterol Hepatol* 3, 622-32.

Rothenbacher, D. & Brenner, H. (2003). Burden of Helicobacter pylori and H. pylori-related diseases in developed countries: recent developments and future implications. *Microbes Infect* 5, 693-703.

Salama, N. R., Shepherd, B. & Falkow, S. (2004). Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *J Bacteriol* 186, 7926-35.

Schuster, S., Fell, D. A. & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol* 18, 326-32.

Scott, D. R., Weeks, D., Hong, C., Postius, S., Melchers, K. & Sachs, G. (1998). The role of internal urease in acid resistance of Helicobacter pylori. *Gastroenterology* 114, 58-70.

Seshadri, S., Benoit, S. L. & Maier, R. J. (2007). Roles of His-rich hpn and hpn-like proteins in Helicobacter pylori nickel physiology. *J Bacteriol* 189, 4120-6.

Sixsmith, D. G., Watkins, W. M., Chulay, J. D. & Spencer, H. C. (1984). In vitro antimalarial activity of tetrahydrofolate dehydrogenase inhibitors. *Am J Trop Med Hyg* 33, 772-6.

Skouloubris, S., Labigne, A. & De Reuse, H. (1997). Identification and characterization of an aliphatic amidase in Helicobacter pylori. *Mol Microbiol* 25, 989-98.

Skouloubris, S., Labigne, A. & De Reuse, H. (2001). The AmiE aliphatic amidase and AmiF formamidase of Helicobacter pylori: natural evolution of two enzyme paralogues. *Mol Microbiol* 40, 596-609.

Skouloubris, S., Thiberge, J. M., Labigne, A. & De Reuse, H. (1998). The Helicobacter pylori UreI protein is not involved in urease activity but is essential for bacterial survival in vivo. *Infect Immun* 66, 4517-4521.

Smith, J. I., Drumm, B., Neumann, A. W., Policova, Z. & Sherman, P. M. (1990). In vitro surface properties of the newly recognized gastric pathogen Helicobacter pylori. *Infect Immun* 58, 3056-60.

Spiegelhalder, C., Gerstenecker, B., Kersten, A., Schiltz, E. & Kist, M. (1993). Purification of Helicobacter pylori superoxide dismutase and cloning and sequencing of the gene. *Infect Immun* 61, 5315-25.

Suerbaum, S. & Michetti, P. (2002). Helicobacter pylori infection. *N Engl J Med* 347, 1175-86.

Suerbaum, S., Thiberge, J. M., Kansau, I., Ferrero, R. L. & Labigne, A. (1994). Helicobacter pylori hspA-hspB heat-shock gene cluster: nucleotide sequence, expression, putative function and immunogenicity. *Mol Microbiol* 14, 959-74.

Tannaes, T. & Bukholm, G. (2005). Cholesteryl-6-O-acyl-alpha-D-glucopyranoside of Helicobacter pylori relate to relative lysophospholipid content. *FEMS Microbiol Lett* 244, 117-20.

Tannaes, T., Bukholm, I. K. & Bukholm, G. (2005). High relative content of lysophospholipids of Helicobacter pylori mediates increased risk for ulcer disease. *FEMS Immunol Med Microbiol* 44, 17-23.

Tannaes, T., Dekker, N., Bukholm, G., Bijlsma, J. J. & Appelmelk, B. J. (2001). Phase variation in the Helicobacter pylori phospholipase A gene and its role in acid adaptation. *Infect Immun* 69, 7334-40.

Telford, J. L., Ghiara, P., Dell'Orco, M., Comanducci, M., Burroni, D., Bugnoli, M., Tecce, M. F., Censini, S., Covacci, A., Xiang, Z. & et al. (1994). Gene structure of the Helicobacter pylori cytotoxin and evidence of its key role in gastric disease. *J Exp Med* 179, 1653-58.

Thiele, I., Vo, T. D., Price, N. D. & Palsson, B. O. (2005). Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-

scale characterization of single- and double-deletion mutants. *J Bacteriol* 187, 5818-30.

Todoroki, I., Joh, T., Watanabe, K., Miyashita, M., Seno, K., Nomura, T., Ohara, H., Yokoyama, Y., Tochikubo, K. & Itoh, M. (2000). Suppressive effects of DNA vaccines encoding heat shock protein on Helicobacter pylori-induced gastritis in mice. *Biochem Biophys Res Commun* 277, 159-63.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C. (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* 388, 539-47.

Vanet, A. & Labigne, A. (1998). Evidence for specific secretion rather than autolysis in the release of some Helicobacter pylori proteins. *Infect Immun* 66, 1023-7.

Walsh, E. J. & Moran, A. P. (1997). Influence of medium composition on the growth and antigen expression of Helicobacter pylori. *J Appl Microbiol* 83, 67-75.

Wang, G., Rasko, D. A., Sherburne, R. & Taylor, D. E. (1999). Molecular genetic basis for the variable expression of Lewis Y antigen in Helicobacter pylori: analysis of the alpha (1,2) fucosyltransferase gene. *Mol Microbiol* 31, 1265-74.

Warren JR, M. B. (1983). Unidentifi ed curved bacilli on gastric epithelium in active chronic gastritis. *Lancet*

Yeh, I., Hanekamp, T., Tsoka, S., Karp, P. D. & Altman, R. B. (2004). Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res* 14, 917-24.

Zhang, H., Zhang, X., Liu, M., Zhang, J., Li, Y. & Zheng, C. C. (2006). Expression and characterization of Helicobacter pylori heat-shock protein A (HspA) protein in

transgenic tobacco (Nicotiana tabacum) plants. *Biotechnol Appl Biochem* 43, 33-8.

Zuber, U. & Schumann, W. (1994). CIRCE, a novel heat shock element involved in regulation of heat shock operon dnaK of Bacillus subtilis. *J Bacteriol* 176, 1359-63.