A systems biology approach sheds new light on the regulation of acid adaptation in *Escherichia coli* BW25113 and MG1655 strains

By

ANNA STINCONE

A thesis submitted to

The University of Birmingham

For the Degree of

Doctor of Philosophy

School of Biosciences

The University of Birmingham

March 2012

Acknowledgements

In my years in Birmingham I had the pleasure to meet many friends and co-workers. I would like to thank them for being there to support me and I am sorry if I cannot mention all of them in this page. However, they will always be in my thoughts.

The most special thanks are for my supervisor, Dr Francesco Falciani. I have spent several hours, days, months, discussing, sometimes fighting with him about my project. He has been more than a supervisor in the last for years. If today I can consider myself a scientist, it is also because of him.

I would like to thank Dr Peter Lund, which helped me in extricating myself in the regulative meanders of *E. coli*.

Thanks to all my colleagues, they supported me in entering the magic world of Systems Biology. To my friend Nil, for her great support and useful advices in science and in life and for the nights spent talking and laughing. To Phillip, great companion during the long days in the lab. Thanks for letting me understand that I should use Excel in order to not to waste my time… Thanks to Wazeer, which allowed me to consider him my "little brother", hence to annoy him… Thanks to Jaanika, for letting me re-discover the pleasure of long walks, to Rita, the most reassuring presence in the office and to Helani, which made me laugh even when I was desperate. And to Kim and Peter for their support. I would like to thank also Ayesha and Nazish, for being so helpful with me at the beginning of the project.

Special thanks are for my beloved parents, because they never stop to trust in me, even when the circumstances were not in favor.

To my dear Pierre-Philippe, because he has always been there, even when the distance was not our best ally. Thanks for helping me to acknowledge my fears, instead of rejecting them; to live with them instead of trying to overcome them.

I wish to thanks all my friends in Italy and in UK: Camilla, Natalia, Mariagiovanna, Sandra, Francesco T., Francesco C. and Salvo. My friend "Chica" Teresa, Francesco, Elsa, Giacomino, Antonio M., Ludovico, Iryna, Angela, Stefano, Keyko, Federica, Richard, Silvia, Ricardo, Giorgio, Aya, Marianne, Fabio, Manuela, Daniel and Berrak and all the others that, unfortunately, I am not able to cite now.

In the end, I want to thank the Stincone's clan: my sister Antonella, my brothers Silvio and Maurizio; Ottavio, Antonella and Anna. And all my nephews and my niece: Gianni, Gianluca, Davide, Mattia, Gianmaria, Francesco and Ilaria.

And to Ilaria I want to dedicate my thesis. Not because she is my favourite niece (...she is the only one). She taught me the greatest lesson that I have received in all my life. She made me understand that when our beliefs are strong, anything is possible. The important is to think that we are able to change our life.

Author's declaration

Anna Stincone declares that:

I am the sole author of this thesis. The work presented here is original work, completed by me during the period of registration as a postgraduate student. The exception to this statement relates to the microarrays and metabolomics experiments on the *E. coli* MG1655 strain, performed by Dr A. S. Rahman and Dr S. Manzoor, and the State Space Model, developed by my supervisor Dr F. Falciani.

Abstract

The ability of *E. coli* to survive in extreme acid conditions is an important component of its physiology and it is potentially linked to pathogenesis. So far, efforts in understanding the molecular basis of acid resistance have focused on mechanisms related to proton scavenging.

In my work I hypothesized that acid resistance may require the coordinate regulation of a broader spectrum of molecular pathways. I tested this hypothesis by using an Integrative Biology approach based on gene regulatory network inference.

In my study I have profiled the *Escherichia coli* (*E. coli*) K-12 BW25113 strain using microarray technology and I have analysed a multi-omics dataset representing the transcriptional and metabolic responses of the MG1655 *E. coli* strain. An initial high-level model in the BW25113 strain representing the interaction between two component systems regulators and effectors functions was built using the ARACNE methodology. My model supported the view that acid resistance involves a mechanism based on the transcriptional switch between the expression of genes encoding aerobic and anaerobic enzymes and controlled by the two-component system (TCS) regulator OmpR. Model validation confirmed this hypothesis and provided the means to predict the ability of a given strain to survive exposure as a function of the molecular response.

This model allowed me to predict that the MG1655 strain would be more sensitive to acid than the related BW25113 strain. Acid exposure induced an opposite response in this strain by repressing most of the anaerobic enzymes in favour of the aerobic metabolism. I have developed a dynamical model to represent several smaller sub-networks by using State Space Models (SSM). The model was interrogated to identify key mechanisms of acid resistance and revealed three potential regulators of acid adaptation in the MG1655 strain: OmpR, YehT and DcuR.

I concluded that OmpR has a key role in acid adaptation in both strains and that, unlike previously thought, the ability to reassess the balance in the expression of bioenergetics genes is more important for survival than proton detoxification.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

E. coli	*Escherichia coli*
S. enterica	*Salmonella enterica*
AR/Ars/ARs	Acid Response system/s
AR1	Acid Response system 1
AR2	Acid Response system 2
AR3	Acid Response system 3
AR4	Acid Response system 4
GAD	Glutamate dependent system/ AR2
AFI	Acid Fitness Islands
TF/TFs	Transcription Factor/s
TCS/TCSs	Two-Component System/s
OCS	One Component Systems
HK	Histidine Kinase
RR	Response Regulator
Δ"*gene name*"	Strain with relative gene KO
PI	Propidium Iodide
BOX	bis-(1,3-diethylthiobarbituric acid) trimethine oxonol
gfp	green fluorescent protein
PBS	Phosphate buffered saline buffer
EDTA	Ethylenediaminetetraacetic acid
PCA	Principal Component Analysis
PC1	Principal Component 1
PC2	Principal Component 2
SAM	Significance Analysis for Microarrays
PCR	Polymerase Chain Reaction
SSM	State Space Model
ARACNE	Algorithm for the Reconstruction of Accurate Cellular NEtworks
CLR	Context Likelihood of Relatedness algorithm
BN	Bayesian Networks
FNR	Fumarate Nitrate Reductase regulator

Chapter 1: Introduction

In the last fifteen years, the advent of genomics has stimulated biologists to think of biological systems as a whole rather than focussing on detailed mechanisms of specific proteins. With the development of functional genomics technologies that allow tens of thousands of variables to be measured in single experiments, an unprecedented amount of large scale biological datasets have accumulated in publicly accessible databases (Hunter et al., 2012). This has stimulated the development of computational approaches to organize and model this information (Noble, 2002).

Systems Biology developed as an integral part of this scientific and technological revolution. Although it is challenging to rigorously define a rapidly developing discipline, we could argue that Systems Biology aims to model the underlying mechanisms that regulate a complex biological system, by combining mathematical and computational modelling with biological measurements (Kitano, 2002).

One of the biggest challenges of this new-born discipline is the integration between large scale Omics dataset, biochemical and physiology measurements in a comprehensive computational model representing the underlying biological network controlling a biological system. Recently, this complex challenge has been approached with a number of computational techniques collectively defined as network inference.

The overarching aim of my PhD project was to use this approach to model acid adaptation in two *Escherichia coli* K-12 strains. Several acidic environments occur in the biosphere, including sulfidic mine areas and marine volcanic vents. The microorganisms that inhabit them are termed ‘acidophiles’ and they can resist and survive to very low pH

values (Baker-Austin and Dopson, 2007). Acid resistance in bacteria is thought to be an important pathogenic determinant and crucial for survival of pathogenic and non pathogenic strains when passing the gastric barrier (Foster, 2004). In order to colonise the mammalian intestine, *E. coli* needs to overcome the acidic barrier of the stomach. Recently, acid adaptation has been an area of active research, leading to the discovery of several molecular mechanisms controlling this important process (Foster, 2004). All mechanisms discovered so far focused on proton scavenging strategies involved in resetting the intracellular pH to neutrality (Richard and Foster, 2004). I have shown that, in addition, several changes happen in the cells, most of them involving aerobic/anaerobic switches, in dependence of the strain. The development of a computational model revealed that OmpR, the regulator of the osmoprotectant Two-component system (TCS) with EnvZ, is a potential regulator of the mechanisms of acid resistance. I hypothesised that TCS could be the key of the mechanisms underlying several stress responses and the further cascade of regulatory events. A gene inference network identified many TCS regulators as putative responsible for signal detection. The main issue at this stage of the work was the prediction of potential TCS regulators acting during acid response. A dynamic model was developed, based on the integration of transcriptomic and metabolomics data collected during acid exposure in MG1655 strain. After analysing the neighbours of the TCS regulators in the obtained network, the results confirmed that OmpR, YehT and DcuR could be potentially involved in acid response of the MG1655 strain.

1.1 Systems Biology approaches for the understanding of *E. coli* acid response

1.1.1 Systems Biology

In Systems Theory, a system is generally defined as an entity which maintains its existence through the mutual integration of its parts (Kohl et al., 2010). Applying the same concept to biology, a biological system involves the co-regulation of cellular components, ensuring the life and survival of the system itself. The main goal of Systems Biology is to offer a consistent understanding of biological systems on the basis of theories focusing on systems-level behaviours (Kitano, 2002). The discipline does not require any restriction about the definition of a System: the word can be referred to a single cell, to a tissue, even to a more complex organism (Kohl et al., 2010).

Systems biology is based on the principle that the information does not flow only in one direction. In characterizing the molecular basis of life, the molecular biosciences have become one of the most successful branches of science, culminating in determination of the human genome sequence in 2001. However, the function of living organisms cannot be addressed satisfactorily by looking at molecules alone, not even if all molecules are studied. To address the function and dysfunction of organisms, a systems approach is needed (Kohl et al., 2010).The central dogma of Molecular Biology implies that most of the information goes from genes to RNA and to proteins, in a non reversible process (Crick, 1970). This paradigm implies the bottom-up chain of events: the flow of the information from DNA to the complete System through the different biological levels (Noble, 2008). Denis Noble claimed that the central dogma of Molecular Biology is not complete for some reasons. DNA contains the genetic information; hence it is possible to know which protein is going to be produced during the regulatory event. However, it is not

possible to quantify the amount of made protein (Noble, 2008). . Hence other controls need to be included in a System Biology study, which are explained in the downward causation (**Figure 1.1**).

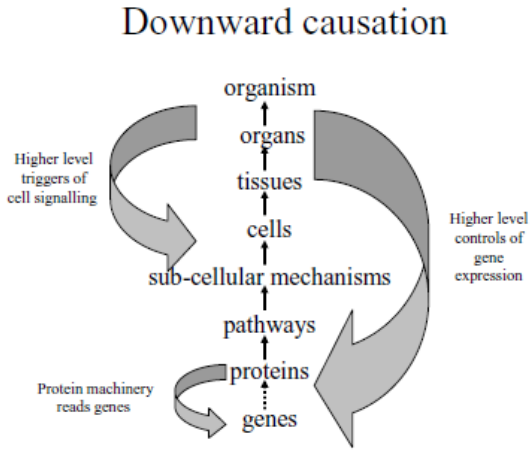


Figure 1.1 Downward causation chain of events

The causal chain of events, typical of the bottom-up methods, was added with the components of the downward causation (top-down). This allows the understanding and control of regulative events not contemplated by the reductionist scheme (Noble, 2008).

The downward causation characterises the top-down structure of a system, starting from a high level (the cell), then going down to lower levels (protein, genes) in order to get to an inverse solution (Kohl et al., 2010). In order to perform Systems biology studies, scientists need to gather information at other levels of the chain previously described, according to the principle that genes are not the only keepers of the information. The development of new technologies in molecular biology allowed the collection of multiple data, from gene expression levels to protein and metabolites quantifications. After the first genome sequencing in 1995 (Fleischmann et al., 1995), genomic analysis and annotation were required. The development of data repositories to improve the data sharing was the first

step towards the computational approaches characterising Systems Biology. Bioinformatics approaches led to create collections of multi-datasets gene-gene, gene-proteins and biochemical interactions. The KEGG pathways (Kanehisa et al., 2008), EcoCyc (Keseler et al., 2009) and RegulonDB (Gama-Castro et al., 2011) are only few examples of data repositories for *E. coli*. The databases not only provide information about the genomic data, moreover they offer datasets analysis and visualization trough computational tools for data interpretation and description of correlations between mRNA and protein levels (Waters et al., 2006).

High-throughput experiments were increasingly employed, in order to collect genome-level information from genes, mRNA, proteins, metabolites and potential interactions within these components. The measurements and analysis of these biological components lead to the development of Functional genomics.

1.1.2 Functional Genomics

The experiments belonging to this field provide genome-scale measurements for many of the molecular species existing in the cell (Joyce and Palsson, 2006). In general, all these experimental approaches need to be high-throughput, data-driven, holistic and top-down methodologies. The second feature characterising these experiments is the attempt to understand cellular functions as one integrated system more than single components. High-throughput experiments can generate large amounts of data, therefore statistical and computational analysis is required (Zhang et al., 2010).

The description of the cellular network that the omics data provide for a given time or condition can be classified into three categories, according to Joyce and Palsson: components, interactions and functional states (Joyce and Palsson, 2006) (**Figure 1.2**).

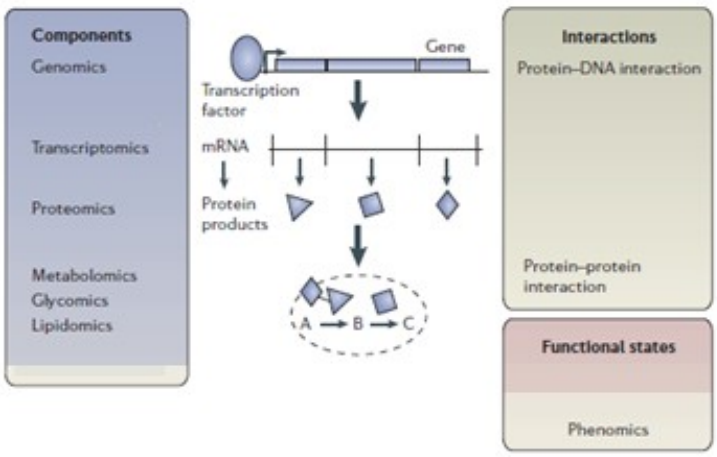


Figure 1.2 The omics data (adapted)

Omics data can be categorised in three classes, comprehending components, Interactions and functional states, describing the main entities and interactions in the cells (Joyce and Palsson, 2006).

Components data contain the information relative to a specific molecular content inside the cells. Interactions data identify the connectivity within the molecular species. Functional states reveal the cellular phenotypes. The components data are largely collected by the scientific community, they give information about specific molecules inside the cells and they are easy to integrate. Integration of high-throughput data is an important feature of current Systems Biology, since one class of omics data is not enough to explain the complexity of a system regulation (Zhang et al., 2010). Here I am going to describe some of the data belonging to the class of the components.

Genomics define the studies about genomic sequences. Since 1995 the number of sequenced genomes rapidly increased, allowing comparative genomics studies in order to find new gene regulatory elements (Kellis and Rinn, 2010). Moreover, they facilitated the identification of transcription factor binding sites in genomic sequences and protein-coding

sequences encoded by the Open Reading Frames (ORFs) (Joyce and Palsson, 2006). The use of high-throughput sequencing can increase studies of comparative genomics and decrease costs and times for the experimental procedures (Hall, 2007). One of the first outstanding genome sequencing was performed in 1997, when the sequence of *E. coli* K-12 was presented by the Human Genome Consortium (Blattner et al., 1997).

Transcriptomics is also defined as genome wide expression profiling and aim to measure the mRNA molecules in a population of cells (Zhang et al., 2010). Unlike Genomics, Transcriptomics focuses on the quantitative measurements of the dynamic expression of mRNA abundance and the associated variation between different conditions (Ye et al., 2012). Since the late 1990s a large number of genome-wide studies examined the dynamics of gene expression in many biological systems and in consideration of different conditions (Joyce and Palsson, 2006). The high-throughput transcriptomics strategies involve firstly the identification of the significant changes in gene expression under diverse conditions, or in comparison between different organisms. The most used techniques for Transcriptomics analysis are the oligonucleotides microarrays for cDNA and the more precise chip-based Reverse Transcript (RT)-PCR, which measures gene expression for several genes simultaneously and at higher sensitivity than microarrays (Stedtfeld et al., 2008).

The most recent approaches involve the next-generation sequencing on cDNA converted from the whole transcriptome (Gilbert et al., 2008) or directly on the RNA molecules, without needing in additional steps of amplification (Ozsolak et al., 2009). Combined with the chromatin immune-precipitation (ChiP) procedure, transcriptomics can also reveal genome-wide location and function of DNA binding proteins (Uyar et al., 2009).

Recently, this technology has become common in many research laboratories; therefore they can be employed in molecular diagnosis studies as well as microbiological systems stress responses (Zhang et al., 2010). As regards *E. coli*, last decade has been characterised by studies focusing on the mechanisms underlying several stress conditions. In 2005 Maurer et al. described the transcriptional profiling of the W3110 strain under acid and basic conditions, compared to pH 7. It was observed that 763 genes were differentially expressed between the three conditions. In basic conditions, the response was characterised by a reduction in the gene expression of the flagellar assembly and increase of the ATPase components, while low pHs were accelerating the processes of acid consumption and proton export (Maurer et al., 2005). Rapid acid shift experiments revealed that, beside the amino acid decarboxylase known to be activated during acidic conditions, also genes encoding the succinate dehydrogenase and biofilm associated genes were acid-induced (Kannan et al., 2008). One of the most recent works was comparing the mechanism of response in two *E. coli* strains, MG1655 K-12 and Sakai O157:H7. The study identified a strains' specific acid response and an increased resistance of the pathogenic strain, with consequent persistence in environments for food production. Many of the changes induced by acid exposure were involving several metabolic genes in both strains during stationary phase acid exposure (King et al., 2010). Transcriptomics are therefore largely and intensively used across the scientific community. However, some scientists argued about the limitations of these experimental approaches, since protein abundance in the cells cannot be predicted by the mRNA levels (Maier et al., 2009). Nevertheless, many regulatory steps are involved in the process which goes from mRNA to proteins, such pre and post-translational modifications (Watson et al., 2004). Therefore the data obtained through these methodologies cannot always be considered on its own, but should be

integrated with data from other biological sources, such as Proteomics and Metabolomics, for the understanding of the overall processes.

Proteomics aims to identify and quantify the cellular levels of proteins and modifications made to proteins. Proteins are important components of the cellular structure; they can be important metabolic enzymes and be involved in the signalling pathways (Graham and Olson, 2007). The identification of protein abundance in the cells is based on high-throughput methodologies with simultaneous measurements of the all proteins expressed in the cells. Based on the current knowledge, the Proteomics experiments involve strategies based on the separation and visualization of proteins: two-dimensional PAGE, mass spectrometry, and multidimensional separations with micro-capillary liquid chromatography (Zhang et al., 2010). This methodology has been extensively used to explore microbial metabolism, differentiation and relationship with environments (Lacerda and Reardon, 2009). Proteomic observations were also performed in *E. coli* during acid exposure. One of the studies, for instance, demonstrated the induction of periplasmic proteins during acid exposure, which could work as potential transporters for metabolic molecules. These findings were integrated with a gene expression analysis, revealing changes in the amino acid catabolism and energy metabolism (Stancik et al., 2002).

Metabolomics represent the branch of the omics field which goal is the characterization of the dynamic cellular metabolites response to environmental stimuli or genetic perturbations (Raamsdonk et al., 2001). Since the metabolome represents the result from the cellular integration of other structural components, it provides a functional readout of the cellular states (Joyce and Palsson, 2006). Metabolomics analysis are performed through gas chromatography time-of-flight mass spectrometry (GC-TOF), high-performance liquid chromatography mass spectrometry (LC-MS) or capillary electrophoresis mass

spectrometry (CE-MS) instruments, nuclear magnetic resonance (NMR) spectroscopy, and more recently vibrational spectroscopy (of which the resolution and sensitivity are considered as being lower than mass spectrometry). Metabolomics analysis can also be performed through a combined application of several technologies together in order to achieve high coverage and better identification (Kell, 2004, Dunn, 2008). Compared with transcriptomics and proteomics, technologies enabling to profile metabolites are less mature and their measurement accuracy also needs further improvement (Cascante and Marin, 2008). However, microbial metabolomics studies could be a powerful tool in the comprehension of microbial metabolism. Metabolomics amplify changes in the proteome and provide a better representation of the phenotype of an organism than any other method (Cascante and Marin, 2008). Because of the large dynamic range of metabolite diversity that requires detection, modern techniques must capture hundreds of distinct chemical species. Microbial Metabolomics studies using isotope-labelled intermediate metabolites with dynamic metabolic flux modelling, have been considered for investigations of large-scale metabolic systems, and the term '**Fluxomics**' has been coined to describe this new discipline (Forster et al., 2002, Toya et al., 2007). Improvements are still needed for the methodologies and for the data analysis software to improve the accuracy of the measurements (Dunn, 2008). Metabolomics techniques are becoming popular tools for studying the cellular state of many systems, including microbes (Wang et al., 2006), as well as in pharmacology and toxicology (Robertson, 2005) and in human nutritional studies (Gibney et al., 2005). Several metabolomics studies have been reported in recent years for various microbes including *E. coli* and *S. cerevisiae* (Tweeddale et al., 1998, Raamsdonk et al., 2001, Castrillo et al., 2003, Garcia et al., 2008, Soga et al., 2006). One of the most recent studies on *E. coli* response to several stress conditions (cold and

oxidative stress, heat shock and carbon starvation) highlighted a conserved set of response within the different environmental conditions. Co-occurring responses of the transcriptomic and metabolomics levels were observed as peaks of the maximal changes following the considered perturbations. This co-occurrence was involving genes and metabolites functionally associated and it was happening during the first minutes of exposure to the stress condition (Jozefczuk et al.). This relationship between transcripts and metabolites is no always true. In fact, for most organisms there is not direct relationship between cellular metabolites and genes (Zhang et al., 2010). Therefore data integration becomes a valid alternative to fill the gaps of the cellular understanding, when considering only one source of data as the main information.

Integrated omics approaches are becoming common tools for the understanding of global regulatory mechanisms as well as complex metabolic networks. They require many efforts in order to improve not only the experimental protocols but also the computational methodologies (Steinfath et al., 2007). In the last years many researchers have made great efforts in order to study the properties of omics data and to develop methods to integrate them.

The integration between Transcriptomics and Proteomics data try to solve the problem related to the mRNA and protein abundance. In fact, as previously said, proteins can be subjected to pre and post-translational modifications. The integration of these data allows a better coverage of the understanding of the metabolic changes in the cells, but it is also useful for cross-validation purposes (Nunez et al., 2006). It can also reveal cellular mechanisms, which cannot be detected by the use of a single omics dataset. Some of those studies involved experiments for the understanding of physiological and metabolic changes of *E. coli* during high cell density cultivation (Yoon et al., 2003). In this work it has been

observed that the patterns at gene expression levels and transcriptomic levels were showing a similar trend.

In addition to the biological factors to be considered in the process of integration, there are also the limitations of available statistical tools. Other limitations involve the processes of normalization and transformation of the data, plus experimental measurement errors (Nie et al., 2006). One of the methods for data integration, the zero-inflated Poisson regression model, for example considers undetected proteins and corrects the measured proteins abundance by considering the mRNA levels (Nie et al., 2008). An alternative is the use of multivariate statistical methods to study gene expression and protein abundance data, to visualise and explore relationships between gene and proteomic expression data (Fagan et al., 2007).

The multivariate analysis methods can be also applied in the integration of Transcriptomics and Metabolomics data. In one of the first attempts to study the correlation between these two methods, the principal component analysis was used to compare the two sets of data. Then, in order to determine the relationship between genes and metabolites the Spearman correlation was used (Urbanczyk-Wochniak et al., 2003). The integration of multi-omics datasets requires pre-processing of each dataset, with normalization procedures and missing value inputs. The advent of the omics data made possible not only the understanding of the cells behaviour. Recently, a study performed on integrated transcriptomics and metabolomics data found strong relationships within the transcripts profiles and the metabolites of metabolic functions in response to four different environmental conditions in E. coli (Jozefczuk et al.). Through several mathematical approaches, it is also possible to make predictions on putative regulatory mechanisms, for instance with gene regulatory networks inference (Gupta et al., 2011).

1.1.3 Gene regulatory networks inference

Gene regulatory networks are important structures to be considered in order to perform a Systems Biology study. They can be loosely defined as collections of molecular species and their relative interactions (Karlebach and Shamir, 2008). A gene regulatory network aims to capture dependences between the activity of gene functional products and transcription factors, post-translational modifications of proteins. The structure of a gene regulatory network is composed by nodes, which represents genes or proteins, and edges, representing molecular interactions (Hecker et al., 2009). Computational methodologies enabled the discovery and analysis of these networks, from observational data. These processes not only can help in the understanding of the regulatory mechanisms; they allow making predictions about new regulatory pathways. One of the goals of Systems Biology is the validation of these predictions for the better understanding of the system organization.

Systems Biology aims to understand the functions connected and modulated by preferred target genes and proteins (Kitano, 2002), by integrating experimental and theoretical techniques.

Kitano has defined the system level understanding as based on 4 properties:

1 System structures: the network of gene interactions and biochemical pathways.

2 System dynamics: the analysis on the system behaviour.

3 Control method: a mechanism which controls the cells.

4 Design method: the strategies to construct biological systems.

Identification of gene-regulatory networks was always considered a major challenge. The methods for creating a network model include performing a series of experiments to

identify specific interactions and conducting extensive literature investigations (Kitano, 2002). The experiments were based on the hypothesis-driven approach (**Figure 1.3**).

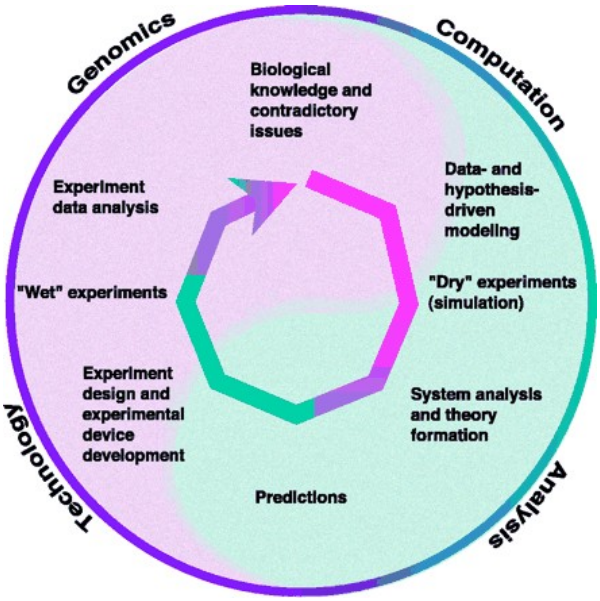


Figure 1.3 Hypothesis-driven research in systems biology (Kitano, 2002).

Performing a hypothesis-driven analysis needs to be specific about the experiments to carry out, which need to be carefully planned in advance. The research cycle behind the hypothesis driven research is realised with the creation of a model representing the phenomenon of interest. Those models can be created either automatically or manually and represent postulates that need to be experimentally validated (Kitano, 2002). Building a gene regulatory network implies an algorithm which fits a mathematical model to the experimental data, selected on the basis of the model architecture and the quantity and quality of the data (Hecker et al., 2009). In the last decade several methods for inferring gene networks have been proposed. Gene network inference (also called reverse engineering) aims to identify regulatory interactions from high-throughput data with the aid of computational methods (Bansal and di Bernardo, 2007). Recently a group of

Systems Biologists organised an annual meeting to define objective criteria to assess the validity of these algorithms, the DREAM (Dialogue on Reverse Engineering Assessment and Methods) project (Marbach et al., 2009). One of the key aims of DREAM is the development of community-wide challenges for objective assessment of reverse engineering methods for biological networks. The challenge was structured in order to evaluate the inference methods to predict the presence of regulatory interactions between genes. The used inference methods in the context were including correlation-based methods, information-theoretic methods, Bayesian network predictions and methods based on dynamical models (Marbach et al., 2009).

Depending on the data and on the experimental conditions, several approaches can be considered. For example, a system can be perturbed, genetically or by external conditions, at a given time. Or more, a system can be analysed in response to external stimuli during a time course experiment. In the first case, a model is required allowing the gene inference of steady state data, therefore a static model. In the second case, a dynamic model will be the best approach to consider. For steady state data ARACNE and other mutual information based methods can be used.

ARACNE is an algorithm designed for the identification of direct transcriptional interactions (Margolin et al., 2006a). The algorithm defines each edge (representing a direct regulatory gene interaction) as a statistical dependency between microarrays gene expression profiles (Margolin et al., 2006a). The candidate gene interactions are identified by measuring gene pair wise mutual information. The algorithm does not depend entirely on the mutual information values, but mostly on the accuracy of the estimation of mutual information ranks, calculated with a Gaussian Kernel estimator, which calculates the probability density function of the dataset. The application of the Data Processing

Inequality principle is also able to remove the indirect connections found in the network. The Data Processing Inequality calculation is used to remove the weakest interaction from a triplet of gene interactions. That is, if a MI value is available between each of three possible pairings of three markers, the weakest interaction of the three will be removed from the output. (Margolin et al., 2006a). Therefore the goal of ARACNE is not based on discovering all the possible gene interactions but the ones that have a high likelihood of being the result of direct interactions.

The context likelihood of relatedness (**CLR**) is another algorithm which uses transcriptional profiles of an organism across a diverse set of conditions to systematically determine transcriptional regulatory interactions (Faith et al., 2007). The algorithm is an extension of the relevance network method, for identifying transcriptional networks interactions and was used for the first time to identify a genome wide regulatory network in E coli K12. It was validated by comparison with the RegulonDB database and was proven to be slightly more effective than ARACNE, at least in discovering the context of Transcription factor target interactions (Faith et al., 2007). However, a recent review reported some limitations of this methodology (De Smet and Marchal, 2010).

Bayesian networks, particularly in their popular implementation Banjio (Yu et al., 2004), are algorithms requiring the estimation of probability density distribution and therefore require a larger amount of data. The applications are based on the fact that gene expression values can be described by random variables. They can be applied to both time course and steady state measurements. In order to reverse-engineer a gene regulatory network, it is important to find the Bayesian network that better fits the dataset (Ortega et al., 2008). Hence, the critical step is the model selection; for this purpose the network construction is

always performed by applying a learning algorithm, which fits the output of the mathematical model to the data (Hecker et al., 2009)

Methods based on the Ordinary Differential Equations (ODE) are also used to model time-course data (Bansal and di Bernardo, 2007). When using this methodology, the rate of a change in the concentration of a gene product at time t is described as the function of all the gene products in the network. Therefore they are used to model the dynamic behaviour of a gene regulatory network in a quantitative manner. In order to fit the experimental data, some parameters they need to be defined, for instant the function can be linear or non-linear. In case of a linear model, a linear algebraic equation can be applied to describe gene expression kinetics. In case of complex dynamic behaviours non-linear methods are preferred (Ortega et al., 2008, Hecker et al., 2009).

New technologies allowed measuring the expression of a large number of genes simultaneously; however they failed to model all possible transcription factors contributing to genetic interactions. Bayesian Networks can include hidden factors, but considering discretized instead of continuous data. The application of state space modelling to reverse engineer transcriptional networks from highly replicated expression profiling data was used in order to this issue, in order to contain gene expression measurements as continuous variables and that can model unknown factors as hidden variables (Rangel et al., 2004, Beal et al., 2005).

Recently it has been proposed a method based on the State Space Model (**SSM**) to infer module-based gene networks, which allows modelling thousands of genes with the currently available datasets (Hirose et al., 2008). SSMs are used for modelling time series data and are a subclass of Bayesian networks. They are highly considered because of the

high computational efficiency and the reduction of the noise. SSM considered two groups of variables: hidden variables (include aspects of the evolution process) and observed variables (microarray measurements) (Wu et al., 2011). Recently, SSM have been used for statistical inference of transcriptional module-based gene networks, applied on replicates of time course gene expression profiles (Hirose et al., 2008). The length of time course data is usually a limiting factor to infer gene regulatory networks. Hirose et al. proposed to explore genetic networks of transcriptional modules; the modules are sets of genes involved in the same pathway (Hirose et al., 2008).

Gene regulatory network inference algorithms are becoming really accurate, at least when their performance is measured against simulated data. When inferring a gene regulatory network, the appropriate model architecture has to be chosen and this means to consider some constraints, such as sparseness of the network (Hecker et al., 2009). In a previous work it was demonstrated that constraints on robustness and complexity suggest that an optimal network, for a given function, might be preordained to only a few (sparse) topologies (Leclerc, 2008). Hence sparseness is an indicator of the fact that few genes can modulate the expression of other genes. Enforcing the sparseness constraint reduces the number of parameters to be considered and improves the quality of network inference (Hecker et al., 2009); this means a penalization in the model reduction. Hence, other constraints (such as scale-freeness and modular design of regulatory networks) were considered (Hecker et al., 2009). The integration of several datasets and prior-knowledge could improve the gene regulatory inference for more exact and more interpretable models. Several studies are in process, in order to obtain more integrative methods, for exploring integrated omics data (Hecker et al., 2009). The aim of gene network inference is the

understanding of biological developmental processes, diseases studies, drug discovery and understanding of the bacterial high-level organization.

1.1.4 Network inference in Microbiology

One of the first examples of gene networks modelling in *E. coli* was obtained through the Dynamic Bayesian Network, using prior biological knowledge and time course datasets (Ong et al., 2002). The use of gene expression data based on tryptophan metabolism (Khodursky et al., 2000) was important for the model structure, in order to validate the Bayesian network. 15 key genes, known to be affected by the presence or the absence of tryptophan were tested for the model. The model was able to detect that the 15 key genes were correlated to most of the genes belonging to the tryptophan operon, as predicted on the basis of the biological knowledge (Ong et al., 2002).

Modelling microorganism's responses to stress conditions is another challenge of Systems Biology. In 2007, a high-throughput analysis was performed on mutants from the Keio collection (Baba et al., 2006), belonging to glycolytic pathways, during carbon starvation conditions and other perturbations (Ishii et al., 2007). An average expression index was used to define the changes in wild type and the selected mutants, in order to describe the severity of the perturbation on the metabolic pathways. It was observed that in response to the main perturbations, *E. coli* can control the response by increasing the expression of key enzyme genes in order to overcome the stress (Ishii et al., 2007). The availability of many datasets and different experimental sources (gene expression, metabolomics, etc.) allowed the development of several methods for the integration of multiple information. One of the latest models was obtained through the combination of different methodologies (ODE framework combined with a multi-objective optimization),

multiple datasets and the knowledge literature base: "Network-Inference with Multi Objective Optimization"(NIMOO) (Gupta et al., 2011). The method was able to identify many interactions, using data of a compendium of gene KO exposed to pH 5.5, i.e. the known negative interaction between gadX and gadW. Moreover, it was able to detect a key interaction involving the two-component system PhoP/PhoQ, which is a known upstream regulator of acid adaptation (Gupta et al., 2011).

The understanding of *E. coli* acid response could provide information about the survival processes of the microorganism during food processing. Therefore, the identification of potential targets responsible for the resistance to hostile environments could enable strategies for controlling food industry procedures.

1.2 Acid response in Escherichia coli

1.2.1 Gram-negative bacteria and *Escherichia coli* K-12

Bacteria can be divided into Gram-positive and Gram-negative groups on the base of the results of a technique of coloration performed by Gram in 1884 (Bartholomew and Mittwer, 1952). The original procedure involved a gentian violet solution, a solution of iodine potassium, alcohol to decolorize cells and Bismark brown as counterstain (Bartholomew and Mittwer, 1952). The violet coloration is typical of Gram positive bacteria' cell wall; the colour is lost by Gram negative bacteria because of the presence of an outer membrane which does not allow the entrance of the dye (Mitchell and Moyle, 1954). Members of the Gram negative group are cyanobacteria, spirochaetes and green sulphur bacteria, and all the Proteobacteria phyla. The phylogenetic classification of proteobacteria is based on RNA sequences and oligonucleotides catalogues (Woese, 1987, Woese et al., 1985) and the first systematic characterization defined them as the purple

bacteria.. The Proteobacteria are divided into six classes, referred to by the Greek letters alpha through zeta, again based on rRNA sequences; one of the last studied was performed with protein sequences alignment (Bern and Goldberg, 2005). The last definition of the phylum was then replaced by a nomenclature at the level of class (Zavarzin et al., 1991). The alpha, beta, delta, epsilon sections are considered monophyletic (Ciccarelli et al., 2006) the Gammaproteobacteria due to the Acidithiobacillus genus is paraphyletic to Betaproteobacteria, according to multi-genome alignment studies (Williams et al., 2010) (**Figure 1.4**). The classes of proteobacteria were once considered as subclasses (e.g. α-subclass of the Proteobacteria), but they are now regarded as classes (Olivier et al., 2005).

Gammaproteobacteria is the most studied class belonging to Proteobacteria and includes Enterobacteriaceae, Vibrionaceae and Pseudomonadacae. *Escherichia coli*, *Salmonella*, *Yersinia pestis* and *Shigella* are examples of microorganism belonging to the Enterobacteriaceae family, which includes several bacteria living in the intestine of mammals. Enterobacteriaceae can be commensal members of the host gut flora, helping in maintaining the physiology of the system. However, some members of the family can produce endotoxins, inducing local and systemic infections (Milner, 1963). The members of the family possess flagella for motility, though some of them are not motile, are facultative anaerobes and they do not form spores (Sanderson, 1976) (**Figure 1.4**).

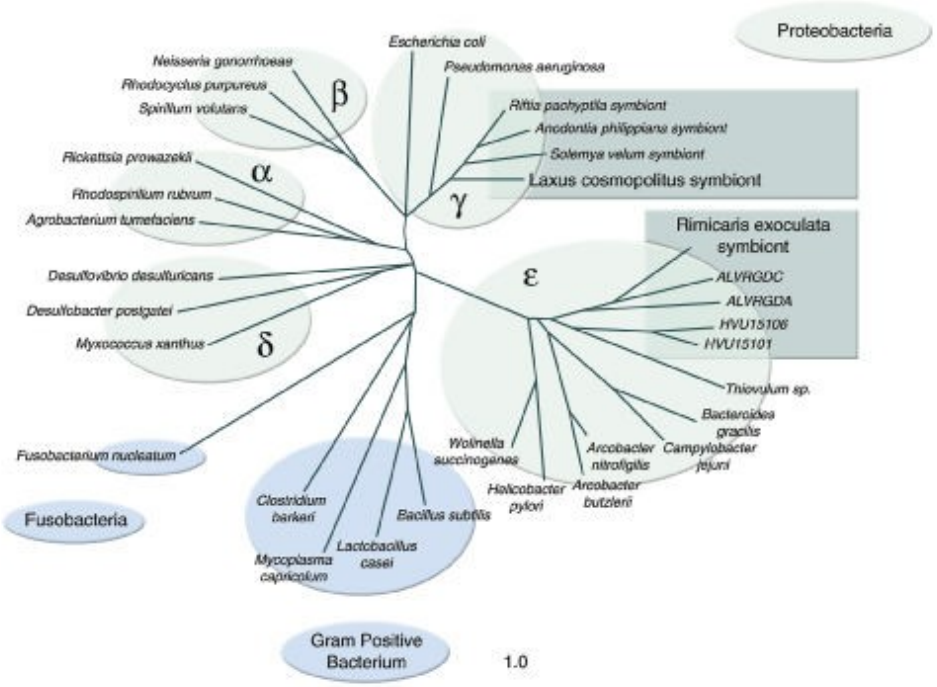


Figure 1.4 Division of the Proteobacteria classes (American Society of Microbiology website).

Escherichia coli is a rod-shaped bacterium, found in the intestine of mammals, as most of the Enterobacteriaceae, occasionally cause of infection outbreaks, because of the virulence genes carried by pathogenic strains (Sanderson, 1976) (Swerdlow et al., 1992) (Vogt and Dippold, 2005). The bacterium was discovered in 1885 by Theodor Escherich, in the faeces of healthy individual and was firstly named "*Bacterium coli*", because of the localization in the colon of the hosts; later was renamed after his discoverer (Deisingh and Thompson, 2004). The bacterium can live in several conditions; however the optimal temperature is 37˚C and pH close to neutrality. It is a facultative anaerobe microorganism and can use mixed-acid fermentation to produce lactate, acetate, succinate, hydrogen and carbon dioxide. *E. coli* has been used in the last sixty years for works on genetics,

molecular biology, biochemistry, making the microorganism a model for many studies in these fields. The most employed strain of the species is K-12, which for its versatility was accepted as the best candidate for all genome sequencing (Blattner et al., 1997).

E. coli MG1655 and BW25113 strains are probably the most studied and used strains. In terms of their strain history, it is know that MG1655 (Genotype F- λ- ilvG- rfb-50 rph-1) and BW25113 (Genotype Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), lambda-, rph-1, Δ(rhaD-rhaB)568, hsdR514) strains are both derived from the ancestral K-12 strain (Baba et al., 2006). The divergence happened at the second step of evolution (**Figure 1.5**): BD792 is the direct ancestor of BW25113 which is the result of a 13 steps of genetic manipulations process.

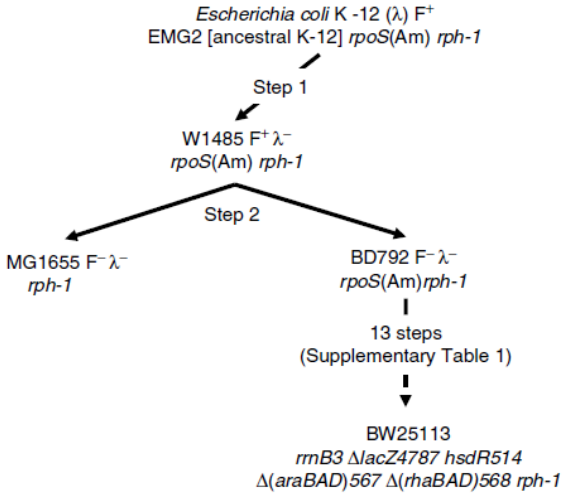


Figure 1.5 Derivation of E. coli K-12 BW25113 (Baba et al., 2006).
Strain BD792, like MG1655, is a two-step descendent of ancestral E. coli K-12. BW25113 was derived from BD792 in a series of steps involving generalized transduction and allele replacements, which included introducing the pseudo reversion rpoS (Q33) allele from MG1655 into a predecessor of BW25113.

Many studies have confirmed that the wild type BW25113 showed a similar behaviour, in physiological conditions, to another strain, also derived from the ancestral K-12 of MG1655: W3110 (Hua et al., 2003) (**Figure 1.6**).

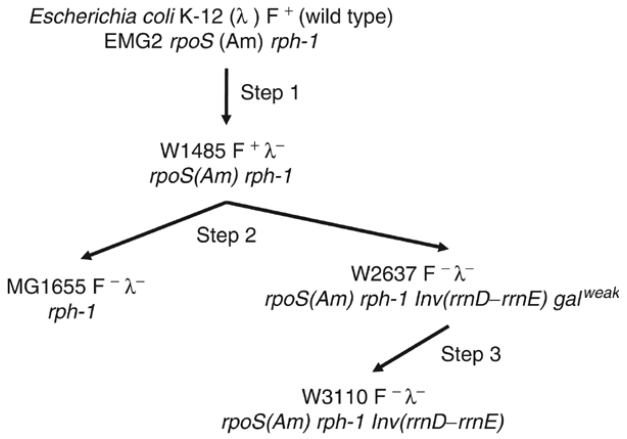


Figure 1.6: The relationships of E. coli K-12 MG1655 and W3110 (Bachmann et al., 1972; Hayashi et al., 2006).

The differences between W3110 and MG1655 strains consist of only 10 genes, on the basis of the annotation made in 2005: 4464 genes in MG1655 against 4474 genes in W3110 (Hayashi et al., 2006). The genes missing in the BW25113 compared to the MG1655 are the *araBD*, the *rhaBAD* and *lacZ* operons (Orth et al., 2011). On the basis of the pedigree, the differences between the two strains analysed in this work should be minimal.

E. coli K-12 is also the ancestor of many pathogenic strains, in which category is located the EHEC SAKAI O157:H7. The time separation between the K-12 and SAKAI is 4.5 million years, it has been found that all the lineages have acquired virulence factors in parallel (Reid et al., 2000) (**Figure 1.7**).

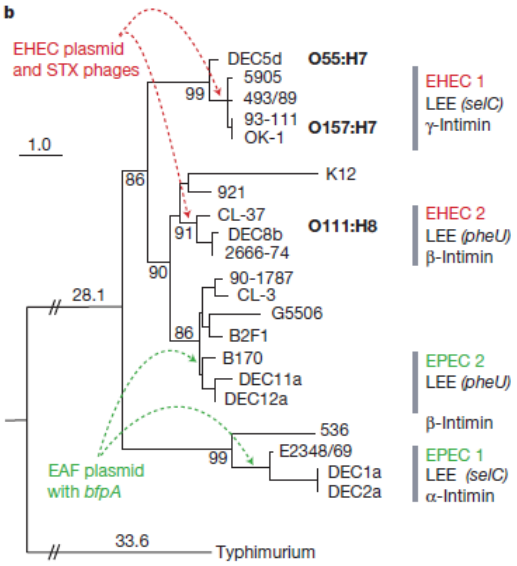


Figure 1.7 Rooted phylogeny of pathogenic strains

The tree is based on concatenated sequences of 6 loci (a total of 5.0 kb) rooted with homologous sequences from *S. enterica Typhimurium* (Reid et al., 2000).

Natural selection has produced several genetic mechanisms, facilitating acclimation to external stimuli (Moxon et al., 1994). These virulence attributes are frequently encoded on genetic elements that can be mobilized into different strains to create novel combinations of virulence factors, or on genetic elements that might once have been mobile (Kaper et al., 2004). The best combinations of virulence factors identified strong pathotypes, which can generate three different syndromes: enteric/diarrhoeal disease, urinary tract infections (UTIs) and sepsis/meningitis. Six categories were identified within the intestinal pathogens: Enteropathogenic *E. coli* (EPEC), Enterohaemorrhagic *E. coli* (EHEC), Enterotoxigenic *E. coli* (ETEC), Enteroaggregative *E. coli* (EAEC), Enteroinvasive *E. coli* (EIEC) and Diffusely Adherent *E. coli* (DAEC) (Nataro and Kaper, 1998) (**Figure 1.8**).

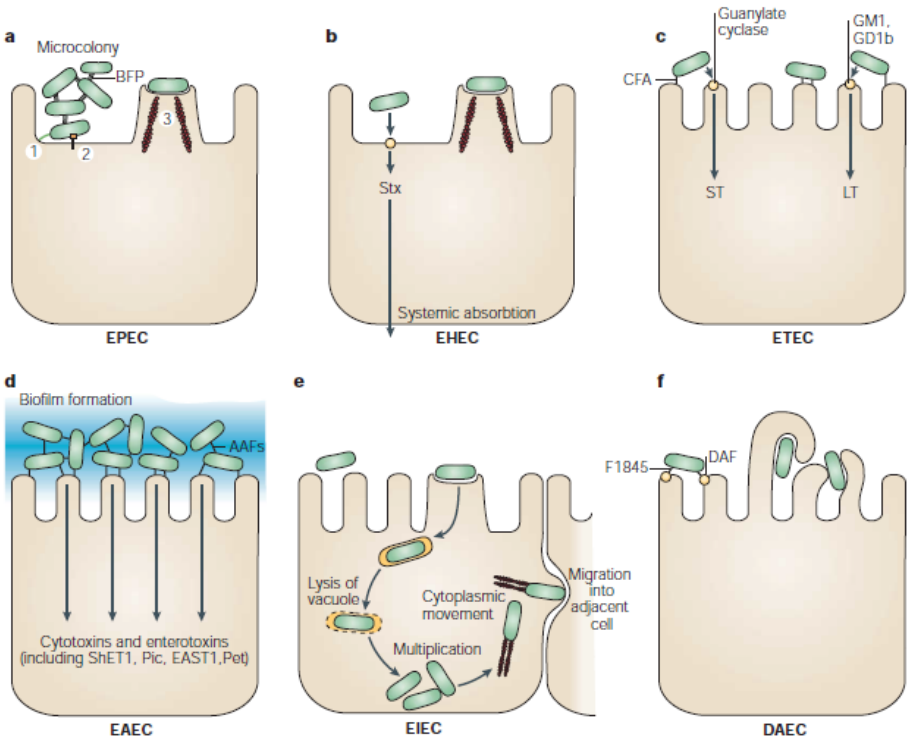


Figure 1.8 Pathogenic *E. coli* categories.

The six recognized categories of diarrhoeagenic *E. coli* each have unique features in their interaction with eukaryotic cells. Here, the interaction of each category with a typical target cell is schematically represented. **a**) EPEC adhere to small bowel enterocytes, but destroy the normal microvillar architecture, inducing the characteristic attaching and effacing lesion. Cytoskeletal derangements are accompanied by an inflammatory response and diarrhoea. **b)** EHEC induce the attaching and effacing lesion in the colon. The distinguishing feature of EHEC is the production of Shiga toxin (Stx). **c)** ETEC adhere to small bowel enterocytes and induce watery diarrhoea by the secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins. **d)** EAEC adheres to small and large bowel epithelia in a thick biofilm and produces secretory enterotoxins and cytotoxins. **e)** EIEC invades the colonic epithelial cell, lyses the phagosome and moves through the cell, spreading laterally to the other cell. **f)** DAEC elicits a characteristic signal transduction effect in small bowel (Kaper et al., 2004).

The characterization of the pathotypes was done by considering clonal groups sharing K (Polysaccharide) O (lipopolysaccharide, LPS) and H (flagellar) antigens which define serogroups (O antigen only) or serotypes (O and H antigens) (Kaper et al., 2004). The infection scheme used by pathogenic *E. coli* is characterised by the colonization of a mucosa, resistance to host defence, multiplication and host damage. Colonization occurs through adherence factors, which in the EHEC could also involve outer membrane

proteins. Moreover, some strains can secrete toxins which can affect several host processes, i.e. Shiga toxins of the EHEC can affect protein synthesis and kill endothelial and epithelial cells (Melton-Celsa et al., 1998).

Within the pathogenic strains, EHEC is the most known for the outbreaks caused by contaminated food. First recognised cases were involving consumption of poorly cooked meat (Vogt and Dippold, 2005, Riley et al., 1983), lately also other foods were described as potentially dangerous due to infection by these bacteria (Tzschoppe et al., 2012). The key virulence factor for EHEC is Shiga toxin, which is also known as verocytotoxin. The bacteria belonging to this group were firstly recognised as a human disease in 1982, when two outbreaks occurred in Oregon and Michigan (Johnson et al., 1995, Vogt and Dippold, 2005); they are able to cause bloody and non-bloody diarrhoea and the haemolytic uremic syndrome (HUS) (Nataro and Kaper, 1998). The infection process usually happens by contamination of meat with faeces or intestinal contents after slaughter (Pierard et al., 1997). Another outbreak occurred in 2011 after vegetable contamination (Werber et al., 2012). This pointed out other aspects about infective diseases which should be regarded, because of the changes in animal production and in consideration of the increased international food trade (Sofos, 2008). Food preservation is involved in the process of infection control and became even more important when it was observed that many pathogenic bacteria can easily overcome external stress conditions, such as temperature, low pH, disinfectants etc (Samelis et al., 2003). Moreover, bacteria can easily adapt themselves to non optimal conditions, therefore becoming a big issue for the food industry management (Leyer et al., 1995, Altekruse et al., 1997). Optimization of the Risk-assessment procedures and research based on the microorganisms' adaptation to food

preservative conditions (non-optimal conditions) are increasing the knowledge about the bacterial mechanisms of response to the hostile environment (Sofos, 2008).

The resistance to acidic pH shifts observed in many pathogenic (and non pathogenic) *E. coli* strains has been considered a potentially important pathogenic factor. Therefore, in the last two decades, several research groups have intensively studied mechanisms of acid adaptation (Mellmann et al., 2011). The aim of these studies involves the comprehension of the regulation processes in order to improve food production industry, but also for the understanding of the effects of a natural acidic barrier, the stomach, encountered by bacteria during the process of infection.

1.2.2 Chemistry and definition of acids

Citing the Arrhenius definition, an acid is a substance which, when dissolved in water, can increase the concentration of H_3O^+ in aqueous solution. In pure water a small part of the molecules are in constant association-dissociation equilibrium, on the basis of the following equation:

$$H_2O_{(l)} + H_2O_{(l)} \leftrightarrow H_3O^+_{(aq)} + OH^-_{(aq)}$$

The reaction of an acid is always indicated by the following formula:

$$HA \leftrightarrow H^+ + A^-$$

in which H^+ represents the acid and A^- is the conjugate base. The equilibrium of an acid and base in solution is defined by *K*, the equilibrium constant, which is the expression of the equilibrium concentrations between the two components. The K_a is the acid dissociation constant, which could be considered an equilibrium constant for a weak acid equilibrium, calculated with the following formula:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

which is more frequently used as

$$pK_a = -\log 10\ K_a$$

since the many levels of magnitude of the K_a constant. K_a is also a measure of the acid strength: strong acids usually have smaller values of pK_a.

Hydrochloric acid (HCl), hydriodic acid (HI), hydrobromic acid (HBr), perchloric acid ($HClO_4$), nitric acid (HNO_3) and sulphuric acid (H_2SO_4) are examples of strong acids, since in water they completely dissociate. Acids are usually used for many applications; they play also very important roles in the human body. The HCl present in the stomach is important to break down molecules during the process of digestion, but also constitutes an important barrier for the pathogenic microorganisms (Smith, 2003).

1.2.3 Effects of acid on *E. coli* cells

Gastric acid is an ancient mechanism of protection against infections, probably developed in cartilaginous fishes more than 400 million years ago. HCl is usually secreted in the human stomach by parietal cells, which are controlled by a complex system of endocrine cells and neurons. It is important for the denaturation of proteins, for activating pepsinogen and inactivating the ingested microorganisms (Martinsen et al., 2005).

Exposure to acid can affect several functions in *E. coli*. Among these are changes in membrane structure, protein transport and metabolism (Krulwich et al., 2011). pH homeostasis is really important for the cells during challenging acid conditions. The membranes are not permeable to protons. Proton pumps, such as the ATPase F_OF_1, allow the entry of protons (Senior, 1990). At pH 2.5 undissociated HCl could cross the

membrane as uncharged (Foster, 2004); inside the cells HCl deprotonates and acidifies the cytoplasm (Gutknecht and Walter, 1981). At pH 2.5, *E. coli* internal pH assumes values around 4.5: with this pH decrease, most of the functions will be prevented, because of the ability of the enzymes to work only at specific pH ranges. Therefore events of protein denaturation, DNA and membrane damages will begin to happen in the cells (Foster, 2004).

Therefore pH homeostasis is a very important aspect for prokaryotic and eukaryotic cells. Maintaining pH homeostasis requires several molecular components, such as the proton motive force (PMF). The PMF, defined as the electrochemical gradient of H^+ across the cell membrane, is an important factor for pH homeostasis (Kashket, 1985). PMF in bacteria is generated by proton pumps, which could be respiratory or redox-potential driven pumps or an electron transport chain (Rottenberg and Koeppe, 1989). On the basis of the pH affinity for each class of bacteria, different patterns of a proton motive force are described (**Figure 1.9**). It has been observed that *E. coli* can act as an acidophile microorganism in acid conditions (Foster, 2004).

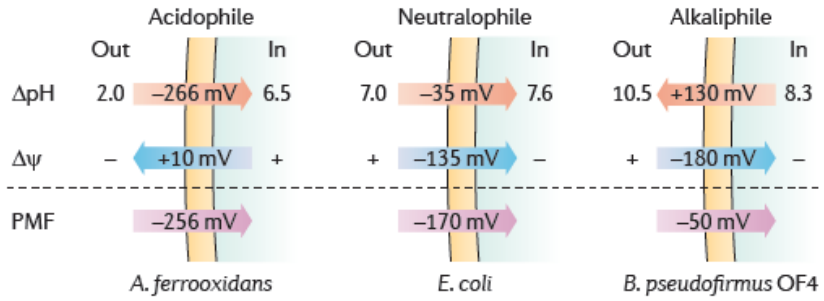


Figure 1.9 The PMF patterns of bacteria growing in different ranges of pH

In figure are shown PMF and Δψ values for the acidophile *Acidithiobacillus ferrooxidans* growing at pH 2.0, the neutralophile *Escherichia coli* growing at pH 7.0 and the alkaliphile *Bacillus pseudofirmus* OF4 growing at pH 10.5 (Krulwich et al., 2011).

PMF is calculated with the following formula:

$$PMF\ (mV) = \Delta\psi - (2.3 \times (RT / F) \times \Delta pH)$$

in which $\Delta\psi$ represents the membrane electrical potential, R the gas constant, T the temperature, F the Faraday constant and ΔpH the pH gradient. In *E. coli* the value for the PMF is between –140 mV and –180 mV. When ΔpH increases, $\Delta\Psi$ is usually reduced to keep stable the PMF. During growth, neutralophiles always have a negative-inside $\Delta\Psi$, which for *E. coli* is usually about -90 mV. Surprisingly, $\Delta\Psi$ was considerably decreasing when stationary phase *E. coli* cells were exposed to low pHs. When glutamate or arginine were added to media, $\Delta\Psi$ was then flipping to a positive value (+0 mV with glutamate, +80 mV with arginine), which is the same strategy adopted by acidophiles microorganism for pH homeostasis (Foster, 2004). Acidophiles maintain internal pH near neutrality; they have a large ΔpH and must compensate for this large chemical force by reversing $\Delta\Psi$, making it inside positive to prevent PMF from becoming too negative. Excessive PMF could create an excessive draw of external H^+, creating something similar to a short in a battery (Foster, 2004).

1.2.4 Mechanisms of acid resistance

Most of the theories underlying acid adaptation in *E. coli* are centred on the role of proton detoxification systems in re-establishing a neutral intracellular pH. Here I will review these studies and later explain how the approach proposed in this thesis has contributed to improve general knowledge in acid response.

Acid resistance is the mechanism which enables microorganism to survive at low pHs, such as gastric acidity and volatile fatty acids produced during fermentation (Giannella et

al., 1972, Giannella et al., 1973). Different species have adapted in order to overcome the host defences, for example *Vibrio cholerae* attacks the host cells in large numbers (Foster, 2004). Other species have developed cellular components which allow resisting the acidic barriers, i.e. *Salmonella enterica* and *E. coli*. Cultures of *E. coli* challenged at pH 2 for several hours were able to survive and to recover when exposed to neutral pH (Hall et al., 1995, Lin et al., 1995). In order to comprehend to which extent enteric microorganism could survive strong acidic conditions, many experiments were performed in the last years, using a multitude of different conditions (Foster, 2004). Interestingly, one of the best approaches involved the utilization of diverse media of growth, therefore mimicking external conditions for the cells. When MG1655 strain was challenged in rich media, such as Luria Bertani broth, in which concentration of nutrients is extremely high, they were showing a high resistance to acid, which was lost when the experiments were performed in minimal glucose media (Small et al., 1994, Lin et al., 1996). The utilization of different media induced scientists to identify the effects of acid in important metabolic components. For this reason, the experiments were focusing on important enzymes (for *E. coli* and *Salmonella*), and the results highlighted the great contribution to resistance of important decarboxylases (Epps and Gale, 1942). Three of these enzymes and their related amino acids were characterised and identified as components of regulatory machineries essential for acid response: Acid Response systems (**ARs**). Four ARs have been recognised in *E. coli,* three of them were discovered with the external supply of amino acids, therefore they are considered amino acid dependent; the only AR which is not amino acid dependent is also the less known to date.

AR1: This system was discovered when cells were grown in Luria Bertani broth, buffered at pH 5.5, and then challenged at pH 2.5 in minimal medium. The activation of

the system is dependent on the alternative σ^S factor (RpoS), therefore when cells are in stationary phase of growth (Foster, 2004). It has been seen that this system, because of the growth conditions, is glucose repressed, hence the involvement of another important regulator, Crp (cAMP receptor protein) (Castanie-Cornet et al., 1999). Moreover, the system seems to be activated when glutamate (or glutamine) is present in the medium, but if those amino acids are added during the challenge from external sources, AR1 is inhibited.

The AR1 mechanism which helps the cells to prevent acid stress damages is not clearly understood yet. It has been observed the potential involvement of the H^+ pump (Richard and Foster, 2004), the F_OF_1 ATPase, which generates ATP from ADP and P_i, using the protons entering the cytoplasm (Tanabe et al., 2001), as happens in *Streptococcus* (Martin-Galiano et al., 2001). The pump could also work in the opposite direction, therefore extruding protons by hydrolyzing ATP (Hicks and Krulwich, 1986). It is not clear whether it can work as H^+ extrusion system or to satisfy the energy requirements for the system to work (Foster, 2004). Even though the pump was important for the protection acquired through AR1, it has been observed that gene KO encoding the FoF1 ATPase were still showing some acid resistance (Richard and Foster, 2004).

Recently it has been discovered that the ATP hydrolysis activity of the pump is slight at pH less than 5 and the K_m (the Michaelis constant for substrate affinity) was 0.6 mM for ATP (Kobayashi and Anraku, 1972), which corresponds to 1.8 nmol ATP per milligram of protein. In these conditions, it would be really difficult for the ATPase to extrude protons because of thermodynamic reasons (Sun et al., 2011).

AR2 (GAD SYSTEM): AR2 is the best understood of the acid response mechanisms. It was discovered by accident, together with the AR3, during studies on AR1 (Lin et al., 1995) with external supply of amino acids, glutamate and arginine. AR2 is activated when cells, which were grown in reach media, where then challenged with low pH in minimal media with the exogenous glutamate (Foster, 2004, Richard and Foster, 2004). The system is characterised by the presence of two decarboxylase isozymes, GadA and GadB, which replace the α-carboxyl groups of glutamate with a proton recruited from the cytoplasm (Castanie-Cornet et al., 1999, Smith et al., 1992, De Biase et al., 1999) (**Figure 1.10**), with the production o CO_2 and GABA (γ-amino butyric acid). Moreover, a glutamate/GABA antiporter, GadC, is involved in the process of resistance by extruding the decarboxylates products to increase the glutamate concentration (Richard and Foster, 2004).

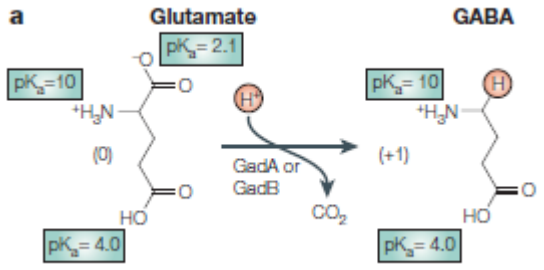


Figure 1.10 Consumption of protons during decarboxylation of glutamate.

Decarboxylation of glutamate to produce GABA. Numbers in green indicate pKa values of ionizable groups. The numbers in the parentheses are the charge values of the compound during the reactions (Foster, 2004).

The damages of acid stress could also involve the translational processes and protein folding, especially for the periplasmic proteins, which are strongly and primarily exposed to the pH shift. Two chaperones were discovered to help in the protection against acid

condition, which are periplasmic proteins and can both become dimers at pH 2.5: HdeA and HdeB (Gajiwala and Burley, 2000, Mates et al., 2007).

The regulation of the system is mainly based on the activity of the central regulator GadE (YhiE), a member of the LuxR-family, which binds a gadbox 63 bases upstream the starting sites of the *gadA* and *gadBC* genes (Ma et al., 2003). Moreover, 10 regulators are involved in the modulation of the mechanism, on the basis of the acid conditions (Foster, 2004). A big difference in the activation circuit of the GAD genes has been observed when using LB medium or glucose minimal medium in mild pH conditions, whereas GadE is always activated at pH 2.5. The activation or inhibition of *gadE* depends on a 750 bp region, located upstream of the starting codon. This region contains three different promoters' sites, which are targets of different regulators and determine activation and also auto inhibition of the regulator (Sayed et al., 2007). Three different ways of GadE activation are known to date (**Figure 1.11**).

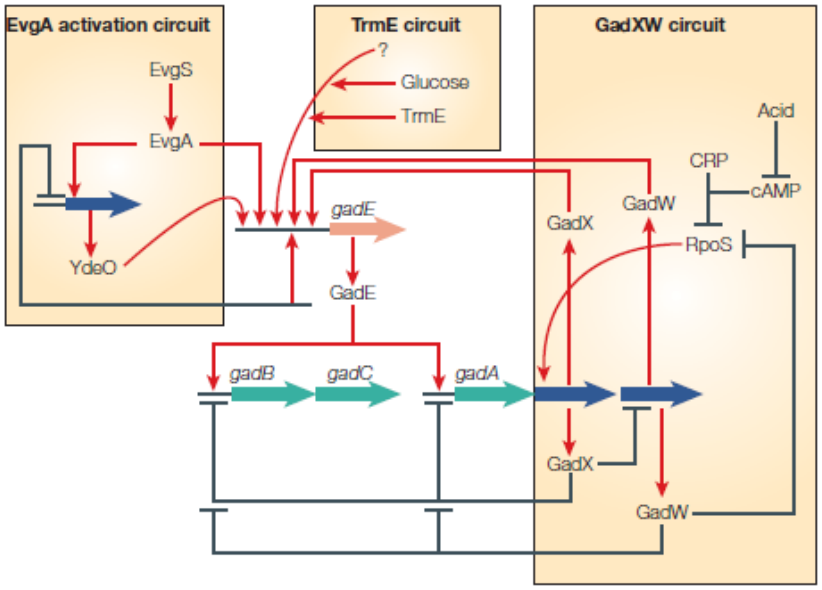


Figure 1.11 Three circuits involved in the activation of the GAD system.

Large arrows represent genes, regulatory genes are coloured blue. Smaller red arrows indicate positive control, black lines denote negative control (Foster, 2004).

The first circuit involves the Two-component system EvgA/EvgS and AraC-like regulator YdeO: when over-expressed, EvgA can indirectly activate the regulator YdeO, which in turn activates GadE, but in physiological conditions EvgA could directly activate GadE (Ma et al., 2004). The second activation circuit involves the regulators CRP, RpoS and the two AraC-like regulators GadX and GadW. GadX and GadW genes are downstream gadA, however they are transcribed from independent promoters (Tucker et al., 2003). They directly activate GadE and indirectly *gadA* and *gadBC* (Tramonti et al., 2002). In some conditions, it has been seen that both regulators can bind the decarboxylase genes promoters and could also be involved in repression events. GadW can inhibit RpoS expression, hence GadX expression. In a GadW mutant, GadX is usually overexpressed; hence the circuit is based on the inhibitory effects of each of the two-regulators towards the other (Foster, 2004) CRP is a repressor of RpoS, but its action is reduced during acidic conditions, because of its own repression, hence RpoS gene can be expressed and activate the GAD response (Ma et al., 2003). The third factor involved in the activation of AR2 is TrmE (MnmE), which prevents *gadA* and *gadBC* expression when cells grow in LB medium. TrmE is also indirectly involved in the regulation of GadE, through a possible stationary phase regulator. The three circuits were added in 2006 of new regulators and potential modulators of the response, i.e. Two-component systems, which will be further discussed. One of the regulators recently discovered is a small-RNA regulator, GadY. It has been observed that this regulator is directly involved in the regulation of the glutamate-dependent system, by positive regulation of the regulator GadX (Opdyke et al., 2004).

AR3: The third acid response system is amino acid dependent, as the GAD system. The decarboxylase present in the system AdiA, reduces the cytoplasmic proton concentration by the decarboxylation of arginine, with the production of agmatine and CO_2. The antiporter AdiC extrudes the produced amine in order to increase the amino acid pool of arginine for the system to work (Iyer et al., 2003) (**Figure 1.12**).

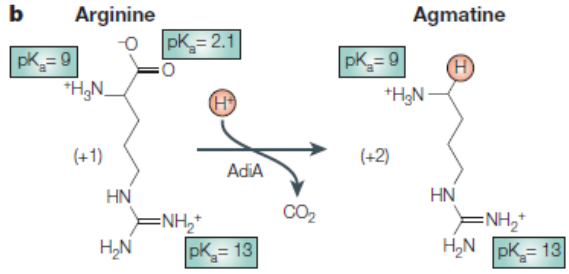


Figure 1.12 Consumption of protons during decarboxylation of arginine.

Decarboxylation of arginine to produce agmatine. Numbers in green indicate pKa values of ionizable groups. The numbers in the parentheses are the charge values of the compound during the reactions (Foster, 2004).

The use of decarboxylase for AR2 and 3 requires that internal pH values should be as close as possible to the optimum pH for the enzymes to work, hence about 4-4.5 (Capitani et al., 2003); the systems help the cells in maintaining the pH homeostasis. The protons from undissociated HCl are largely used by the decarboxylase, while it was proposed that the Chloride ions are transported outside the cells through a Chloride transporter (ClC), which also prevents the hyperpolarization of the membrane (Iyer et al., 2003). Potential membrane measurements however demonstrated that *E. coli* can actually reverse its $\Delta\psi$when cells are exposed to pH 2.5: protons enter the cells and the ClC channel could import Cl^- in order to balance the positive charge of the cells (Richard and Foster, 2004) (**Figure 1.13**).

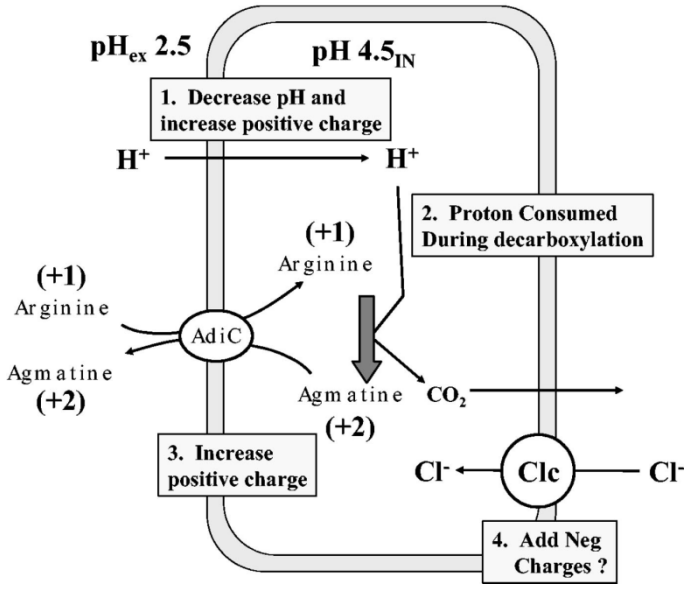


Figure 1.13 Model for the arginine-dependent acid resistance.

When the internal pH drops, the arginine decarboxylase consumes protons, producing agmatine, increasing the positive charges inside the cells. The ClC channel would be therefore required for increasing negative charges by intruding Cl^- (Richard and Foster, 2004).

The regulation of AR3 is also similar to the GAD system: AdiY, an AraC-like regulator, is directly involved in the activation of the decarboxylase and antiporter genes. RpoS is also involved in the regulation, but it does not directly activate the genes of the system. Moreover, under anaerobic conditions, it has been seen that CysB can be an activator of the system (Shi and Bennett, 1994).

AR4: The fourth and last acid response system in *E. coli* is based on the contribution by the amino acid lysine. The system usually works during carbohydrate fermentation, under anaerobic conditions and also phosphate starvation, through the decarboxylase reaction of the CadA decarboxylase, and consequent production of cadaverine (Yamamoto et al., 1997). The antiporter associated to the system is CadB, which extrudes the amine, as

previously seen for the AR2 and 3 antiporters. The regulation of the system is exerted by CadC, but in a LysP dependent manner (Neely et al., 1994). In presence of exogenous lysine, LysP, which inhibits *cadC* expression, is not active, therefore the regulator can trigger the system to work when pH values are low (Tetsch et al., 2011).

AFI: Recently a region of 12 genes at 79 centisomes was observed, called Acid Fitness Island (AFI), which also participate in acid resistance (Mates et al., 2007). Two genes belonging to this cluster (*slpA* and *gadA*) are unique to *E. coli*, though they are relatively close to *Shigella* genome. The 12 genes are usually activated under acidic conditions; when mutated, the microorganism can lose the ability to survive at pH 2.5 (Hommais et al., 2004, Mates et al., 2007). In specific conditions at acidic pHs, the cells can extrude toxic metabolites and three genes belonging to the AFI (*slp*, *yihF* and *hdeA*) were required for resistance to the toxic effects. The study also discovered that the resistance to acidic conditions was not only achieved by exogenously added glutamate: *E. coli* was perfectly able to obtain the required amino acid through *slp* and the glutamate transporter gene, also belonging to the AFI. Moreover, it has been observed that cells at stationary phase become more resistant to acid because of the involvement of two AFI genes: *yhiD* and *hdeD*, both membrane proteins (Mates et al., 2007).

1.2.5 The mechanism of acid adaptation

In pathogenic and non pathogenic *E. coli* strains the ability survive to strong acid conditions is enhanced when cells are pre-exposed to mild acid conditions. This mechanism, known as acid adaptation or acid tolerance, is also common to other enteric bacteria, such as *Salmonella* (Foster, 1999). However, during stationary phase conditions, *E. coli* cells become further more resistant than *Salmonella* (Bearson et al., 1998). This

process is used by *E. coli* in order to maintain the internal pH as close as possible to neutrality, which is not possible for not adapted cells, since their internal pH values shift to 4-4.5. The role of acid adaptation in pathogenic strains has been discovered after analysing their survival in some foods and weak organic acids, which can also induce cross-protection against other stress conditions (Leyer and Johnson, 1993). Acid tolerance in *Salmonella* involves a two-stage process: the first one at a mild pH condition (pre-shock) and the second one at pH values smaller than 4 (post-shock) (Hill et al., 1995). It has been observed that during the pre-shock 18 polypeptides are affected, 12 are induced and 6 are repressed; those changes reduce the possibility to get denatured acid sensitive proteins. Moreover, the pre-shock stage induces the expression of acid shock proteins (Foster, 1993). The studies performed on *Salmonella* identified five regulator loci including Fur and OmpR (Hill et al., 1995). In *E. coli* pathogenic and non pathogenic strains, the regulation of acid adaptation is mainly based on the contribution of RpoS and CRP (Foster, 2004). The GAD system is also involved in the process, depending on the growth conditions and the medium (Foster, 2004). Recently it has been discovered the involvement of the outer membrane protein TolC (Deininger et al., 2011), which is part of the EvgA regulon (Masuda and Church, 2002).

The mechanism of acid adaptation is extremely important for the food industry, since the conditions required for food preservation, such as apple cider, include temperature shifts and low pH environments (Sofos, 2008)

The control of acid adaption is mainly based on the regulation of the systems for acid resistance, however, only few works were focusing on finding out the mechanisms enabling *E. coli* to sense the pH shift. New mechanisms could involve different elements from the canonical response, such as transporters and two-component systems, which can

act as sensors and activators of the response to stress conditions (Soncini and Groisman, 1996).

1.3 The TCS signalling pathway

1.3.1 Structure and mechanisms of Two-component systems

Bacteria commonly use two-component signal transduction pathways to couple environmental stimuli to adaptive responses (Gao and Stock, 2009). TCS are the most common signal transduction mechanism in bacteria (Zhou et al., 2003). The mechanism is based on a stimulus-response coupling which allows sensing and responding to external changes. The typical structure for a TCS consists of a histidine protein kinase (HK) and a response regulatory protein (RR) (Stock et al., 2000) (**Figure 1.14**).

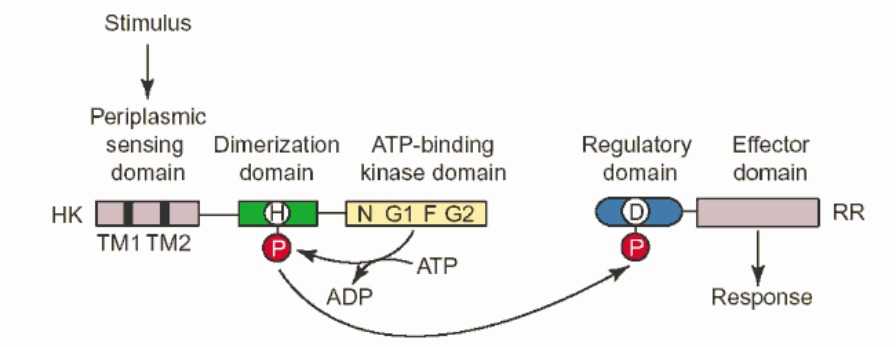


Figure 1.14 Two-component phosphotransfer mechanism.

A typical two-component phosphotransfer system consists of a dimeric transmembrane sensor HK and a cytoplasmic RR and transmembrane segments (TM1 and TM2). Conserved sequence motifs N, G1, F and G2, are located in the ATP-binding domain. The autophosphorylation happens on a specific conserved His residue (H). The phosphoryl group (P) is then transferred to a specific Asp residue (D) in the conserved regulatory domain of the RR (West and Stock, 2001).

Three phosphorylation events are involved in the functioning of TCS:

1. Autophosphorylation: HK-His + ATP ↔ HK-His~P + ADP

2. Phosphotransfer: HK-His~P + RR-Asp ↔ HK-His + RR-Asp~P

3. Dephosphorylation: RR-Asp~P + H_2O ↔ RR-Asp + P_i

The three reactions require divalent metal ions with Mg^{2C+} as the relevant cation in vivo (Stock et al., 2000).

The events happening following external stimuli are regulatory effects exerted by phosphorylation of the RR component. The activity of the kinase is modulated by the signals sent to the sensing domain; the HK is involved in a process of autophosphorylation at a conserved His residue in the kinase core. Subsequently, with a stoichiometric reaction the phosphoryl group is transferred from the HK to an Asp residue in the RR regulatory domain. In the amino acid sequence of prokaryotic and eukaryotic HKs kinase core (about 350 amino acids) there are five conserved amino acids; the His represents the most important part of the core (**Figure 1.14**) (Stock et al., 2000).The stimuli can be detected by the N-terminal sensing domain of the HK, on the basis of a specific stimulus interaction. The phosphorylation of RR, on the conserved Asp, can happen through different molecules which act as donors, such acetyl phosphate and imidazole phosphate (Lukat et al., 1992). Two domains are present in each RR: a conserved N-terminal regulatory domain and a variable C-terminal effector domain, however the majority of the RR are transcription factors with DNA-binding effector domains (Mizuno and Tanaka, 1997). These domains can be classified into three major families, in which OmpR, NarL and NtrC are the most representative (Mizuno and Tanaka, 1997). The effector domains are important because of their activation/repression ability, binding specific sequences which are recognised in

dependence of the stimulus. The great diversity of the binding sites allowed the specialization of the TCS for several stimuli.

The HK input domains are well known to date, contrarily to N-terminal sensor domains, which are different in sequence, membrane topology and domain arrangement. This great variability could probably reflect different principles in stimulus sensing and transduction of the event. The signal domains could be therefore classified on the basis of their localization and therefore on the stimuli detected and transducted.

Periplasmic sensing HK: The TCS belonging to this group represent the most studied and well known to date. They are formed by the N-terminus periplasmic sensing domain and by a C-terminus cytoplasmic transmitter domain, connected by transmembrane helices. The linking domain is usually used as an additional criterion for the classification of the members of the group. EnvZ, PhoP and TorS belong to this group and they can be considered as prototypical examples.

EnvZ with its correspondent RR, **OmpR**, belong to the most studied and known TCS in *E. coli*, which promptly respond to osmolarity changes (Stock et al., 2000). An event of trans-autophosphorylation on the conserved His-243 of EnvZ determines the activation of the TCS; the phosphoryl group is then transferred to the Asp-55 of the RR OmpR (Yang and Inouye, 1993). This event determines the regulation of the outer membrane proteins OmpF and OmpC, which then allow the passage of small hydrophilic molecules through the membrane (Nikaido and Vaara, 1985)EnvZ is 450 amino acids long, it consists of a transmembrane dimer in the inner membrane of *E. coli* and it is connected to OmpR are by a flexible linker. The RR OmpR showed, through X-ray crystallography, a helix-turn-helix structure. It was observed that in vivo and in vitro studies OmpR-P binds to the -100 to -38

region of *ompC,* one to the -380 to -361 and -100 to -39 regions of *ompF* (Bergstrom et al., 1998).

The role of OmpR is well known in literature to date, in *E. coli* and *Salmonella*. OmpR not only can exert a regulatory action during osmotic stress conditions, it is also involved in the regulation of biofilm formation and curli fimbriae, moreover can inhibit the expression of flagellar genes (Vidal et al., 1998). The **PhoP**/**PhoQ** TCS has an important function for the pathogenesis in Salmonella and other enteric bacteria, by increasing the concentrations of lipid A, including the ones for antimicrobial peptide resistance (West and Stock, 2001). PhoQ, which is the sensor kinase, is activated at low concentrations of Mg^{2+} and repressed by high concentrations of divalent cations (Gunn and Miller, 1996). The sensor contains a periplasmic PAS domain, which structure matches with CitA and DcuS sensors. The binding of the cations happens in the acidic surface of the cytoplasmic membrane; the Mg^{2+} cations form bridges between this region and the acidic phospholipid membrane (Kim and Cho, 2006). When the antimicrobial peptides break this bridge, PhoQ becomes active, therefore allowing its auto-phosphorylation. The corresponding RR, PhoP, belong to the subfamily OmpR/PhoB, and it is characterised by a winged-helix DNA-binding domain (Bachhawat and Stock, 2007). In vitro PhoP is a dimeric protein in both phosphorylated and not phosphorylated states. In vivo studies indicated that the unphosphorylated and phosphorylated PhoP dimers have structural relationships that allow similar binding of their C-terminal effector domains to the target DNA. The fact that only phosphorylated PhoP dimers can induce or repress gene transcription suggests that phosphorylation could require transcriptional components in association with other regulators (Birck et al., 2003). Recently PhoP was found to regulate virulence and membrane potential in the uropathogenic *E. coli*; moreover alterations in the expression of

genes encoding respiratory chain components also occurred in the *phoP* mutant (Alteri et al., 2011). A second group belonging to periplasmic HK group is characterised by **NarX**/**NarQ** like sensors. In *E. coli* and Salmonella the regulation of anaerobic respiration is exerted by two nitrate reductase TCS: NarX/NarL and NarQ/NarP. A sequence of 18 conserved amino acids was found flanking the two transmembrane sites of the sensors; however only one transmembrane site is required for the nitrite and nitrate detection. NarL is a nitrite/nitrate dual regulator which activates the expression of the genes involved in the nitrate respiration. The response regulator NarL belongs to the LuxR/UhpA family and contains the N-terminal receiver domain and a C-terminal DNA-binding domain. In unphosphorylated NarL, the DNA-binding site of the output domain is blocked by the receiver domain (Baikalov et al., 1996). The phosphorelay signal transduction pathway induces the separation of the two domains and the formation of a new dimerization interface in the C-terminal domain. The RR NarP is another regulates of many aerobic and fermentative genes, dependent on nitrate and nitrite concentrations. It has been observed that between the two RR, NarX and NarL, exists 44% homology in the structure (Baikalov et al., 1996) and they can cooperate in the regulation of the anaerobic processes (West and Stock, 2001).

HK linked to transmembrane regions: In this group the transmembrane regions have a predominant role, since they are directly involved in the stimuli detection. The grouping is mainly based on their functioning and on their structure. These sensors are not present in *E. coli*, however two phylogenetically unrelated groups of cell envelope stress-sensing HKs show similarities to members of this group, in consideration of their size and domain organization: PmrB/BasS-like HKs mediating resistance to cationic antimicrobial peptides in *E. coli* and *Salmonella* (West and Stock, 2001). **PmrB**/**BasS**-like HKs possess a

periplasmic linker of 30 to 35 amino acids between the transmembrane regions. It has been demonstrated that PmrB senses ferric iron through two conserved ExxE (involved in iron uptake) motifs in this short extracytoplasmic sensor domain. Therefore, PrmB-like proteins are periplasmic-sensing HKs with a HK linked to transmembrane region.

Cytoplasmic sensing HK: The sensing domain of these HK is located in the cytoplasm. The sensor region could be partly membrane integrated, associated to membrane proteins or permanently located in the cytoplasm.

The membrane anchored HK groups contains almost 100 members, in which it can be catalogued the sensor **KdpD**, which controls turgor in *E. coli* under limiting K^+ concentrations, with its associated RR **KdpE**. Sensing of osmolality occurs indirectly by measuring the intracellular parameters K^+, ATP concentration, and ionic strength, which respond to osmolality changes (Bang et al., 2002). The TCS controls the K^+ uptake by sensing also ATP concentrations (West and Stock, 2001). The transmembrane helices were not found important for the sensing, however they are required for the correct positioning of the sensing region in the cytoplasm. The **ArcA**/**ArcB** TCS has a very important role in the facultative anaerobic metabolism of *E. coli* in response to O_2. ArcA can repress genes of the aerobic metabolism and activate genes of the fermentative pathways (Iuchi and Lin, 1988). ArcB can sense the O_2 levels through an intermediate molecule of the anaerobic respiration, but in non stimulating conditions it could also work as a phosphatase of ArcA (Georgellis et al., 2001).

1.3.2 TCS Cross-talking events and stress responses

Events of cross-talking (cross-regulation and cross-phosphorylation) are defined such as mechanism in which one TCS can control another regulatory system (Wanner, 1992).

The phosphorylation events between the Histidine kinase of the HK and the RR are usually specific for each TCS, however, it was observed that some phosphorelay systems are not limited to only one regulator, but can have multiple targets (Goulian, 2010)A typical example for this “one-to-many” trend of regulation is the CheA sensor, which can easily phosphorylate two response regulators: CheY and CheB (Kirby, 2009). NarX and NarQ can also both regulate the phosphorylation of NarL and NarP. In contrast, unwanted events of cross-phosphorylation could negatively affect the specificity of the signal (Laub and Goulian, 2007), therefore it was observed that HK prefers the interaction with its cognate RR. Cross-phosphorylation could be interpreted as an integration of signals to control multiple targets, but only few examples have been reported to date (Laub and Goulian, 2007). The cross-talk is based on the lack of specificity of the interaction but could also play a role in inhibiting the phosphorylation from a non cognate HK. Additionally, the RR could also suppress cross-talking events by competing with other RR, in order to prevent non specific signals (Siryaporn and Goulian, 2008). Recently it was discovered an evidence of cross talking between the TCS EnvZ/OmpR and CpxA/CpxR, by studying the mechanism of suppression of this cross talking which can be described by the 3 proposed mechanisms (Siryaporn and Goulian, 2008). The first mechanism involves the suppression by EnvZ, the second is based on the fact that phosphorylation of OmpR by CpxA is negligible when EnvZ is present; the third mechanism involves the suppression by CpxR (Siryaporn and Goulian, 2008). Kinetic studies revealed that the events of cross-phosphorylation usually happening in vitro, are not visible in vivo because of the double phosphatase activity of the sensors, fast enough to remove the phosphates that get transferred via non-cognate interactions (Groban et al., 2009). Those mechanisms are also considered buffering mechanisms and could emerge from the ability of the histidine kinase

to both phosphorylate and dephosphorylate the response regulator (Alves and Savageau, 2003). Many TCS can regulate other TCS through auxiliary proteins which modulate the activity of either the HK or the RR; these connectors can mediate positive and negative feedback on the TCS.

TCS are in control of many stress responses in bacteria and important for sensing several environmental conditions. A phenotypic analysis on the gene KO of *E. coli* TCS (HK and RR) revealed that some TCS can sense many extracellular changes (Zhou et al., 2003). For example, CpxA/CpxR, which is known to control cell envelope stress, was found to respond also to ethylene glycol sensitivity; EnvZ/OmpR mutant increased the use of carbohydrates and showed resistance to cephalosporin. The TCS PhoP/PhoQ showed increased use of fructose and mannitol and RssB, which regulates the RpoS degradation process, was found to decrease the use of C_4 di- and mono-carboxylates and amino acid N sources (Zhou et al., 2003).A summary of the findings is shown in **Table 1.1**.

Table 1.1 Phenotypes that two-component and other mutants have in common (Zhou et al., 2003).

Mode of action	Phenotype	Genes deleted
Alanine deamination	Increased sensitivity to L-alanine at pH 9.5	*cpxRA, ntrBC, luxS*
Alkaline pH sensitivity	Increased sensitivity to pH 10	*arcA, arcB, luxS, phnC-P*
Amino acid analog	Increased resistance to β-chloro-L-alanine	*arcA, arcB, ntrBC*
Anion transport	Increased sensitivity to sodium nitrite	*arcB, rssB*
Anion transport	Increased sensitivity to sodium tungstate	*arcB, baeSR, rssB*
Antimicrobial	Increased sensitivity to myricetin	*arcA, arcB, baeSR*
C source	Decreased utilization of C_4 acids	*dcuRS, rssB*
C source	Increased utilization of D-fructose and D-mannitol	*ompR-envZ, phoPQ*
C source	Increased utilization of D-melibiose	*barA, uvrY*
C source	Increased utilization of β-methyl-D-glucuronic acid	*arcA, rpoS*
Cation transport	Increased sensitivity to cobalt chloride	*arcA, arcB, ompR-envZ, qseCB*
Cation transport	Increased sensitivity to nickel chloride	*baeSR, qseCB*
Cell wall inhibitor	Increased sensitivity to cefoxitin	*phnC-P, phoH*
Cholinergic antagonist	Increased sensitivity to pridinol	*phoH, rstAB*
DNA damage, antifolate	Increased resistance to hydroxylamine	*cbrA, uvrY*
DNA synthesis inhibitor	Increased sensitivity to nitrofurazone	*cbrBC, ntrBC, rcsB, uvrY*
Folate antagonist	Increased sensitivity to trimethoprim	*rcsB, yojN-rcsB*
Lipophilic chelator	Increased sensitivity to 1,10-phenanthroline	*arcA, arcB, cusRS, luxS*
Lipophilic chelator	Increased sensitivity to 5,7-dichloro-8-hydroxyquinaldine	*cbrA, luxS, phn*
Lipophilic chelator	Increased sensitivity to 5-chloro-7-iodo-8-hydroxyquinoline	*arcB, narP*
Lipophilic chelator	Increased sensitivity to 8-hydroxyquinoline	*arcB, psiE*
Membrane agent, outer	Increased sensitivity to polymyxin B	*atoSC, uvrY*
Membrane perturbant	Increased sensitivity to membrane-active agents	*arcA, arcB*
Membrane, cationic detergent	Increased sensitivity to methyltrioctylammonium chloride	*arcA, arcB, atoSC, uvrY*
N source	Decreased utilization of Leu-Trp	*cpxRA, rssB*
N source	Decreased utilization of δ-amino-*N*-valeric acid	*ntrBC, rssB*
Osmotic sensitivity	Increased sensitivity to ethylene glycol	*cpxRA, ompR-envZ, yojN*
Osmotic sensitivity	Increased sensitivity to NaCl	*atoSC, rcsB, yojN, yojN-rcsB*
P sources	Decreased utilization of organophosphates as P sources	*phoA, phoBR*
Protein synthesis inhibitor	Increased resistance to troleandomycin	*phn, phoH*
Protein synthesis inhibitor	Increased sensitivity to amikacin	*cpxRA, ntrBC*
Protein synthesis inhibitor	Increased sensitivity to dihydrostreptomycin	*atoSC, cpxRA, uvrY*
Protein synthesis inhibitor	Increased sensitivity to Geneticin (G418)	*arcA, cpxRA, ntrBC*
Protein synthesis inhibitor	Increased sensitivity to hygromycin B	*cpxRA, kdpFABCDE, rpoS*
Protein synthesis inhibitor	Increased sensitivity to paromomycin	*arcA, ntrBC, phoBR*
Protein synthesis inhibitor	Increased sensitivity to tobramycin	*cpxRA, ntrBC, phoBR*
Protein synthesis inhibitor	Increased sensitivity to troleandomycin	*arcB, rstAB*
Respiration inhibitor	Increased sensitivity to iodonitrotetrazolium violet	*atoSC, rcsB, uvrY*
Respiration inhibitor	Increased sensitivity to thioridazine	*arcA, arcB, ntrBC, ompR-envZ*
S source	Decreased utilization of methane sulfonic acid	*cpxRA, rssB*
Transport, toxic cation	Increased sensitivity to lithium chloride	*cpxRA, ompR-envZ*

These findings lead to the hypothesis that a potential regulatory network can connect the TCS, in order to satisfy the need of the cells to respond to the external stimuli. High-throughput technologies and Systems Biology approaches can help in developing models in order to explain the intricate system of responses and interactions within TCS (Soncini and Groisman, 1996, Zhou et al., 2003). A first evidence of transcriptional interaction was found in a gene expression analysis performed on the entire gene KO of the TCS in *E. coli* (Oshima et al., 2002). Functional analysis performed on each mutant pointed out the main changes at metabolic, energetic and in general, cellular levels in Luria Bertani broth and

under aerobic conditions. The cellular pathways affected could be determined by pattern matching. More than half the TCS mutants showed major alterations in a small number of genes. The most dramatic changes were detected in mutants of **ArcB/ArcA**, the osmoresponsive **EnvZ/OmpR** system and the response regulator **UvrY**. It has been also discovered that multiple TCSs are involved in the regulation of cellular functions such as RpoS regulon, flagellar synthesis and maltose transport; most importantly, several TCSs would control other TCSs (Oshima et al., 2002). Evolutionary processes might prefer to simplify the signalling cascade of events by involving only few components, which can detect several external conditions. Nevertheless, environmental bacteria which can adapt to different metabolic conditions were found to code for more TCSs than microorganisms living in a uniform habitat, for example bacteria that need to adapt to host organisms (Beier and Gross, 2006). For example, the role of the TCSs in virulence is still poorly understood, however few examples of involved TCSs in infection processes were found in different pathogenic bacteria.

1.3.3 Two-component systems and acid resistance

I have previously described the role of the two-component system (TCS) EvgA/EvgS in acid response, which initiates a transcriptional cascade of events modulating acid resistance genes, encoding the regulators EvgA, YdeO, and GadE (Eguchi et al., 2011)The function of Two-component systems in acid response is becoming even more intriguing since the role of EvgA/EvgS was explained. However, an additional finding pointed out the role of PhoP/PhoQ in the regulation of AR1 in both *E. coli* and *Salmonella*. The involvement of PhoP in the regulation of the AR2 genes was discovered when it has been observed that the expression of GadW was PhoP dependent (Eguchi et al., 2004, Zwir et

al., 2005). Moreover, through foot-printing experiments, both *hdeA* and *gadW* were identified as PhoP targets. This study highlighted also another aspect, which could potentially spread the interest of the Two-component systems in the study of the acid response regulation. EvgS/EvgA is able to enhance the expression of a protein, which connects the previous cited Two-component system and PhoP/PhoQ: SafA. SafA is a small inner membrane protein which can bind the sensor PhoQ, after being activated by EvgA (Eguchi et al., 2007). The role of TCS connectors is not limited to the previous finding. A circuit of connectors has been lately discovered, which involves the SafA protein and IraM, a small protein which inhibits proteolysis of RpoS through RssB (Eguchi et al., 2011) (**Figure 1.15**).

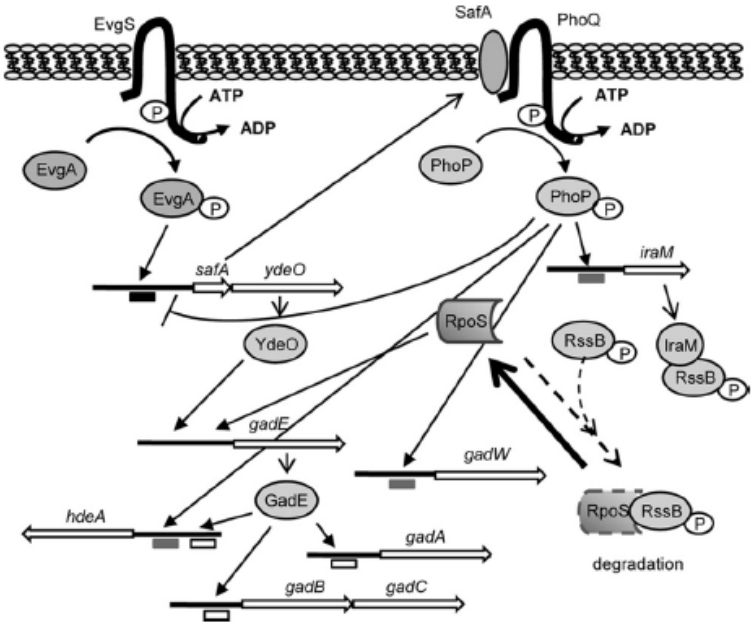


Figure 1.15 Model of regulation network of AR genes initiated by the EvgS/EvgA system.

EvgS/EvgA induces the EvgA-YdeO-GadE cascade of AR gene regulation and SafA, which then activates the PhoQ/PhoP system. PhoP/PhoQ then induces the expression of GadE and GadW and through the connector IraM regulates RssB, which controls RpoS proteolysis (Eguchi et al., 2011).

Recently, the involvement of another TCS was studied in the activation process of AR2, with the possible involvement of GadE. RcsB was seen to activate genes required for acid resistance during stationary phase (Castanie-Cornet et al.), which promoters could be either dependent or independent of GadE (Johnson et al., 2011). The regulator OmpR, which is involved in the response to osmotic stress conditions, was also observed as potential regulator of the response to mild acidic conditions in *Salmonella* (Bang et al., 2000) and in *E. coli* (Schwan, 2009, Stincone et al., 2011).

The model input-output characterising the TCS response has always been considered of great interest for the understanding of the biological systems, last decade has been characterised by several studies on the potential involvement of TCS in sensing external adverse conditions and regulating the response to them, i.e. acid stress response.

Chapter 2: Characterization of the BW25113 strain response to acid

2.1 Introduction

The acidic barrier of the stomach represents a strong challenge for many pathogenic enterobacteria (Giannella et al., 1972). The ability of some *E. coli* strains to survive exposure to strong acid conditions is potentially relevant for pathogenicity (Benjamin and Datta, 1995, Conner and Kotrola, 1995, Foster, 2004). For this reason, the molecular and physiological response to acid stress has been the subject of intense investigation (Foster, 2004). Four acid stress response systems (ARs) that can protect *E. coli* from low pH are known to date (Foster, 2004, Richard and Foster, 2003, Richard and Foster, 2004). Three of those systems are amino acid dependent (glutamate, arginine, and lysine); one of them is instead characterised by a mechanism which involves the FoF_1 ATPase and the two master regulators CRP and RpoS (Foster, 2004). *E. coli* also shows acid adaptation, characterised by enhanced resistance to low pH following exposure to mild acidic conditions (Foster and Hall, 1990, Boot et al., 2002). This mechanism is mediated by the up-regulation of acid shock proteins, including components of the acid response systems described above.

Exposure to acid induces a sudden drop of the intracellular pH (Foster, 2004), which despite the beneficial effect of the ARs, has profound effects on the physiology of the cell. For example, high intracellular proton concentration may induce uncoupling of oxidative phosphorylation resulting in alteration of energy metabolism (Richard and Foster, 2004). More generally, genes involved in energy metabolism, transport and amino acid biosynthesis are known to be modulated at both mRNA (Jozefczuk et al.) and protein level (Stancik et al., 2002) suggesting that a much broader spectrum of adaptation pathways may

be modulated in response to acid exposure. Therefore, the understanding of acid stress response goes beyond the study of the ARs.

A number of genome wide expression profiling studies, performed in W3110 strain, representing different acid stress conditions have been recently published (Maurer et al., 2005, Hayes et al., 2006, Kannan et al., 2008). Although they have contributed to the identification of novel genes transcriptionally regulated during acid exposure, they do not yet provide with a comprehensive model of *E. coli* acid resistance. Therefore, my work is focused on the general understanding of the mechanisms involved in acid adaptation (required also for acid resistance), in order to develop and validate a model for the identification of new potential regulators.

In this chapter the effects of acid adaptation on the *E. coli* BW25113 strain are shown, with particular attention to the modulated pathways and their corresponding components.

2.2 Materials and Methods

2.2.1 Bacterial strains

All the experiments described here were based on *E. coli* K-12 BW25113, which is directly derived from BD792, itself a two-step descendent of the *E. coli* K-12 ancestral strain (Hayashi et al., 2006, Bachmann, 1990). All mutant strains analyzed in this study originated from the Keio collection (Baba et al., 2006) and were checked by PCR to verify the presence of the deletion before being used. For this purpose a combination of locus and kanamycin specific primers were used as described in the original publications (Datsenko and Wanner, 2000).

2.2.2 Culture conditions

Bacterial strains were cultured in Luria-Bertani (LB) medium (Sigma Aldrich, USA) supplemented with kanamycin (50μg/ml). In all our experiments, pH was adjusted to neutrality using sterile 5 M NaOH. In the acid adaptation and acid shock experiments pH was adjusted using sterile 1 M HCl. Media were buffered with MES (final solution 10%).

In order to maintain optical density within a narrow range (1.85 – 2.15) and to keep growth rate constant during the experiment, I used a medium replenishment strategy in which medium and cells were removed and replaced by an equal volume of pre-warmed medium (37°C) at regular intervals (every 5 minutes).

Cultures were grown for 16 hours from a single colony at 37°C in a shaking incubator at 200 rpm in 10 ml of LB adjusted to at pH 7. 200 ml of pre-warmed LB medium at pH 7 in a 1 litre conical flask was inoculated with overnight culture to a starting OD600nm = 0.1. The culture was grown at 37°C and 200 rpm until it reached OD600nm = 2. At this point I initiated the medium replenishment procedure by removing 10 ml of culture and adding 10 ml of pre-warmed medium at pH 7 every 5 minutes. Using this procedure I kept the culture at OD600nm = 2 and neutral pH for 1 hour. This ensures that cells reached a steady state before addition of acid (**Figure 2.1**). The technique, aiming to reproduce the conditions of continuous cultures, allows the cells growth at a constant rate and in a constant environment (Herbert et al., 1956).

The pH of the culture was then shifted to pH 5.5 by addition of 14 ml of 1M HCl while maintaining a constant OD for the duration of the experiment (1 hour). After addition of acid, replenishment was performed as before but with buffered medium at pH 5.5.

2.2.3 Phenotypic analysis of mutant strains using flow cytometry

Wild type and mutant strains were phenotypically characterized using flow cytometry analysis (Hewitt et al., 1999, Shapiro and Nebe-von-Caron, 2004). In this application I used propidium iodide (PI) and bis-(1,3-diethylthiobarbituric acid) trimethine oxonol (BOX) to respectively monitor membrane permeability and polarization. Briefly, bacterial cells were added to 2 ml of freshly made, filtered-sterilised (0.2μM filter) PBS supplemented with EDTA, PI and BOX (final working concentrations of 4 mM, 5 μg/ml and 10 μg/ml respectively).

Percentages of healthy and stressed cells were derived from the fluorescence emitted by the cell populations after staining, in relation to reference samples of alive or dead (ethanol-treated) cells (**Figure 2.2**).

2.2.4 Expression profiling by microarray

Ten ml samples of cultures were killed by adding them to a phenol-ethanol solution (final concentration of 19% phenol and 1% ethanol). Samples were left on ice for 20 minutes and then centrifuged (4°C and 5000 rpm) for 10 minutes to recover cell pellets. Stabilized cells were recovered and stored at -80°C. RNA was isolated using the Quiagen RNeasy® kit (Quiagen, USA) according to the manufacturer's instructions. Ten micrograms of input RNA was labelled with using Cy5 labelled dCTP (Amersham Biosciences, USA) using the CyScribe Post-Labelling Kit (Amersham Biosciences, USA), and purified using CyScribe purification Kit (Amersham Biosciences, USA) according to the manufacturer's instructions. Operon *E. coli* Ultra GAPS microarray slides (Corning, USA) were hybridized overnight with 80 pmol of labelled cDNA. The slides were washed in

AdvaWash automated washing station (Adavlytix, USA) and scanned with the ScanArray® GX (PerkinElmer®, USA), using the ScanArray® software.

2.2.5 Data Analysis

Data processing: The single channel array data were normalized using quantile normalisation (Irizarry et al., 2003) in order to correct for systematic errors.

Clustering and Principal Component Analysis (PCA): Multivariate exploratory analysis of the gene expression datasets was performed using a combination of clustering and PCA. In order to identify clusters of genes with similar expression profiles I have used SOTA, a clustering methodology based on an appropriate distance function (Bolstad et al., 2003) with Pearson correlation coefficient as a similarity measure. The relationships between the transcriptional states of the different cell populations were represented using the first two PCs defined by PCA (Bolstad et al., 2003). The number of components was chosen to represent at least 80% of the total sample variance. Bacterial survival and death kinetics, defined by flow cytometry analysis, were visualized using a standard average linkage hierarchical cluster analysis (Raychaudhuri et al., 2000) with Pearson correlation as a similarity measure. Both clustering and PCA were performed using the TMEV software application (Raychaudhuri et al., 2000, Bar-Joseph et al., 2001).

Identification of differentially expressed genes: Differentially expressed genes in the time course experiment were selected using a fold-based rule (the absolute value of the log2 of the ratio between gene expression intensities between pH 5.5 and pH 7 larger than 1.5 at any time point. Note that this corresponds to a 3-fold increase or decrease in a linear scale). This identified 2137 differentially expressed genes. Whenever experimental replicates were available, differentially expressed genes were identified using Significance

Analysis for Microarrays (SAM) (Tusher et al., 2001) as implemented in the TMEV software application (Saeed et al., 2006, Saeed et al., 2003) by employing a 10% FDR threshold, unless otherwise specified in the text.

Functional annotation of the gene lists: Gene lists obtained by cluster or differential expression analysis were assessed for over-representation of KEGG (Ogata et al., 1998) and Gene Ontology (Ashburner et al., 2000) functional terms using the open-source software DAVID (Dennis et al., 2003, Huang da et al., 2009) (http://david.abcc.ncifcrf.gov/). The GO analysis was performed with the lowest level of GO terms. In all cases I used an FDR<1% threshold to define a statistically significant enrichment.

2.3 Results

2.3.1 *E. coli* adaptation to acid involves a rapid but transient transcriptional response

The approach I took in this study was to first characterize the dynamics of the transcriptional response of *E. coli* to acid adaptation. In order to do this, I monitored cultures that were kept in a balanced state of growth by removal of culture and replenishment with pre-warmed medium at regular intervals, a procedure that maintains cells in a constant transcriptional state (**Figure 2.1**).

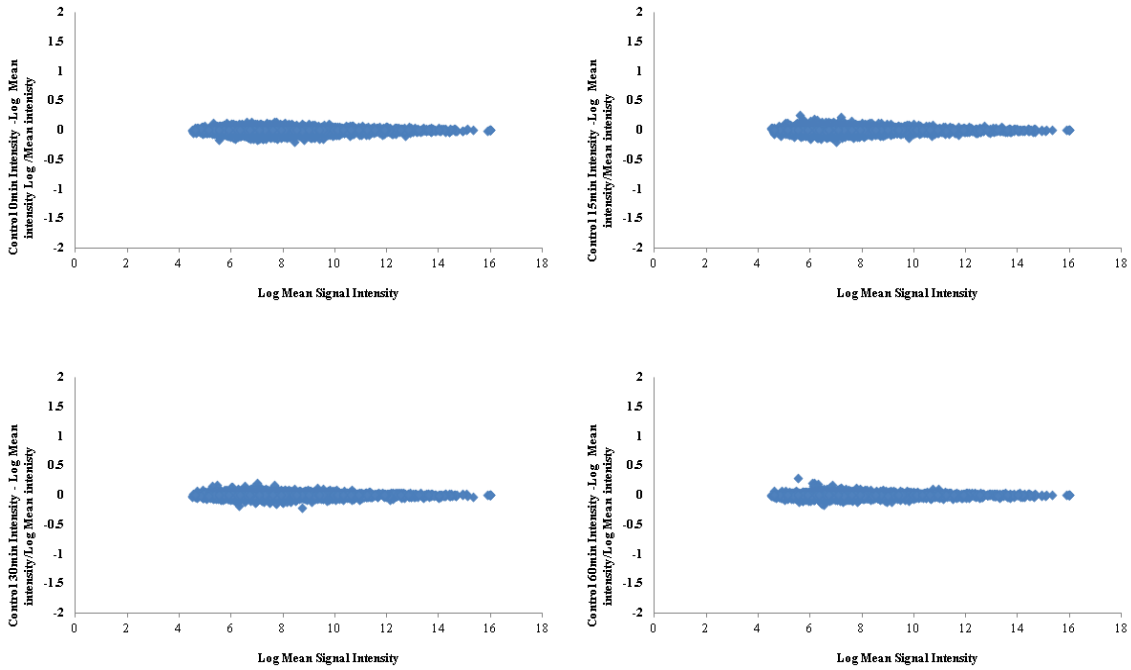


Figure 2.1 Variability between samples at pH 7 in continuous culture experiments.

The above graph shows plots of control pH 7 culture collected after reaching steady state as described in materials and methods at a) 0 min, b) 15 min, c) 30 min and d) 1 hour of growth at pH7 in a microarray experiment. On the X-axis the mean log signal intensity across the four time points is plotted against the log signal microarray intensity of each replicate subtracted from the mean divided by the mean log signal intensity. The graphs clearly show that the variation between replicates is about 5%. The median of coefficient of variation between the control replicates is 10%.

The effect of acidification on gene expression was analysed using microarrays with RNA samples from six time points: prior to treatment, and 30 seconds, 5, 15, 25 and 60 minutes after exposure to pH 5.5. A high level representation of the changes in the cells' molecular state, performed using PCA, revealed that the process is defined by a rapid (5 seconds to 15 minutes) but transient response leading, after one hour, to cells whose transcriptional state is similar to cells grown at neutral pH (**Figure 2.2**). However, the return to pH 7 conditions does not concern the proteome.

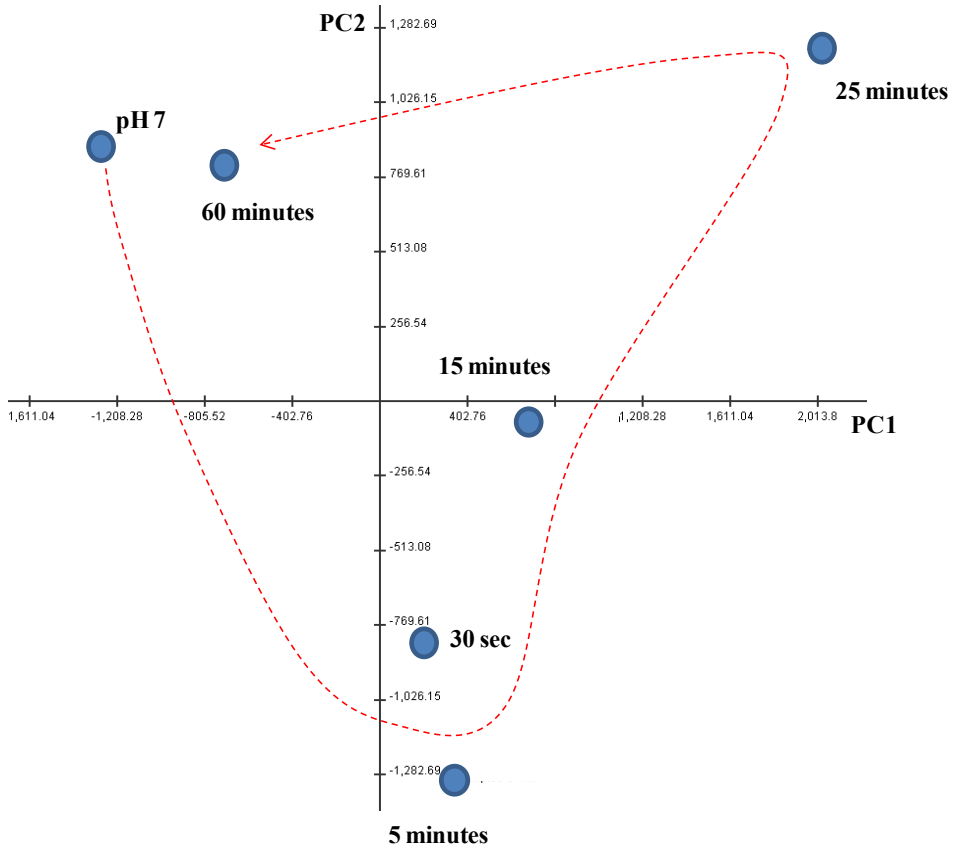
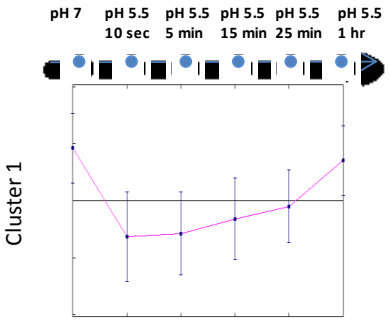


Figure 2.2 Principle component analysis of the transcriptional response of BW25113 to acid adaptation.

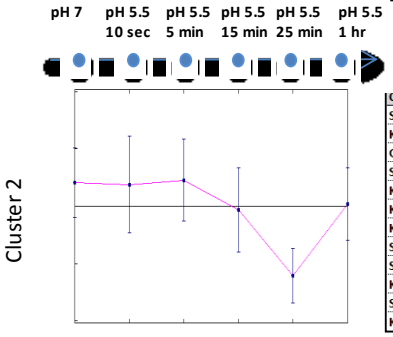
This analysis shows the change in the transcriptional response of BW25113 in the first hour of acid adaptation. The x and y axes represent respectively the first and second components in the PCA.

A more detailed analysis of the gene expression dynamics both by visual inspection of the data and using clustering revealed that the transcriptional events during this period of adaptation could be summarised by four main clusters of gene expression profiles (**Figure 2.3**)



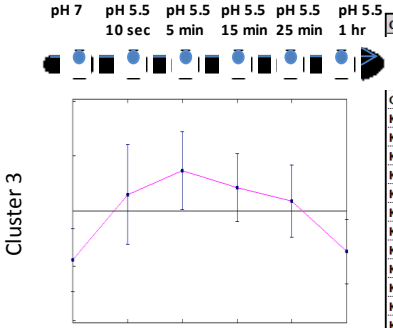
Cluster 1

Category	Term	Count	PValue	Benjamini	ASRS	AFI
GOTERM_CC_FAT	GO:0005618~cell wall	85	8.48E-22	2.09E-20		
GOTERM_BP_FAT	GO:0009060~aerobic respiration	32	2.95E-21	2.02E-18		
KEGG_PATHWAY	ecd00020:Citrate cycle (TCA cycle)	17	4.71E-17	8.02E-14	lysU,	
KEGG_PATHWAY	ecv00190:Oxidative phosphorylation	18	1.86E-16	1.26E-13	cysB,	
KEGG_PATHWAY	eck00620:Pyruvate metabolism	15	7.04E-11	6.66E-09	phoP,	
SP_PIR_KEYWORDS	amino-acid transport	13	3.27E-09	2.79E-08	atpB,	gadA,
KEGG_PATHWAY	eum02010:ABC transporters	26	6.85E-09	5.07E-07	gadX,	gadX
GOTERM_BP_FAT	GO:0042398~cellular amino acid derivative biosynthetic process	18	4.72E-09	2.31E-07	atpD,	
KEGG_PATHWAY	eum00260:Glycine, serine and threonine metabolism	11	4.45E-08	2.23E-06	gadA,	
KEGG_PATHWAY	ece00630:Glyoxylate and dicarboxylate metabolism	10	1.07E-06	4.15E-05	atpI,	
KEGG_PATHWAY	ecr00400:Phenylalanine, tyrosine and tryptophan biosynthesis	8	3.97E-06	1.28E-04	rpoS	
KEGG_PATHWAY	ecd00230:Purine metabolism	13	1.36E-05	3.73E-04		
SP_PIR_KEYWORDS	DNA binding	28	3.96E-20	1.35E-18		



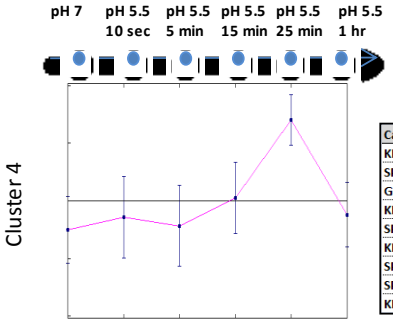
Cluster 2

Category	Term	Count	PValue	Benjamini	ASRS	AFI
SP_PIR_KEYWORDS	transport	134	4.33E-49	2.00E-47		
KEGG_PATHWAY	ecd02020:Two-component system	29	4.81E-21	6.46E-18	adiC,	
GOTERM_BP_FAT	GO:0009061~anaerobic respiration	24	8.37E-13	4.79E-10	adiA,	
SP_PIR_KEYWORDS	nitrate assimilation	9	1.50E-09	3.45E-08	gadB,	hdeA,
KEGG_PATHWAY	ecg00910:Nitrogen metabolism	9	6.41E-07	3.59E-05	ydeO,	hdeD,
KEGG_PATHWAY	ecd02010:ABC transporters	22	2.81E-10	2.22E-08	hdeD,	yhiD.
KEGG_PATHWAY	ecd03010:Ribosome	10	2.79E-06	1.21E-04	gadC,	hdeB,
SP_PIR_KEYWORDS	electron transport	20	1.17E-11	3.23E-10	hdeA,	slp
SP_PIR_KEYWORDS	iron transport	11	3.75E-10	9.44E-09	hdeB,	
KEGG_PATHWAY	ecj00650:Butanoate metabolism	8	1.06E-05	3.65E-04	cadA	
SP_PIR_KEYWORDS	phosphate transport	7	2.25E-08	3.67E-07		
KEGG_PATHWAY	ecd02040:Flagellar assembly	7	9.79E-05	0.0023858		



Cluster 3

Category	Term	Count	PValue	Benjamini	ASRS	AFI
)TERM_CC_FAT	GO:0009274~peptidoglycan-based cell wall	87	9.51E-24	2.00E-22		
_PIR_KEYWORDS	metal-binding	86	5.07E-30	2.50E-28		
GG_PATHWAY	ecw02020:Two-component system	23	9.19E-12	1.38E-08		
GOTERM_BP_FAT	GO:0006091~generation of precursor metabolites and energy	55	2.94E-12	1.02E-09		
KEGG_PATHWAY	eck00520:Amino sugar and nucleotide sugar metabolism	13	4.89E-09	4.32E-07		
KEGG_PATHWAY	ecd00010:Glycolysis / Gluconeogenesis	10	7.99E-07	2.67E-05		
KEGG_PATHWAY	ecd02060:Phosphotransferase system (PTS)	13	1.50E-09	1.50E-07	atpF,	
KEGG_PATHWAY	ecd02010:ABC transporters	22	5.09E-08	2.94E-06	atpH,	
KEGG_PATHWAY	ecx00480:Glutathione metabolism	7	4.32E-06	1.06E-04	evgA,	
KEGG_PATHWAY	eco00052:Galactose metabolism	10	3.54E-07	1.33E-05	gadE,	
KEGG_PATHWAY	ect00230:Purine metabolism	12	1.46E-05	2.97E-04	atpA,	gadE
KEGG_PATHWAY	ecg00190:Oxidative phosphorylation	12	2.04E-08	1.39E-06	phoQ,	
KEGG_PATHWAY	ecx00051:Fructose and mannose metabolism	8	4.46E-05	7.61E-04	rpoA,	
KEGG_PATHWAY	eco00030:Pentose phosphate pathway	8	2.30E-05	4.36E-04	atpG,	
KEGG_PATHWAY	ect00240:Pyrimidine metabolism	9	1.42E-04	0.0021164	crp, lrp	
KEGG_PATHWAY	ecd00620:Pyruvate metabolism	10	2.61E-06	7.12E-05		
KEGG_PATHWAY	ece00630:Glyoxylate and dicarboxylate metabolism	9	4.30E-06	1.08E-04		
KEGG_PATHWAY	ect02030:Bacterial chemotaxis	6	1.95E-04	0.0027019		



Cluster 4

Category	Term	Count	PValue	Benjamini	ASRS	AFI
KEGG_PATHWAY	eco02010:ABC transporters	29	6.63E-16	7.78E-13		
SP_PIR_KEYWORDS	metal-binding	57	1.93E-13	5.49E-12		
GOTERM_BP_FAT	GO:0009061~anaerobic respiration	21	2.75E-10	8.16E-08		
KEGG_PATHWAY	eci02020:Two-component system	15	1.73E-06	1.26E-04	gadW,	gadW,
SP_PIR_KEYWORDS	nucleotide binding	22	2.08E-16	7.88E-15	trmE,	yhiF
KEGG_PATHWAY	ecv00564:Glycerophospholipid metabolism	7	1.82E-05	9.23E-04	adiY	
SP_PIR_KEYWORDS	amino-acid transport	11	3.42E-07	3.35E-06		
SP_PIR_KEYWORDS	lipid metabolism	5	1.17E-04	6.76E-04		
KEGG_PATHWAY	ecj00071:Fatty acid metabolism	5	4.19E-04	0.0121765		

Figure 2.3 Cluster analysis of gene expression profiles in response to acid adaptation.

Genes which showed the largest transcriptional response to pH 5.5 were clustered using SOTA (Irizarry et al., 2003). The four different clusters identified by the analysis are shown on the left by heat maps where green and red correspond respectively to expression levels below and above the mean gene expression values. Clusters 1 and 3 show genes which are transiently down-regulated

whereas cluster 2 and 4 show genes which are transiently up-regulated. The table to the right of the heat maps shows GO and KEGG functional terms significantly enriched in each cluster. (Count: number of genes for each pathway; Benjamini: p-value correction; ASRs: Acid Stress Response Systems genes; AFI: Acid Fitness Islands genes).

The first two clusters represented the earliest response to stimulation, including down-regulated and up-regulated genes (**Figure 2.3**, cluster 1 and 2 respectively). In these two clusters, the largest changes in absolute value were detected five minutes after acid exposure. The second set of two clusters represented genes that are more gradually down or up-modulated in response to acid exposure pH 5.5 (**Figure 2.3**, cluster 3 and 4 respectively) and where the largest change in absolute value was seen 25 minutes post-treatment.

I looked to see whether genes in each cluster were significantly enriched for any specific GO and KEGG functional terms (**Figure 2.3**). Genes represented in the significant terms were then mapped on the BW25113 genome and, as an additional level of quality control, the transcriptional response of genes in operons was analyzed to check that all genes in a given operon showed that same transcriptional response (**Figure 2.4**: A, B, C, D).

OXIDATIVE PHOSPHORYLATION

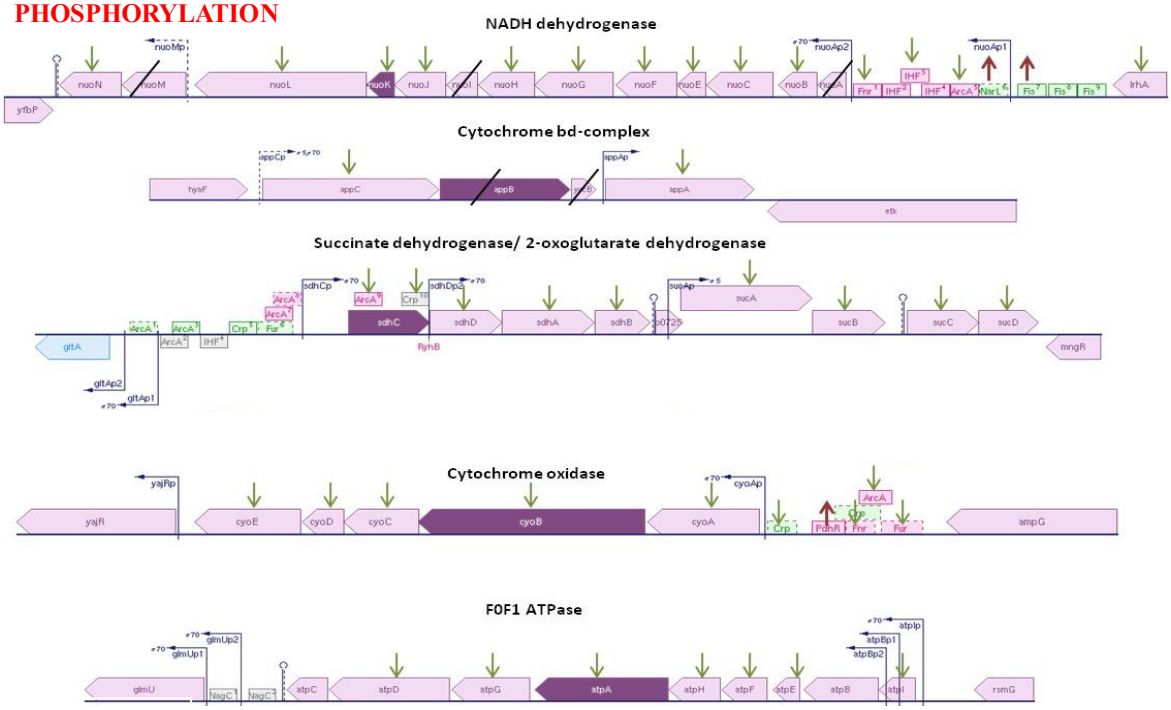


Figure 2.4 (A)

ANAEROBIC RESPIRATION

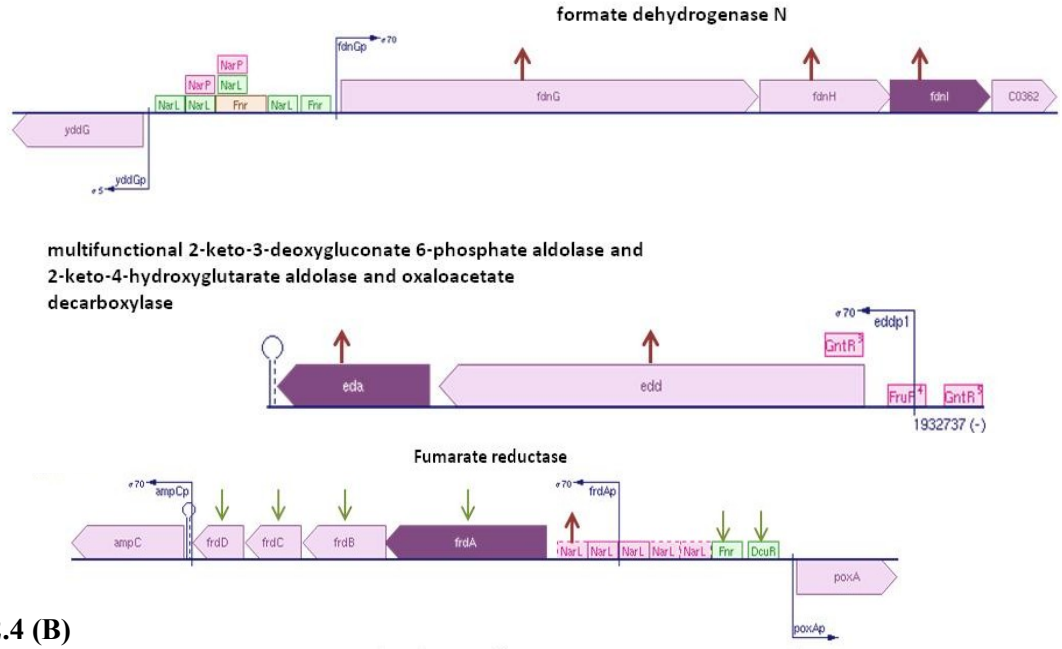


Figure 2.4 (B)

Glycolysis/Gluconeogenesis

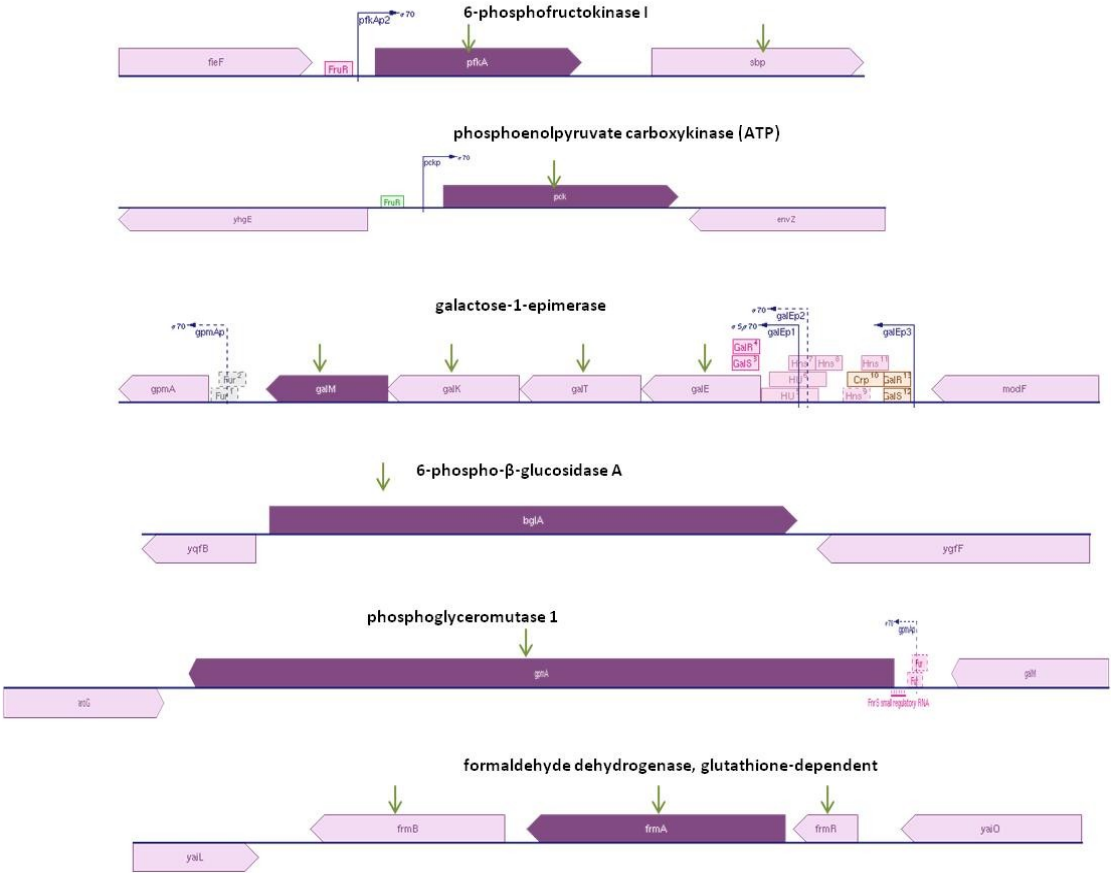


Figure 2.4 (C)

ABC Transporters

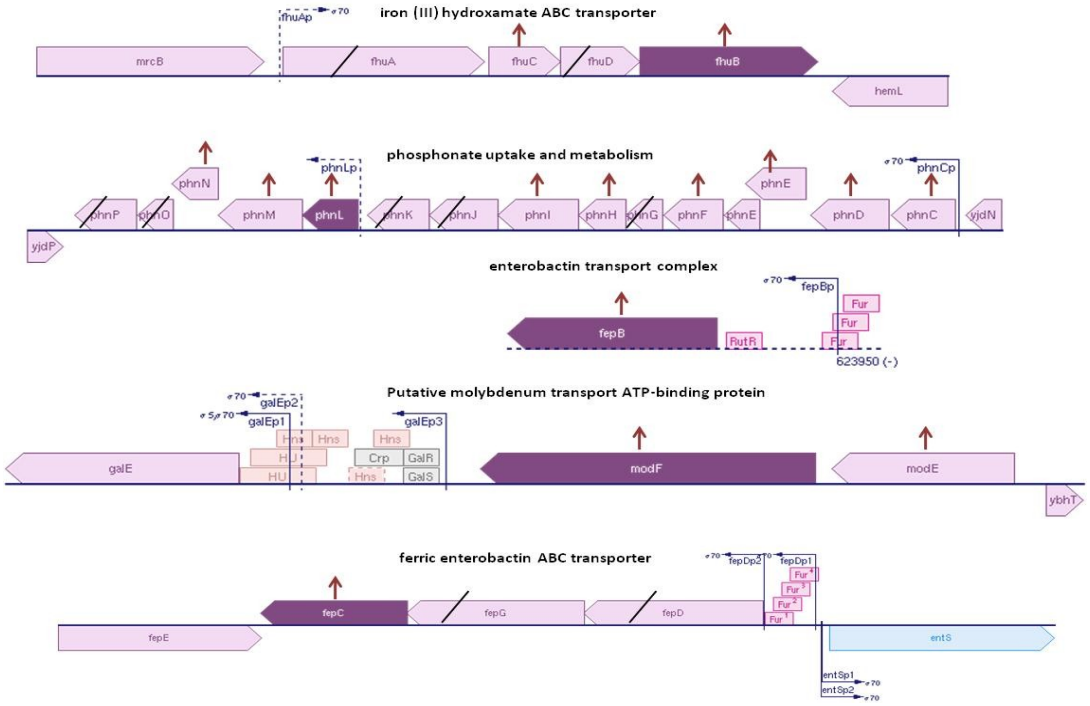


Figure 2.4 (D)

Figure 2.4 Identification of the genes modulated during acid adaptation at operon level.

I have considered the most modulated functions: (A) Oxidative phosphorylation, with some of the representative enzymes; (B) Anaerobic respiration; (C) Glycolysis/Gluconeogenesis; (D) ABC transporters. In the figure are shown some of the components belonging to the modulated functions and their regulation at the operon level. The red arrows are indicative of up-regulation, vice versa for the green arrows.

Consistent with our current understanding of acid adaptation, I observed the up-regulation of many of the genes associated with the glutamate, arginine and lysine dependent acid response systems, and members of the AFI such as *gadW*, *gadX*, *hdeA*, *hdeD*, *hdeB*, *yhiD*, *slp* and *yhiF*. The exceptions found in this preliminary analysis were the genes coding for the isozyme GadA, and the GadX and the GadE regulators, since during early stationary phase they are supposed to be active, following the expression of RpoS (Weber et al., 2005). In my data, *rpoS* was induced at pH 7, suggesting that the considered

conditions could be considered an early stationary phase. Genes coding for the FoF1 ATPase complex (thought to be a component of the amino acid-independent acid stress response system AR1) were also down-regulated. In addition, the transcriptional response to acid exposure involved the concomitant up-regulation of genes involved in anaerobic respiration (GO:0009061) and down-regulation of genes involved in aerobic respiration (GO:0009060). Genes involved in the regulation of the cell wall (GO:0005618) and translation (GO:0006412) were down-regulated whereas genes involved in membrane transport (GO:0055085 ecd02010) were up-regulated in response to acid exposure (**Table 2.1** and **Figure 2.4**).

Table 2.1 Gene expression in response to acid adaptation.

The table list genes belonging to some of the pathways that are down (oxidative phosphorylation, glycolysis and translation) and up-regulated (anaerobic respiration and transport). The p-values obtained from DAVID functional annotation analysis are shown in the right side of the table alongside the direction of change.

Function	Complex	Genes	Regulation	p-value
Oxidative phosphorylation	NADH Dehydrogenase	nuoA	↓	1.18E-11
		nuoB	↓	
		nuoC	↓	
		nuoE	↓	
		nuoF	↓	
		nuoK	↓	
	FoF1 ATPase	atpB	↓	
		atpE	↓	
		atpH	↓	
		atpA	↓	
Glycolysis	Galactose-1-epimerase	galE	↓	3.58E-04
		galT	↓	
		galK	↓	
		galM	↓	
	Phosphoglycerate kinase	epd	↓	
		pgk	↓	
Translation	Glycil-tRNA synthetase	glyS	↓	3.07E-05
		glyQ	↓	
	Lysyl-tRNA synthetase	lysS	↓	
		dsbC	↓	
Anaerobic respiration	Formate dehydrogenase	fdnG	↑	1.20E-06
		fdnH	↑	
		fdnI	↑	
	Trimethylamine N-Oxide reductase	torC	↑	
		torA	↑	
		torD	↑	
Transport	Sulfonate/Nitrate/Taurine ABC Transporter	ssuA	↑	1.20E-06
		ssuB	↑	
		ssuC	↑	
		ssuD	↑	
	Probable proton-driven drug efflux system	yjeP	↑	
		hsrA	↑	

Between the genes up-regulated, I also found several genes coding for membrane proteins involved in osmoregulation (**Figure 2.5**).

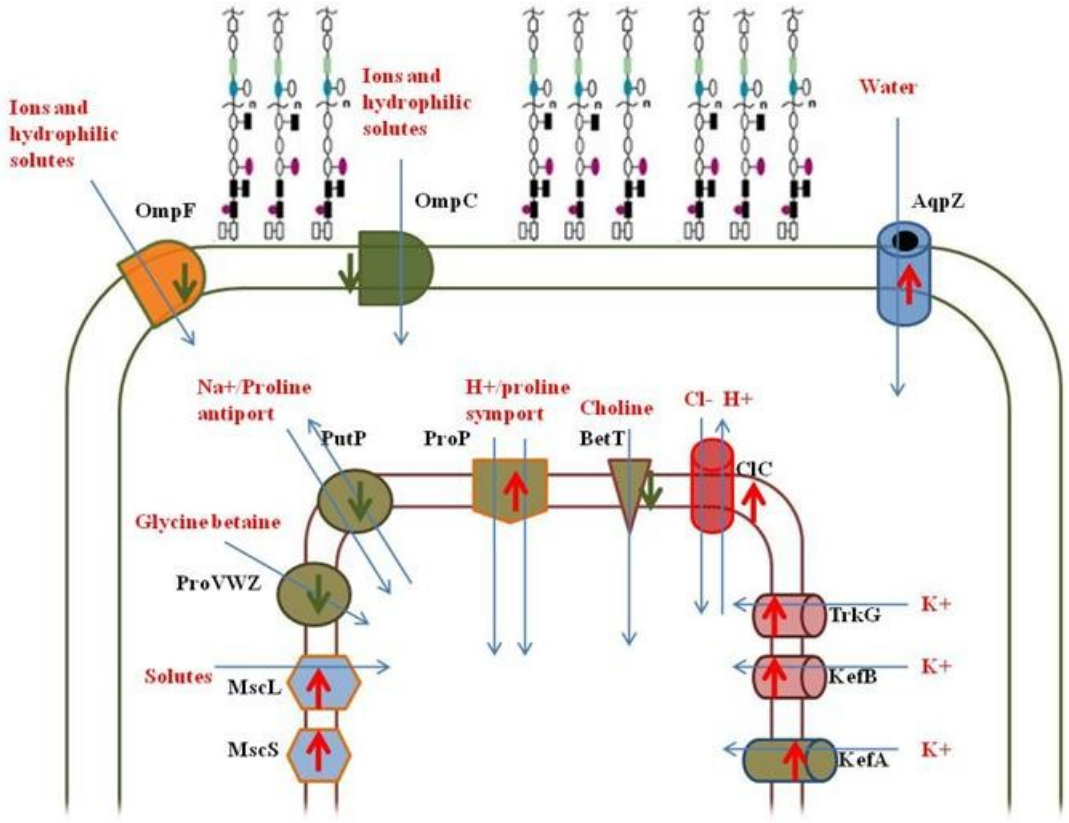


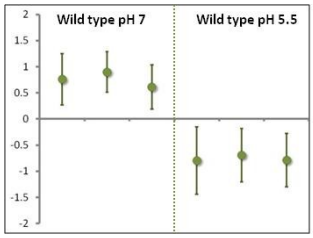
Figure 2.5 Schematic representation of the genes involved in the regulation of the osmotic response.

The proteins of the inner membrane (IM) and periplasmic membrane (PM) are represented in the figure. The green arrows indicate the effect of down-regulation; the red arrows instead are for the up-regulation after acid adaptation.

More specifically, a number of osmoprotectant transporters involved in the response to hyperosmotic shock were transiently up-regulated together with the genes *mscL* and *mscS* coding for the mechano-sensors involved in the response to hypo-osmotic shock. As the analysis of the time course data was carried out using one replicate of the time course, I also compared three replicates of wild type control cells with cells 15 minutes after acid exposure. Statistical analysis of the data identified 1871 differentially expressed genes. The

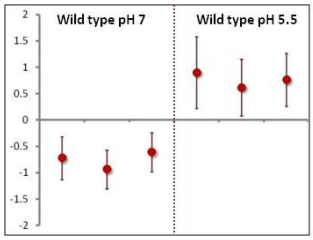
result was largely overlapping with the results of the time course analysis (83% of the genes identified in the time course experiment were also identified by the single time point analysis, (**Figure 2.6** and **Figure 2.7**), showing our time course data to be robust.

SAM 10% FDR - 973 negative genes



Category	Term	Count	PValue	Benjamini	ASRS	AFI
SP_PIR_KEYWORDS	transport	115	6.76E-20	9.16E-19	crp, atpD, phoQ, phoP, lrp, cysB, evgA, rpoS, atpC	
KEGG_PATHWAY	eum03010:Ribosome	25	1.79E-16	1.26E-13		
KEGG_PATHWAY	ecx00230:Purine metabolism	28	6.69E-15	8.07E-13		
KEGG_PATHWAY	ecc00240:Pyrimidine metabolism	24	9.95E-16	2.12E-13		
KEGG_PATHWAY	ecv00970:Aminoacyl-tRNA biosynthesis	16	3.15E-13	2.14E-11		
KEGG_PATHWAY	ecq00250:Alanine, aspartate and glutamate metabolism	15	3.00E-11	1.24E-09		
KEGG_PATHWAY	ecr00061:Fatty acid biosynthesis	10	1.07E-09	3.25E-08		
SP_PIR_KEYWORDS	cell cycle	20	6.08E-12	5.44E-11		
KEGG_PATHWAY	ecg00190:Oxidative phosphorylation	16	2.63E-09	7.43E-08		
KEGG_PATHWAY	eco00270:Cysteine and methionine metabolism	12	1.56E-07	2.82E-06		
KEGG_PATHWAY	ecr02010:ABC transporters	32	1.54E-08	3.74E-07		
KEGG_PATHWAY	ecg00010:Glycolysis / Gluconeogenesis	16	4.79E-10	1.56E-08		
KEGG_PATHWAY	ecd00620:Pyruvate metabolism	23	1.97E-17	3.35E-14		
KEGG_PATHWAY	eck00500:Starch and sucrose metabolism	11	6.57E-06	8.58E-05		
KEGG_PATHWAY	ecg02020:Two-component system	26	2.78E-08	6.05E-07		
KEGG_PATHWAY	eco00520:Amino sugar and nucleotide sugar metabolism	16	8.13E-09	2.09E-07		
KEGG_PATHWAY	ect00260:Glycine, serine and threonine metabolism	11	4.76E-06	6.35E-05		
KEGG_PATHWAY	ecj02060:Phosphotransferase system (PTS)	14	4.09E-07	6.67E-06		
KEGG_PATHWAY	ecr00770:Pantothenate and CoA biosynthesis	9	7.00E-06	9.07E-05		
KEGG_PATHWAY	ecx03440:Homologous recombination	10	1.06E-05	1.35E-04		

SAM 10% FDR - 898 positive genes



Category	Term	Count	PValue	Benjamini	ASRS	AFI
KEGG_PATHWAY	ecd02010:ABC transporters	36	4.19E-21	5.81E-18	evgS, hdeA, gadB, hdeD, adiC, cadC, cadA, gadA, gadC, ydeO	slp, gadA, yhiD, hdeA, hdeD
KEGG_PATHWAY	ecj02020:Two-component system	23	7.51E-12	6.51E-10		
KEGG_PATHWAY	ecd00330:Arginine and proline metabolism	12	6.43E-09	2.88E-07		
GOTERM_BP_FAT	GO:0009061~anaerobic respiration	22	8.64E-09	2.66E-06		
KEGG_PATHWAY	eco00540:Lipopolysaccharide biosynthesis	9	4.27E-07	1.44E-05		
SP_PIR_KEYWORDS	ion transport	22	8.98E-08	1.10E-06		
SP_PIR_KEYWORDS	amino-acid transport	18	2.30E-12	5.63E-11		
KEGG_PATHWAY	ecd00052:Galactose metabolism	8	1.69E-05	5.09E-04		
KEGG_PATHWAY	ecg00053:Ascorbate and aldarate metabolism	6	2.54E-05	7.04E-04		
KEGG_PATHWAY	ecg00630:Glyoxylate and dicarboxylate metabolism	7	3.49E-04	0.0073022		

Figure 2.6 SAM analysis of the three replicates at pH 7 compared with the three replicates at pH 5.5 of the wild type strain, 15 minutes time point.

The SAM analysis (10% FDR) found 973 genes down-regulated and 898 genes up-regulated. After a functional annotation analysis, I have found that most of the functions down-regulated are involved in the energy metabolism and the up-regulated are characterised by genes coding for membrane transporters.

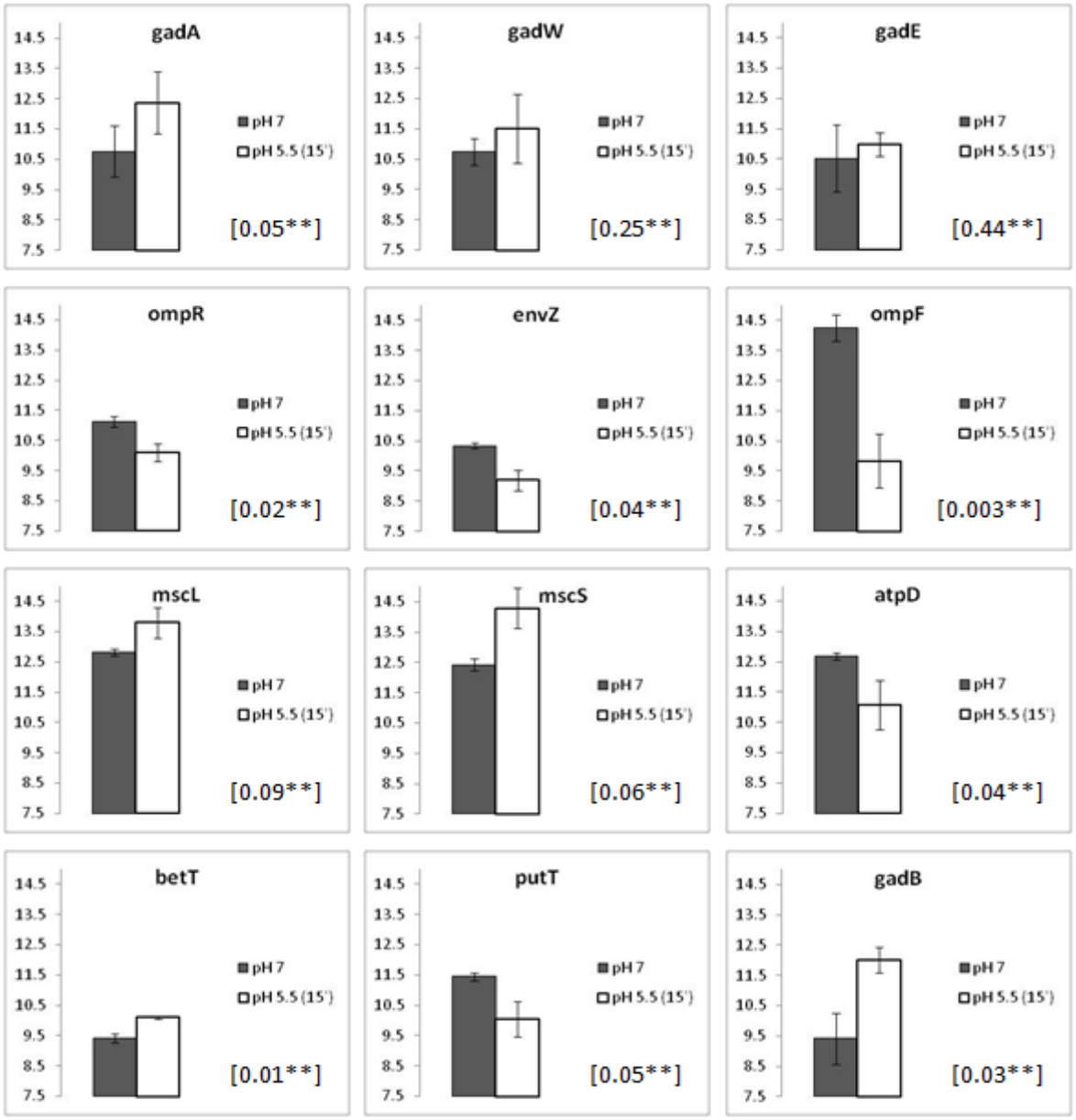


Figure 2.7 Gene expression profiles of selected genes in the three replicates.

Expression levels of each gene are shown at pH 7 and at pH 5.5. The error bars represent standard deviations of the three replicates.

2.3.2 Genes transcriptionally regulated by mild acid are generally required for surviving acid shock

Having performed an initial characterization of the *E. coli* transcriptional response during acid adaptation, I wished to see what fraction of the differentially expressed genes was required for survival during acid shock. In order to address this question, I selected 38 genes representative of the main functions modulated during acid adaptation (**Table 2.2**) and tested the ability of strains where each of these genes was deleted to survive in strong acid conditions (pH 2.5), following the protocol in **Figure 2.8**..

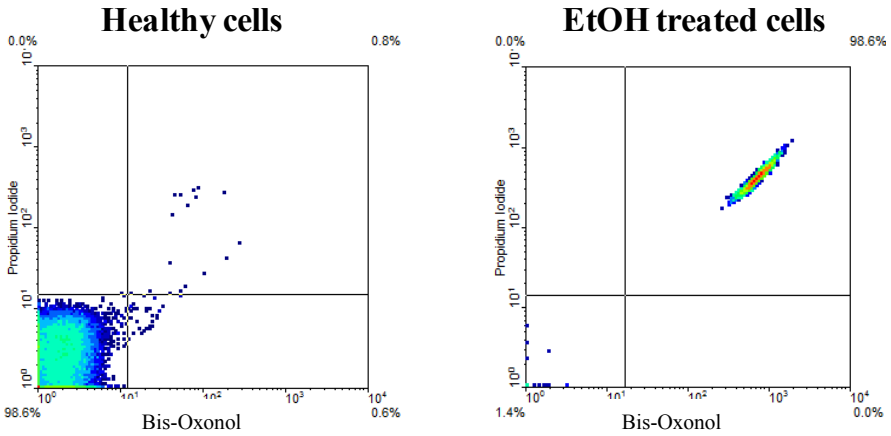


Figure 2.8 Variability between samples at pH 7 in continuous culture experiments.

The above graph shows plots of control pH 7 cultures collected after reaching steady state as described in materials and methods at a) 0min, b) 15min, c) 30min and d) 1hour of growth at pH7 in a microarray experiment. On the X-axis the mean log signal intensity across the four time points is plotted against the log signal intensity of each replicate subtracted from the mean divided by the mean log signal intensity. The graphs clearly show that the variation between replicates is about 5%. The median of coefficient of variation between the control replicates is 10%.

Table 2.2 Selection of the genes from the wild type time course.

Selected gene DOWN-regulated	Function
crp	cAMP receptor protein
fliA	RNA polymerase sigma factor for flagellar operon
fnr	Fumarate and nitrate reduction regulatory protein
gadE	Transcriptional regulator of genes involved in acid resistance
gcvP	Glycine dehydrogenase
hemE	Oxygen sensor protein
ihfA	Integration host factor subunit a
ihfB	Integration host factor subunit b
osmC	Osmotically-inducible protein C
phoH	Phosphate starvation-inducible protein
phoP	Member of the two-component regulatory system phoQ/phoP involved in adaptation to low Mg2+
phoQ	Member of the two-component regulatory system phoQ/phoP involved in adaptation to low Mg2+
pyrF	Orotidine 5'-phosphate decarboxylase
rpoS	RNA polymerase sigma factor for protection against external stresses
sucA	2-oxoglutarate dehydrogenase E1 component
sucB	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex
Selected gene UP-regulated	**Function**
aceE	Pyruvate dehydrogenase E1 component
adiA	Biodegradative arginine decarboxylase
cadA	Lysine decarboxylase
cadB	Cadaverine/lysine antiporter
fis	Factor-for-inversion stimulation protein
gadB	Glutamate decarboxylase b
gadW	HTH-type transcriptional regulator of genes involved in acid resistance
gadX	HTH-type transcriptional regulator of genes involved in acid resistance
gcd	Quinoprotein glucose dehydrogenase
grxA	Glutaredoxin-1
hdeA	Chaperone-like acid stress protein
hdeB	Chaperone-like acid stress protein
hslJ	Heat shock protein
narJ	Nitrate reductase molybdenum cofactor assembly chaperone
narX	Nitrate/nitrite sensor protein
rpoD	The primary sigma factor of E. coli
spy	Spheroplast protein, could be invovled in zinc metabolism
trmE	tRNA modification GTPase
ycaD	Uncharacterized MFS-type transporter
ydeO	HTH-type transcriptional regulator of genes involved in acid resistance
yjbJ	Predicted stress response protein

The mutants were assayed at different time points (5, 30, 60, 90, 120, 150 and 180 minutes) after acid exposure, and alive and dead cells were quantified by flow cytometry analysis. All tested mutants showed a detectable increase in sensitivity to low pH in the assay, though some effects were small (**Figure 2.9**, **Table 2.3** and **Table 2.4**).

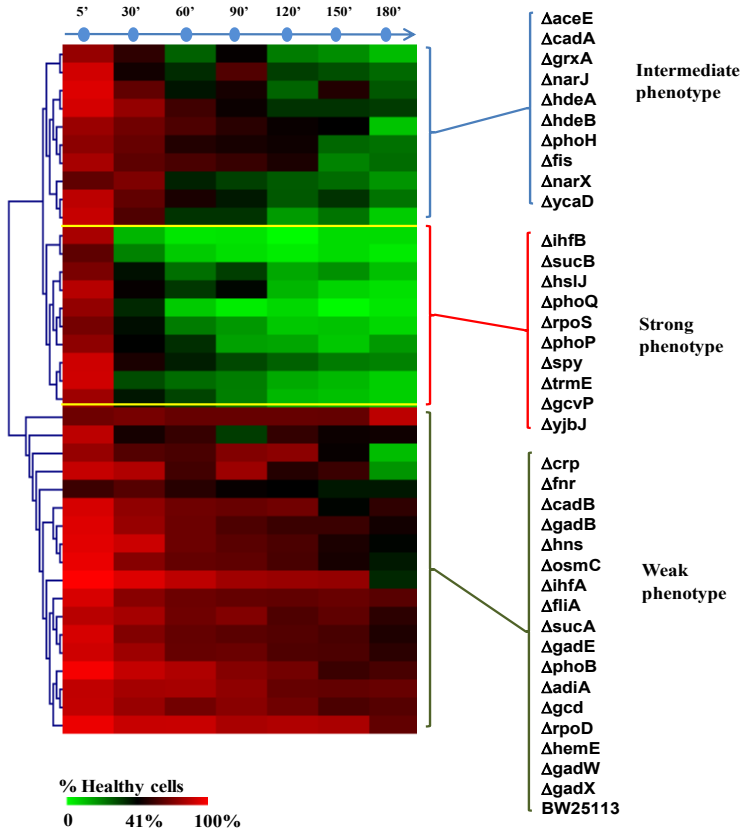


Figure 2.9 Clustering of mutant strains according to their phenotypic response to acid shock.

Survival of exposure to LB at pH 2.5, without prior adaptation, was monitored over time using flow cytometry, taking samples every 30 minutes for three hours. The colour bar indicates the percentage of survival determined by flow cytometry analysis at each time point. Hierarchical clustering of the survival time course data revealed three clusters of strains on the basis of their survival profiles at pH 2.5, defined here as weak, intermediate and strong phenotypes.

Table 2.3 Phenotypes p-values for gene KO and WT not adapted.

NOT ADAPTED	5' Acid	30' Acid	1 hour Acid	1 hr 30' Acid	2 hours Acid	2 hrs 30' Acid	3 hours Acid
ΔaceE	0.01462 5.4E-09	2.2E-10	8.69E-25 7.48E-25	1.25E-07	5E-21 2E-16	2E-18	8.4E-21 6.6E-12
ΔadiA	0.06828 0.06226	0.00876	0.003157 2.93E-10	0.087984	0.0094 2E-16	0.0018	0.37071 6.6E-12
ΔcadA	0.25351	6.2E-05 1.4E-23	2.8E-12 4.83E-27	0.000382	5E-09 6E-08	1E-07	0.00065 0.01769
ΔcadB	0.5429 2E-09	2.2E-13 0.00045	2.54E-16 0.001868	0.002945 0.166394	9E-26	4E-29 9E-29	7.8E-15 1E-14
Δcrp	1.4E-07	0.35534 1.4E-09	0.564475 4.86E-13	0.084494 7.68E-09	0.1488 0.0125	0.4457 0.0589	0.00061
Δfis	0.65632 1.8E-21	0.00231 0.01376	5.87E-08 1.89E-26	0.00017 2.53E-31	3E-06 3E-09	1E-07 1E-08	2.4E-06
ΔfliA	0.55652 0.25034		4.72E-05 0.000286	0.015941	0.0001 2E-29	 2E-32	0.00065 6.3E-20
Δfnr	0.0003 0.08236	3.8E-12	2.57E-14 0.0016	6.21E-13	1E-08 0.0003	1E-07	1.8E-07 0.02016
ΔgadB	0.31098 0.00032	0.35472	0.145061 1.85E-17	0.878042	0.6477 0.0045	0.0004	3.2E-08 0.56451
ΔgadE	0.32649	0.4173	0.891394	0.494995	0.7008 3E-17	0.7152	0.00051 8.6E-20
ΔgadW	0.01411	0.00876	0.001978	0.15801	0.0092	0.0067	0.36766
ΔgadX	0.06828	0.00359	0.000515	0.084125	0.007	0.0014	0.24428
Δgcd	0.09307	4.2E-06	5.6E-09	3.26E-07	3E-06	0.0004	0.00451
ΔgcvP	0.00701 0.7608	9.2E-11 2E-32	4.69E-18 1.94E-31	1.77E-09 5E-34	4E-17	1E-24	6.8E-18
ΔgrxA	0.04931 1	2.1E-07	1.03E-12 9.35E-20	8.31E-09	1E-12 5E-33	8E-07	0.03779
ΔhdeA	0.15328 2.2E-08	6.8E-07 6.7E-06	3.51E-11 1.85E-06	1.12E-11 9.77E-05	1E-19 2E-28	7E-21	4.6E-16 2.1E-19
ΔhdeB	0.04416 2E-09	2.5E-05 2.5E-08	9.72E-14 5.71E-11	3.91E-10 2.69E-07	2E-11 3E-05	4E-15 0.0014	5.1E-08
ΔhemE	0.56312	1	0.499132	0.169731	0.1157 1E-07	0.0006 1E-16	0.2909
Δhns	0.09307 1.1E-29	0.00046 2.2E-06	1.76E-07 2.96E-10	5.81E-07 5.62E-08	4E-10	1E-10	
ΔhslJ	1.1E-09 2.5E-05	2.4E-14	4.48E-18 2.11E-30	1.28E-14	2E-23 3E-07	2E-24 2E-08	4E-13
ΔihfA	0.67395		0.001048	0.000927	1E-05 1E-32		0.02347 7E-26
ΔihfB	0.01567	2.5E-29	1.27E-42	5.02E-32	1E-47	7E-33	
ΔnarJ	0.20821 0.12418		3.02E-08 1.89E-07	2.55E-07	1E-12		0.0001
ΔnarX	0.00016 0.25351	5E-08	6.6E-14 8.44E-11	4.37E-12	1E-14 2E-12	9E-15	1.2E-06 3.5E-08
ΔosmC	0.35778	0.00253	0.000461	0.000889	0.001 3E-21	8E-10	0.00615 3.3E-11
ΔphoB	0.28497	0.00033	9.38E-06	0.004672	0.0016	0.0011	0.77001
ΔphoH	0.00328 0.00027	6E-07 3.2E-05	2.58E-12 2.53E-05	2.72E-10 0.014569	8E-13	8E-14	5.4E-08
ΔphoP	0.0034 6.5E-13	4.2E-12 5.8E-18	1.02E-28 1.65E-22	3.16E-24 2.59E-17	2E-29 0.0019	5E-28 2E-30	1.6E-19
ΔphoQ	0.03514 0.00019	1E-10 1.9E-20	2.27E-17 3.35E-22	3.86E-07 1.21E-14	3E-28 6E-26	3E-32	8.7E-21
ΔrpoD	0.04796 0.26228	0.0024	4.76E-06 6.2E-08	2.85E-05	1E-06 2E-26	0.0002	0.35803
ΔrpoS	0.00062 8.1E-05	2.4E-16	9.71E-40 1.98E-33	2.32E-34	9E-37 0.0004	2E-38	2.8E-25 0.28719
Δspy	5.1E-08 0.00021	5.3E-14	2.17E-18 2.62E-18	7.25E-26	1E-26 1E-28	1E-30	2.6E-09 4.2E-20
ΔsucA	1 1	0.01174 5.3E-06	7.5E-07 1E-06	0.00358 0.000254	3E-05 8E-25	3E-07	0.00139 2.9E-15
ΔsucB	4.9E-08 1.8E-11	2.5E-16 2E-45	3.83E-29 3.78E-51	4.59E-30 7.96E-45	2E-37 0.0001	1E-31	2E-25
ΔtrmE	0.0579	1.2E-05 1.2E-22	2.91E-08 1.98E-33	3.2E-06 5.44E-34	3E-09 6E-49	4E-10	8.7E-05
ΔycaD	0.0013 0.14602	3E-09	2.57E-16 2.78E-21	1.8E-14	5E-22 1E-36	1E-17 6E-35	2E-21
ΔyjbJ	0.00129	8.6E-13	3.74E-19	1.56E-18	1E-28	5E-25	5.3E-16

Table 2.4 Phenotypes p-values for gene KO and WT adapted.

Adapted	5' Acid	30' Acid	1 hour Acid	1 hr 30' Acid	2 hours Acid	2 hrs 30' Acid	3 hours Acid
ΔaceE	0.075344 0.083409	1.03E-11	0.0072696 8.913E-07	3.841E-05	3.9E-11 7E-16		6.54E-15 1.1E-07
ΔadiA	0.878016 1	0.042762	0.6154029 0.0915942	0.6397546	0.2548 0.16437	0.95048	0.230664 0.020962
ΔcadA	1 1.36E-07	1.74E-06 4.1E-09	0.1649808 2.863E-08	0.1018168 8.186E-07	0.7431 9.4E-08	0.18051 1.3E-05	0.006344
ΔcadB	0.17636 0.574604	1.46E-06 1.4E-09	4.116E-07 7.616E-10	1.363E-15 1.861E-10	2.6E-14 1.7E-11	7.3E-08 4.9E-11	2.82E-13
Δcrp	 0.789608	0.308973 0.20964	0.0832009 0.0118623	0.0053511 0.0009768	0.03088 4.3E-05	0.00186 6E-06	0.10117
Δfis	0.004501 0.000542	3.51E-10 4.36E-12	2.595E-14 2.313E-11	1.111E-13 2.45E-09	4.6E-05 2.6E-11	2.2E-15 7E-08	1.35E-22 6.36E-12
ΔfliA	0.546682 0.714445		0.0965822 0.8906121	0.0176328	0.00012 0.46369		1.59E-07 1.16E-06
Δfnr	0.318917 0.682717	0.000818	0.0432246 0.0331891	0.0023893	0.06904 0.0007	0.02061	0.013721 0.000123
ΔgadB	0.012873 0.362411	0.002142	0.0025326 0.0001424	0.0006941	0.12124 0.00052	0.01533	8.17E-05 0.003703
ΔgadE	1	0.077068	0.0573446	0.2011729	0.00946	0.0205	3.16E-14
ΔgadW	0.23513	0.205339	0.8103924	0.1889988	0.06232	0.16803	0.017468
ΔgadX	0.574604	0.029604	0.7726554	0.8361599	0.24237	0.04366	0.534169
Δgcd	1	3.28E-07	2.345E-10	3.43E-09	1.7E-13	0.00517	2.64E-06
ΔgcvP	 0.670637	3.87E-14 2.4E-21	9.035E-17 7.685E-18	2.127E-08 4.834E-17	2.2E-14 6.2E-24	8.4E-16	2E-18
ΔgrxA	0.057137 0.174001	1.07E-07	1.779E-12 4.047E-22	5.483E-13	5.1E-17 5.1E-35	0.0009	2.86E-12 5.05E-31
ΔhdeA	0.544708 5.48E-08	0.000328 4.98E-05	0.0159322 0.0036791	3.699E-05 0.0037073	4.4E-06 0.00053	0.00385 0.29809	7.73E-10
ΔhdeB	6.59E-06 0.107933	2.8E-06 3.28E-15	1.876E-09 8.386E-21	3.827E-09 2.338E-23	7.9E-16 2.2E-26	5.1E-10 2E-20	4.89E-11
ΔhemE	1	1.9E-09	0.0003275	7.779E-09	1.4E-14	8.8E-08	
Δhns	1 2.37E-05	0.934674 0.000203	0.7014699 1.251E-10	0.5832177 1.855E-12	0.06495 3.9E-10	0.39455 0.00478	0.609855
ΔhslJ	9.53E-05 0.001909	2.02E-22	2.709E-23 2.576E-35	1.879E-24	6.8E-24 7.7E-41	2E-18	1.25E-24 1.48E-34
ΔihfA	0.341107		0.5706497	0.0100294	0.02024		5.21E-06
ΔihfB	0.003852	3.11E-31	2.489E-38	7.414E-39	3.5E-43	6.6E-33	
ΔnarJ	0.028442 0.760854		2.514E-10 0.00127	1.331E-10	6.4E-11 0.00012		1.11E-19 2.74E-21
ΔnarX	0.001453 0.648651	2.01E-16	6.157E-17 9.894E-25	7.956E-12	2.2E-24 1.6E-31	1.5E-17	4.91E-16 7.19E-29
ΔosmC	0.661035	2.75E-07	0.0002735	2.585E-05	2.3E-06	0.00609	0.019574
ΔphoB	0.458968	0.060055	0.47284	0.0818702	0.04829	0.74503	0.006774
ΔphoH	0.042711 1.3E-07	5.15E-05 4.19E-07	0.0004196 1.359E-06	0.0001068 7.666E-05	5E-07 2E-07	0.0003	8.83E-08
ΔphoP	0.247232 0.016086	3.51E-05 1.16E-07	1.611E-06 8.401E-11	1.45E-07 7.264E-11	1.2E-10 1.4E-18	0.00012	5.85E-07
ΔphoQ	0.107349 0.029016	8.02E-05 0.000224	1.313E-07 0.0006113	7.916E-18 0.0001233	1.6E-21 0.00012	2.1E-15	5.25E-22
ΔrpoD	0.132226 0.737511	0.001913	0.009447 0.0963165	0.0015525	5.7E-05 0.00063	0.0196	3.23E-05 0.000346
ΔrpoS	0.191247 0.045437	0.005896	2.734E-08 5.821E-07	1.593E-13	2.1E-14 9.7E-07	1.7E-07	8.62E-19 7.88E-07
Δspy	0.450564 0.346179	6.61E-09	1.762E-07 4.915E-08	5.684E-12	1.4E-13 2.2E-17	7.9E-09	1.31E-11 4.17E-19
ΔsucA	0.682717 0.05026	0.000509 1.18E-07	0.0079459 2.774E-05	3.479E-05 3.005E-05	5.9E-09 3.8E-07	0.00179	1.29E-06
ΔsucB	0.182463 1.47E-10	3.52E-05 6.12E-48	3.569E-06 1.833E-43	1.45E-07 1.016E-43	0.00052 7.5E-45	1.5E-06	6.02E-09
ΔtrmE	0.260049 0.000138	0.005123 7.77E-18	0.0501222 1.809E-20	0.0300787 1.52E-25	0.00518 2.1E-28	0.12411 1.1E-22	0.002365 1.82E-21
ΔycaD	0.079489 0.004901	1.09E-15	7.44E-29 3.524E-37	2.407E-31	9.9E-35 3.8E-38	1.3E-27	1.58E-30 2.06E-36
ΔyjbJ	0.217782	1.7E-22	5.91E-30	4.08E-28	1.8E-36	6.2E-26	

Tables 2.3 and 2.4: Mutant strains showing a significant acid response phenotype.

The tables show, for each mutant, the p-values obtained from the Fisher exact test compared to the Wild Type. P-values were computed on the percentages of healthy cells as defined by the flow cytometry. Where two p-values are present, two biological replicates were considered. Table 2.3 shows the values of the not adapted mutants; in Table 2.4 are shown the values of the adapted samples.

In order to define the severity of the phenotype, I have considered a threshold which allowed me to cluster the mutants in three different groups: strong, weak or intermediate phenotype. I have considered 30 minutes of exposure at pH 2.5 the parameter for the definition of the phenotype: strong phenotype mutants show less than 35-40% of healthy cells after 30 minutes at pH 2.5, intermediate phenotype show between 60 and 40% of healthy cells, while the weak phenotype show more than 60% of healthy cells (**Figure 2.9**). The choice of 30 minutes exposure, although arbitrary, relies on the clear difference between the mutants to acid exposure. I did not consider the time point 180 minutes since for the majority of the mutants, the number of healthy cells was nearly close to 0 for the three groups of phenotypes. Cluster analysis of the survival and death kinetic profiles was used to classify the mutant strains into three main groups which displayed strong, weak or intermediate phenotypes (**Figure 2.9**). The wild type strain, included for comparison, shows the weakest phenotype, as expected, as this strain is quite acid resistant under the growing conditions of this assay.

Genes that when mutated gave a strong phenotype include transcriptional regulators (*phoP*, *rpoS* and *ihfB*), a glycine decarboxylase (*gcvP*), the 2-oxoglutarate dehydrogenase (*sucB*), and the two-component system sensor *phoQ*. Strains with an intermediate phenotype included mutants of two acid stress chaperones (*hdeA* and *hdeB*) and nitrate reduction (*narJ* and *narX*). Strains with a weak phenotype included mutants of several transcriptional regulators (*crp*, *rpoD*, *hns*, *phoB*, *ihfA*, *fliA*, *fnr*) and, interestingly, included three of the GAD system specific regulators (*gadE*, *gadW* and *gadX*).

2.3.3 Defective expression of cell wall and energy metabolism genes correlates with increased sensitivity to acid

The experiments described above showed that the inactivation of genes that are transcriptionally regulated during adaptation often results in a significant loss of survival following acid shock. I reasoned that analysis of the transcriptional response to acid exposure in the different phenotypic groups would enable me to formulate hypotheses concerning the molecular pathways that are more important for survival of acid stress. Therefore, I subjected the same 38 mutants used above to a microarray analysis after acid adaptation and asked whether any component of the transcriptional response was predictive of loss of acid resistance. For practical purposes, this experiment was performed on a single time point (15 minutes after exposure at pH 5.5) which is where the highest change in gene expression in the early response clusters of the wild type strain occurred (**Figure 2.3**). A PCA analysis confirmed that the acid adaptation transcriptional program was defective in the mutant strains (**Figure 2.10**). The effects of the mutation compared to the wild type strains were not considered for this purpose.

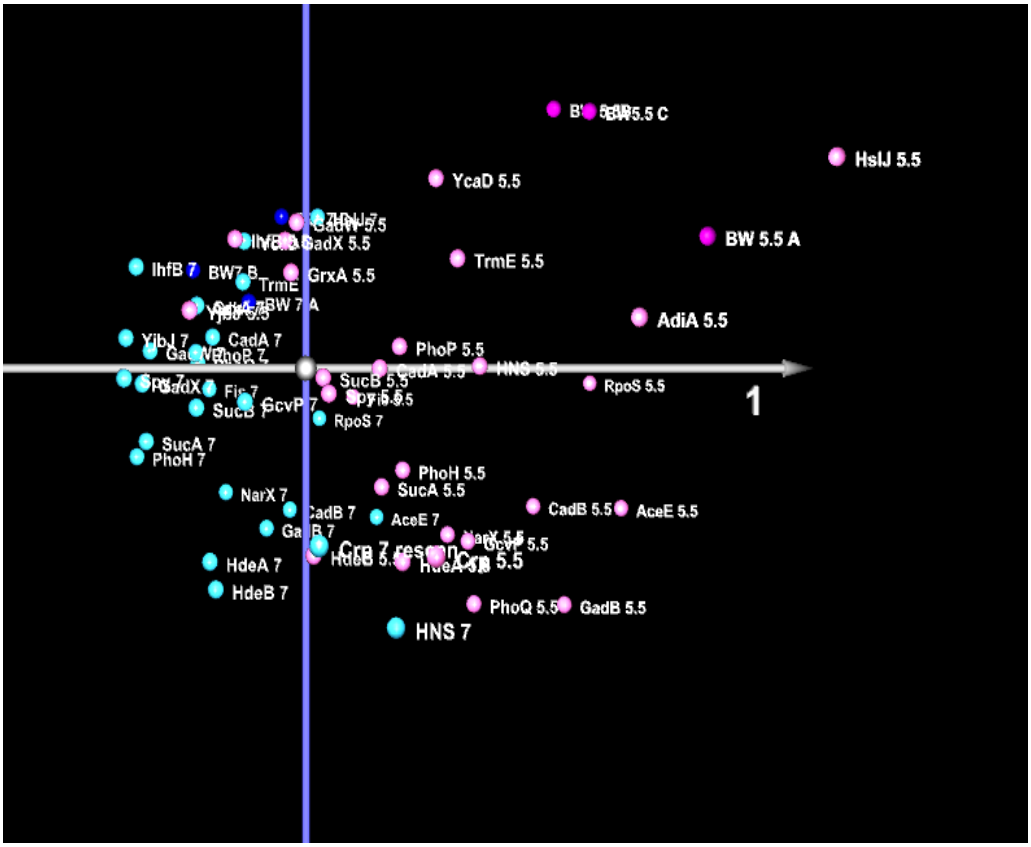


Figure 2.10 Principal component analysis of the gene KO.

The three replicates of the wild type are indicated in blue for the pH 7 and in red for the pH 5.5 treatment. The mutants at pH 7 are represented by the light blue dots, at pH 5.5 by the pink dots. The first component, which defines the shift following acid exposure, is on the X axis.

In order to identify molecular pathways linked to the severity of the phenotype I compared the gene expression profiles (expressed as a log-ratio between expression values at pH 5.5 and pH 7) between the wild type (represented by four replicated experiments) and mutated strains that showed either an intermediate or a strong phenotype. I found 221 genes to be differentially expressed between the three experimental groups (FDR<10%). Among the list of genes up-regulated in the wild type strain I found some of them involved in anaerobic respiration (hyfC, nrfF, menA), sugar transport (*xylE*, *ulaA*, *ycjP*, *sgcC*, *malF*, *agaD*, *agaV*), lipopolysaccharide biosynthesis (*ylbH*, *rfaF*, *rfe*, *rfaJ*, *rhsA*) and purine

metabolism (*yqeA*, *nudF*, *allB*), whereas the list of down- genes was enriched in aerobic respiration genes (including oxidative phosphorylation and TCA cycle) (*ubiA*, *nuoF*, *acnB*, *sdhC*, *acnA*, *sdhD*, *ppc*, *nuoN*, *cyoA*, *cyoB*, *cyoD*), DNA repair (*uvrC*, *xthA*, *nei*, *exoX*) and cell cycle (*mrdB*, *ftsY*, *mukE*, *ftsA*) (**Figure 2.11** and **Table 2.5**).

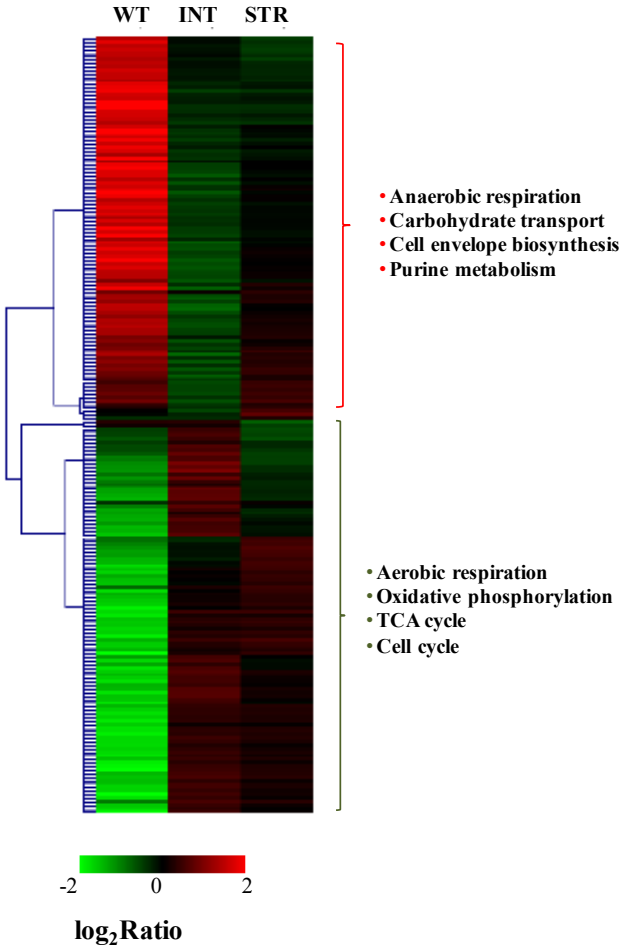


Figure 2.11 Differential gene expression on exposure to acid varies between phenotypic groups.

221 genes were identified which showed significant differential expression between the wild type (WT) and the mutants belonging to the intermediate (INT) and strong (STR) phenotype groups. The annotation on the right refers to functional groups of genes which were down-regulated (green) or up-regulated (red) in the wild type strains at pH 5.5 relative to their expression at pH 7. These are shown in the first column of the heat map. In the second and third columns, the means of the log ratios of expression at pH 7 and pH 5.5 for these same genes are shown for the intermediate and strong groups.

I observed that the large majority of the genes predictive of phenotypic outcome (89%) were also differentially expressed in response to acid exposure (**Figure 2.11**). Remarkably, the ratios of the expression between cells grown at pH 5.5 and at pH 7 were generally reversed in mutant strains that had an intermediate phenotype, while mutant strains with a strong phenotype were characterized by a minor ability to regulate most of these genes.

Table 2.5 Enriched functions of the genes differentially expressed between the three phenotypic groups

The analysis performed between the three groups considering log2 ratios of gene expression values gave me a list of 221 genes. I have performed a functional annotation analysis, results of which are shown in this table. I have selected the functions more represented and the representing genes.

GO_BP Nitrogen compound biosynthetic process	
aroM	Protein aroM
asnC	Asparaginyl-tRNA synthetase; Regulatory protein asnC
bioF	8-amino-7-oxononanoate synthase
cysJ	Sulfite reductase [NADPH] flavoprotein alpha-component
fliY	Cystine-binding periplasmic protein
folE	GTP cyclohydrolase 1
gltI	Glutamate/aspartate periplasmic-binding protein
ilvM	Acetolactate synthase isozyme 2 small subunit
metA	Homoserine O-succinyltransferase
nadC	Nicotinate-nucleotide pyrophosphorylase [carboxylating]
ndk	Nucleoside diphosphate kinase
panD	Aspartate 1-decarboxylase
pheL	Phe operon leader peptide
pheP	Phenylalanine-specific permease
potA	Spermidine/putrescine import ATP-binding protein potA
serS	Seryl-tRNA synthetase
thiD	Phosphomethylpyrimidine kinase
yqeA	Carbamate kinase-like protein yqeA
SP_PIR Lipopolysaccharide biosynthesis	
eptB	Phosphoethanolamine transferase
lpcA	Phosphoheptose isomerase
rfaF	ADP-heptose--LPS heptosyltransferase 2
rfaJ	Lipopolysaccharide 1,2-glucosyltransferase
rfe	Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase
SP_PIR Sugar transport	
agaD	N-acetylgalactosamine permease IID component
agaV	N-acetylgalactosamine-specific phosphotransferase enzyme IIB component 2
malF	Maltose transport system permease protein malF
sgcC	Putative permease IIC component
ulaA	Ascorbate-specific permease IIC component ulaA
xylE	D-xylose-proton symporter
ycjP	Inner membrane ABC transporter permease protein ycjP
GO_BP Anaerobic respiration	
acnA	Aconitate hydratase 1
acnB	Aconitate hydratase 2
hyfC	Hydrogenase-4 component C
hyfE	Hydrogenase-4 component E
menA	1,4-dihydroxy-2-naphthoate octaprenyltransferase
nrfF	Formate-dependent nitrite reductase complex subunit nrfF
nuoF	NADH-quinone oxidoreductase subunit F
nuoN	NADH-quinone oxidoreductase subunit N
KEGG ABC Transporters	
cydC	ATP-binding/permease protein cydC
fhuD	Iron(3+)-hydroxamate-binding protein fhuD
fliY	Cystine-binding periplasmic protein
gltI	Glutamate/aspartate periplasmic-binding protein
malF	Maltose transport system permease protein malF
potA	Spermidine/putrescine import ATP-binding protein potA
GO_BP Aerobic respiration	
acnA	Aconitate hydratase 1
acnB	Aconitate hydratase 2
cyoA	Ubiquinol oxidase subunit 2
cyoB	Ubiquinol oxidase subunit 1
cyoD	Cytochrome o ubiquinol oxidase protein cyoD
nuoF	NADH-quinone oxidoreductase subunit F
nuoN	NADH-quinone oxidoreductase subunit N
ppc	Phosphoenolpyruvate carboxylase
sdhC	Succinate dehydrogenase cytochrome b556 subunit
sdhD	Succinate dehydrogenase hydrophobic membrane anchor subunit
ubiA	4-hydroxybenzoate octaprenyltransferase
KEGG Oxidative phosphorylation	
cyoA	Ubiquinol oxidase subunit 2
cyoB	Ubiquinol oxidase subunit 1
cyoD	Cytochrome o ubiquinol oxidase protein cyoD
nuoF	NADH-quinone oxidoreductase subunit F
nuoN	NADH-quinone oxidoreductase subunit N
sdhC	Succinate dehydrogenase cytochrome b556 subunit
sdhD	Succinate dehydrogenase hydrophobic membrane anchor subunit
KEGG TCA Cycle	
acnA	Aconitate hydratase 1
acnB	Aconitate hydratase 2
sdhC	Succinate dehydrogenase cytochrome b556 subunit
sdhD	Succinate dehydrogenase hydrophobic membrane anchor subunit

2.4 Discussion

2.4.1 A rapid shift between the expression of genes involved in aerobic and anaerobic energy metabolism is a key landmark of acid adaptation

Several lines of evidence that emerge from this work support the view that acid exposure induces a shift between aerobic and anaerobic metabolism and that this is a strong requirement for survival during acid shock. First of all, *E. coli* BW25113 cells under the conditions of our experiment express high levels of mRNA for enzymes involved in aerobic metabolism and lower levels of genes involved in anaerobic energy metabolism. A shift to low pH culture conditions induces a reversal of this pattern. The functional analysis revealed a trend in the modulation of the genes belonging to the anaerobic respiration and fermentation pathways (according to GO terms), which involves the down-regulation of the fumarate dependant respiration and the up-regulation of nitrate and formate respiratory enzymes (**Table 2.6**).

Table 2.6 Regulation of the genes involved in the anaerobic respiration pathway, according to GO terms.

The list includes genes down-regulated in response to acid exposure (left) and up-regulated (right).

DOWN-regulated		UP-regulated	
aceF	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	**aceE**	Pyruvate dehydrogenase E1 component
acnA	Aconitate hydratase 1	**fdnH**	Formate dehydrogenase, nitrate-inducible, iron-sulfur subunit
acnB	Aconitate hydratase 2	**fdnI**	Formate dehydrogenase, nitrate-inducible, cytochrome b556(fdn) subunit
dcuR	Transcriptional regulatory protein dcuR	**fdoI**	Formate dehydrogenase, cytochrome b556(fdo) subunit
dmsA	Anaerobic dimethyl sulfoxide reductase chain A	**glpA**	Anaerobic glycerol-3-phosphate dehydrogenase subunit A
dmsB	Anaerobic dimethyl sulfoxide reductase chain B	**glpB**	Anaerobic glycerol-3-phosphate dehydrogenase subunit B
fdhE	Protein fdhE	**glpC**	Anaerobic glycerol-3-phosphate dehydrogenase subunit C
fdnG	Formate dehydrogenase, nitrate-inducible, major subunit	**hyaA**	Hydrogenase-1 small chain
fdoG	Formate dehydrogenase-O major subunit	**hyaB**	Hydrogenase-1 large chain
fnr	Fumarate and nitrate reduction regulatory protein	**hyaE**	Hydrogenase-1 operon protein hyaE
frdA	Fumarate reductase flavoprotein subunit	**hyaF**	Hydrogenase-1 operon protein hyaF
frdB	Fumarate reductase iron-sulfur subunit	**hycE**	Formate hydrogenlyase subunit 5
frdC	Fumarate reductase subunit C	**hycF**	Formate hydrogenlyase subunit 6
frdD	Fumarate reductase subunit D	**hyfD**	Hydrogenase-4 component D
fumC	Fumarate hydratase class II	**hyfE**	Hydrogenase-4 component E
glpE	Thiosulfate sulfurtransferase glpE	**hyfF**	Hydrogenase-4 component F
glpR	Glycerol-3-phosphate regulon repressor	**hyfG**	Hydrogenase-4 component G
gltA	Citrate synthase	**hyfR**	Hydrogenase-4 transcriptional activator
hybB	Probable Ni/Fe-hydrogenase 2 b-type cytochrome subunit	**menA**	1,4-dihydroxy-2-naphthoate octaprenyltransferase
hybC	Hydrogenase-2 large chain	**napD**	Protein napD
hybD	Hydrogenase 2 maturation protease	**napF**	Ferredoxin-type protein napF
hybF	Probable hydrogenase nickel incorporation protein hybF	**narG**	Respiratory nitrate reductase 1 alpha chain
hybO	Hydrogenase-2 small chain	**narH**	Respiratory nitrate reductase 1 beta chain
hypD	Hydrogenase isoenzymes formation protein hypD	**narI**	Respiratory nitrate reductase 1 gamma chain
hypE	Hydrogenase isoenzymes formation protein hypE	**narJ**	Respiratory nitrate reductase 1 delta chain
hypF	Carbamoyltransferase hypF	**narL**	Nitrate/nitrite response regulator protein narL
icd	Isocitrate dehydrogenase [NADP]	**narX**	Nitrate/nitrite sensor protein narX
lldD	L-lactate dehydrogenase [cytochrome]	**narY**	Respiratory nitrate reductase 2 beta chain
lpd	Dihydrolipoyl dehydrogenase	**narZ**	Respiratory nitrate reductase 2 alpha chain
mdh	Malate dehydrogenase	**ndh**	NADH dehydrogenase
menC	o-succinylbenzoate synthase	**nikE**	Nickel import ATP-binding protein nikE
mltD	Membrane-bound lytic murein transglycosylase D	**nirB**	Nitrite reductase [NAD(P)H] large subunit
napC	Cytochrome c-type protein napC	**nrfA**	Cytochrome c-552
narW	Respiratory nitrate reductase 2 delta chain	**nrfE**	Cytochrome c-type biogenesis protein nrfE
nrfD	Protein nrfD	**pflC**	Pyruvate formate-lyase 2-activating enzyme
nuoB	NADH-quinone oxidoreductase subunit B	**pflD**	Formate acetyltransferase 2
nuoC	NADH-quinone oxidoreductase subunit C/D	**torA**	Trimethylamine-N-oxide reductase 1
nuoE	NADH-quinone oxidoreductase subunit E	**torC**	Cytochrome c-type protein torC
nuoF	NADH-quinone oxidoreductase subunit F	**torT**	Periplasmic protein torT

Enterobacteria are predominantly fermentative anaerobes that generate acids. In an anaerobic environment *E. coli* can utilize mixed acid fermentation, and the adaptation to readily decrease external pH, for example by inducing the synthesis of formate hydrogen lyase pathway. The metabolic switches seem to imply that acid triggers a response from aerobic respiration to adaptation to oxygen-limited growth (Hayes et al., 2006). This suggests that *E. coli* BW25113 has evolved the ability to prepare for anaerobic lifestyle by a pre-emptive induction of anaerobic respiration and fermentation. The comparison between the wild type and mutant strains (with intermediate and severe phenotypes) revealed that failure to invert the expression of aerobic and anaerobic metabolism genes correlates with a strongly acid sensitive phenotype. Moreover, many of the strains which show a strong phenotype are mutated in genes involved in energy metabolism (**Figure 2.5**). This indicates that under the conditions described here, effective proton scavenging alone may not be sufficient for survival unless the expression of bioenergetics genes is also modulated during adaptation to low pH conditions. I have seen that the genes transcriptionally controlled by the pH change are essential for the entire process previously described. This implies that the gene products are part of an integrated adaptive process in which transcription regulation depends on the completion of the process itself. There are currently few insights into the mechanism underlying this response. Hypotheses can be made if an *fnr* mutant could possibly be defective in adaptation to low pH. Although FNR is an iron-sulphur protein that is inactivated by oxygen (George et al., 1998, Marteyn et al., 2010) in cultures of moderately high density such as those used in the current experiments, FNR is partially active despite vigorous aeration. It would also be interesting to see whether FNR is critical for adaptation to acid of cultures at a much lower bacterial density.

2.4.2 The relevance of the osmotic stress response to acid adaptation

A potential connection between the activation of anaerobic metabolism (an important feature of acid adaptation in this strain) and osmotic stress has been previously described (Ni Bhriain et al., 1989), where it was linked to changes in DNA topology. Many genes involved in the response to osmotic shock are also modulated in response to acid in our experiments (**Figures 2.6 and 2.8**). Most of the channels involved in the transport of osmoprotectants were down-regulated at pH 5.5. However, the H+/proline symporter (Culham et al., 1993), the choline transporter (Lamark et al., 1991) and two K+ channels (Bossemeyer et al., 1989a, Bossemeyer et al., 1989b) genes were up-regulated after acid exposure, a response that makes adaptive sense in light of the finding that K+ and proline have a beneficial effect on pH homoeostasis in *E. coli* (Kitko et al.). It may also be the case that the increase of amino acids into the cytoplasm (data will be shown further in the MG1655 strain), as part of the acid adaptation response, could drive the cells to a weak condition of hypo-osmolarity. Consistent with this hypothesis, the mechano-sensor channels, MscS (Akitake et al., 2005) and MscL (Sukharev et al., 1993), with the acquaporine AqpZ (Borgnia et al., 1999) and the K+ mechanosensor (Booth et al., 1985) were up-regulated in the first 15 min after acid exposure. Substantial changes are known to take place in the balance of ions across the membrane on acid shock in *E. coli*, leading to reversal of membrane polarity.

2.5 Conclusion

The aim of this chapter was the description of the general response of *E. coli* BW25113 strain during the process of acid adaptation. As seen in previous works, environmental perturbations can drastically affect the metabolic pathways, therefore modulating them in

order to promptly adapt to the new conditions. My data clearly showed a change in most of the aerobic mechanisms at pH 5.5, confirmed by the gene deletion of important regulators and enzymes. These data are the basis for the next step, the identification of potential regulators of acid adaptation through a Systems Biology approach.

Chapter 3: A Network inference approach identifies new regulators of acid response in *E. coli* K-12 BW25113

3.1 Introduction

Until recently, researchers have assumed that the ability of *E. coli* to survive strong acid conditions is dependent on the activation of four ARs (Foster, 2004). The results I reported in Chapter 2 show that, unexpectedly a shift in the pattern of expression of genes coding for the energy metabolism enzymes is the most predictive signature of survival. The control of the energy metabolism could be an important feature of acid resistance. The hydrolysis of the ATP results in proton production while the reduction of electron acceptors leads to the consumption of protons (Jones et al., 1980). If this hypothesis were correct, it would be important to identify the regulators of the metabolic switch and address whether the regulation of the ARs and metabolic enzymes is integrated in the same regulatory network.

I addressed these questions by using a well-validated network inference approach. The networks I identified led to the hypothesis that the two-component system regulator OmpR may be the key regulator of the complex transcriptional program involved in acid adaptation. Experimental validation of this model, based on the analysis of a Δ*ompR* strain, supported this hypothesis and showed that the deletion of this gene induces a much stronger phenotype than any of the genes involved in the gad system. These results

therefore support the hypothesis that the OmpR dependent control of energy metabolism may be a novel important acid response system.

3.2 Materials and Methods

3.2.1 Bacterial strains

All the experiments described here were based on *E. coli* K-12 BW25113, which is directly derived from BD792, itself a two-step descendent of the *E. coli* K-12 ancestral strain (Hayashi et al., 2006, Bachmann, 1990). All additional mutant strains analyzed in this study originated from the Keio collection (Baba et al., 2006) and were checked by PCR to verify the presence of the deletion before being used. For this purpose a combination of locus and kanamycin specific primers were used as described in the original publications (Datsenko and Wanner, 2000).

The green fluorescent protein constructs plasmids used for the flow cytometry assay for the verification of the *gadW/atp* interaction (**Figure 3.1**) were extracted by the promoter library developed by Uri Alon's laboratory (Zaslaver et al., 2004) and inserted in the gene KO strains ΔgadW and ΔgadX from the Keio collection.

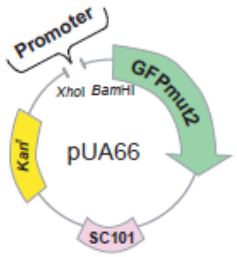


Figure 3.1 Plasmid reporter with gfp insertion (Zaslaver et al., 2004).

3.2.2 Culture conditions

The conditions followed for the experiments were the same as described in chapter 2, section materials and methods.

3.2.3 Phenotypic analysis of Δ*ompR* strain using flow cytometry

The phenotypic analysis of the mutant *ompR* was performed as described in chapter 2, section materials and methods.

3.2.4 pZCompR Plasmid construction

The plasmid construction was performed by Matthew D. Johnson, from Peter Lund laboratory. To generate the complementation plasmid pZCompR the complete *ompR* gene and the native promoter were amplified by PCR from the BW25113 chromosome using primers ompR-348F (GGTTGCTCGAGCGCCCAGACTTGCGGCCCAGG) and ompR+720R (GGTTGGGATCCTCATGCTTTAGAGCCGTCCGG) that introduce unique XhoI-BamHI restriction sites. The fragment was introduced into the multiple cloning site of the low copy number plasmid, pZC320, as described in the original protocol (Shi and Biek, 1995).

3.2.4 Green Fluorescent Protein detection with flow cytometry

Plasmids containing the reporter construct (Figure 3.1) for the genes atpI and gadB were extracted using the MiniPREP® Quiagen Kit. Δ*gadX* and Δ*gadW* were then transformed with the isolated plasmid using the $CaCl_2$ method. A time course experiment was performed at pH 5.5, in the conditions previously described, in order to keep constant optical densities. Samples were collected every 30 minutes from exposure, for a total of 2 hours and 30 minutes experiment. 0.5 ml cells were stabilised with formaldehyde (final concentration 1%). Samples were left on ice for 20 minutes and then stored at -20°C for a

month maximum. 0.5 ml of the stored samples were then centrifuged (4°C and 5000 rpm) for 10 min and washed in PBS for flow cytometry analysis.

For the data analysis, the mean of the fluorescence distributions was considered in three biological replicates for each construct and a T-TEST was performed for significance.

3.2.5 Expression profiling by microarray of Δ*ompR* strain

Expression profiling of the *ΔompR* was performed as described in chapter 2, section Materials and Methods, paragraph 2.2.4 Expression profiling by microarray.

3.2.6 Quantitative PCR

To validate gene expression profiles results, *E. coli* cells were grown and stabilised as for the microarray samples. RNA extraction was performed as described in the expression profiling section. 40 ng of cDNA were analysed with SYBR-Green method, after reverse transcription of the RNA with 2000 units SuperScript II-Invitrogen kit, according to the manufacturer's instructions. The primers were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and verified for specificity with Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The sequences for the primers were as follows: csrA (left primer CGAGTTGGTGAGACCCTCA, right primer AGCCTGGATACGCTGGTAGA); ompR (left primer CGTCGCTAATGCAGAACAGA, right primer GGTCCACTTCTTCCCCTTTC). Invitrogen primers (25 nmol) were used at a final concentration of 10 μM. SYBR-Green mix from ABGene, with ROX as passive reference, was used in a final volume of 10 μl. The analysis was performed with a 7900HT Fast Real-Time PCR System. I have obtained the following values of slope and r^2 for the

efficiency and the accuracy of the measurements: csrA slope = -3.1 cycles/logdecade, r^2 = 0.99; ompR slope = -3.01 cycles/logdecade, r^2 = 0.99. 40 cycles were done with an annealing temperature of 60ºC. Three technical and biological replicates were analysed and the gene expression of ompR was considered relative to the gene expression of csrA, the expression of which does not change under acid conditions (Burton et al., 2010). The three biological replicates were then analysed with the SDS software (Applied Biosystems) and a t-Test was performed for significance.

3.2.7 Data Analysis

Microarray Data processing, data exploration and identification of differentially expressed genes were performed as described in chapter 2, section materials and methods. Flow cytometry analysis of gfp constructs: the mean of the distribution of gfp fluorescence was considered for each replicate. Statistical analysis was performed on the averages of replicates and the standard deviation was considered for reproducibility.

Network Inference: Gene networks were inferred using the software application ARACNE (Margolin et al., 2006a, Margolin et al., 2006b) using the expression matrix representing the transcriptional state of 27 mutant strains and 3 replicates of the wild type strain profiled at pH 7 and pH 5.5 (for a total of 60 arrays) as an input. ARACNE is an algorithm used for the identification of transcriptional interactions between gene products, by identifying statistical interactions based on mutual information (Margolin et al., 2006a). Significant interactions were defined by a p-value threshold of $p<10\text{-}7$ (corresponding to one over 2000 false positive connections). In order to eliminate non-direct interactions I used the inequality principle as implemented in ARACNE with a DPI of 0.1, which will

remove the indirect connection without interfering with the path of the information flow (Margolin et al., 2006b). The resulting networks were visualised using the Cytoscape software application (Shannon et al., 2003).

ARACNE uses a probabilistic measure of dependency (mutual information) to identify potential targets of a given transcription factor. It relies entirely on the analysis of a compendium of expression profiles representing different perturbations of a cell's homeostatic state. Genes are inferred to be transcriptionally coupled when there is a statistically significant correlation between their levels of expression, across all samples in the compendium dataset and after potential indirect connections are removed using the inequality principle criterion (Margolin et al., 2006a). In this analysis I have considered a dataset of 60 experiments representing expression profiles of genes in mutants and wild type strains at pH 7 and pH 5.5, described in the previous chapter. This had the advantage that it represented a collection of mutant strains relevant for acid adaptation while being sufficiently large to allow reliable inference (Daub et al., 2004). In the visualised network, the distance between two genes is a direct measure of the mutual information, therefore the higher is the value of mutual information, the shorter is the distance between two genes. ARACNE cannot detect the direction of a potential interaction; a list of transcription factors was therefore considered in the algorithm.

Promoter analysis: To identify potential binding sites for OmpR, I have used the online tool Virtual Footprint (http://prodoric.tu-bs.de/vfp/) using the Prodoric library and performing the Regulon and the Promoter analysis (Munch et al., 2005). The PWM (Position Weight Matrix) considered were for the OmpC-box and OmpF-box of *E. coli* strain K-12.

3.3 Results

3.3.1 Network inference analysis reveals a new regulator of the global transcriptional response to acid stress

In the previous chapter I reported the intriguing discovery that resistance to mild acidic conditions is also characterised by an anaerobic switch in *E. coli* BW25113.

In order to assess whether this transcriptional signature is truly linked to survival in acid conditions I first set to identify the putative regulators using a network inference approach.

By using the well-validated network inference platform ARACNE I could develop networks representing the neighbourhood of TCSs regulators, which represent the higher level of regulation in *E. coli* (**Figure 3.2**). In order to identify the most likely regulator of acid response I ranked all the *E. coli* TCS regulators on the basis of the number of connections with genes belonging to energy metabolism pathways (**Table 3.1**).

Table 3.1 Potential two-component systems regulators targets as defined by the ARACNE analysis.

Regulators were ranked on the basis of how many connections were found to functions that were modulated during the acid adaptation time course (in red). P-values of the DAVID functional annotation analysis are shown

Gene	Two-component system	Targets based on MI	Function enriched	p-value
ompR	Osmoregulatory Two-component system OmpR/EnvZ	33	GO_BP Intracellular pH elevation (3 genes)	4.50E -05
			GO_CC Cytoplasm (7 genes)	6.80E -03
			KEGG ABC transporters (3 genes)	4.20E -02
			KEGG Gluthatione metabolism (2 genes)	3.90E -02
			KEGG Glycolysis/Gluconeogenesis(4 genes)	3.00E -03
			KEGG Pyrimidine metabolism (5 genes)	3.60E -03
			KEGG Pyruvate metabolism (2 genes)	2.40E -03
			KEGG Two-component system (3 genes)	3.10E -03
dcuR	Two-component system DcuR/DcuS, regulating anaerobic fumarate respiratory system	27	GO_BP Ciliary or flagellar motility (3 genes)	1.10E -03
			KEGG Two-component system (7 genes)	3.10E -05
			KEGG Methane metabolism (3 genes)	9.28E -05
			KEGG Pyruvate metabolism (3 genes)	3.46E -03
baeR	Two-component regulatory system BaeS/BaeR	43	GO_BP Transport (10 genes)	1.25E -06
			KEGG ABC Transporters (11 genes)	3.21E -06
			KEGG Galactose metabolism (5 genes)	1.39E -05
			KEGG Purine metabolism (3 genes)	2.06E -03
narL	NarX/NarL Two-component system, nitrate/nitrite dependent	71	GO_MF Metal ion binding (4 genes)	2.88E -03
			KEGG Glycolysis/Gluconeogenesis (3 genes)	4.63E -04
phoP	Two-component system PhoP/PhoQ, magnesium	48	GO_BP Oxidation reduction (3 genes)	1.63E -03
			KEGG Oxydative phosphorylation (3 genes)	1.63E -03
kdpE	KdpD/KdpE involved in the regulation of the kdp operon	31	GO_BP Nickel ion transport (2 genes)	5.78E -04
			KEGG TWO-component system (6 genes)	4.48E -04
rstA	Two-component regulatory system RstA/RstB	27	GO_BP Translation (4 genes)	5.36E -03
			KEGG Alanine, aspartate and glutamate metabolism (3 genes)	9.81E -04
			KEGG Ribosome (3 genes)	4.27E -05
uvrY	Two-component regulatory system UvrY/BarA	23	GO_BP Two-component signal transduction system (5 genes)	2.48E -05
			GO_BP Transcription (7 genes)	1.45E -04
			KEGG Two-component system (3 genes)	4.27E -05
narP	NarP/NarQ Two-component system, nitrate/nitrite dependent	50	GO_BP Pyrimidine nucleotide biosynthetic process (6 genes)	3.36E -10
			GO_BP Iron-sulfur cluster assembly (6 genes)	1.21E -07
			KEGG Pyrimidine metabolism (7 genes)	1.72E -06
			KEGG Alanine, aspartate and glutamate metabolism (5 genes)	1.57E -05
rcsB	RcsB/RcsC regulates the expression of genes involved in colanic acid capsule	36	GO_BP Transcription (9 genes)	1.08E -04
fimZ	involved in fimbrial expression	34	GO_BP Transcription (7 genes)	7.24E -04
torR	TorR/TorS responds to changes in the concentration of the TMAO	20	GO_BP Metabolic process (5 genes)	1.09E -02
rssB	facilitates and regulates degradation of sigma S	18	GO_MF Protein binding (8 genes)	7.77E -03
yjjR	Two-component system YfjR/YfhK	18	GO_BP Transcription (5 genes)	9.98E -03
uhpA	UhpA/UhpB two-component system	13	GO_CC Integral to membrane (8 genes)	4.33E -03

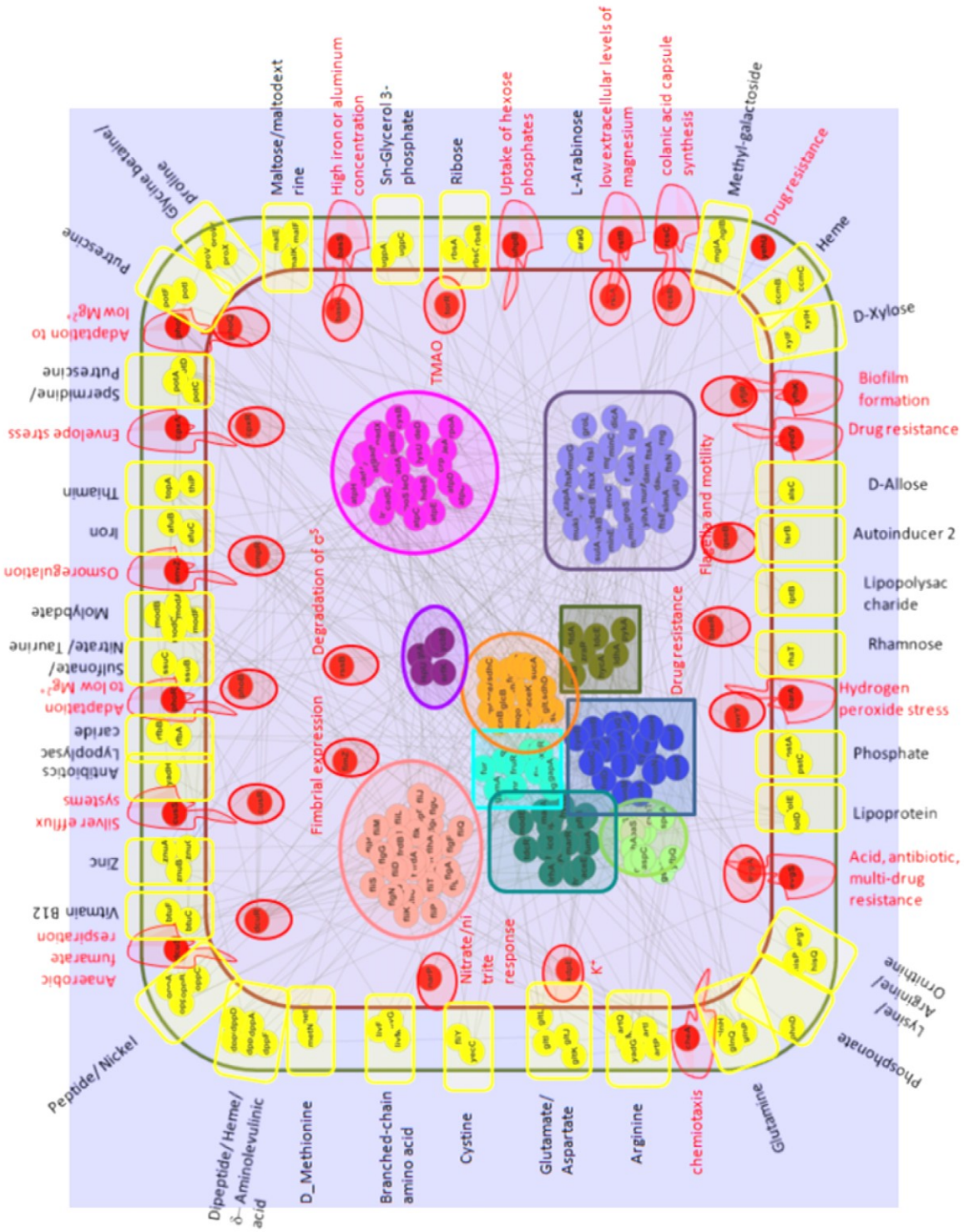


Figure 3.2 Two-component systems network from ARACNE, in a cell structure.

This is a schematic representation of the ARACNE network of the Two-component systems neighbourhood. In red are reported the two-component systems, in yellow the membrane proteins. The others groups are related to genes encoding amino acid metabolism, aerobic and anaerobic respiration pathways.

The highest-ranking gene was OmpR, the regulatory subunit of the osmoregulator two-component system EnvZ/OmpR. OmpR was connected to genes involved in aerobic energy metabolism (pyruvate metabolism and glycolysis), signal transduction, and transport, as well as some of the components of the GAD system (**Figure 3.3** and **Table 3.2**).

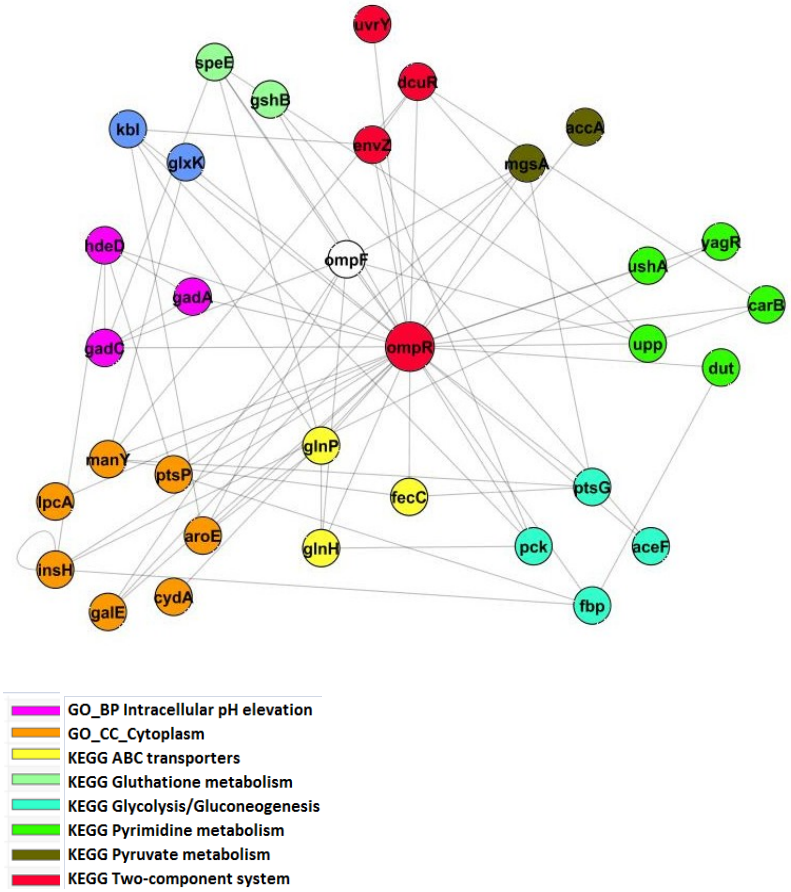


Figure 3.3 Network of regulatory interactions in the neighbourhood of OmpR, inferred using ARACNE.

Nodes represent genes and edges represent inferred connections ($p<10^{-7}$) between them. Genes are colour-coded on the basis of their function.

Table 3.2 SAM comparison performed considering Δ*ompR* at pH 7, 5.5 and the wild type.
The functions enriched by the genes differentially expressed between the wild type at pH 7 and the mutant at pH 7 (column 1), wild type pH 5.5 and mutant at pH 5.5 (column 2) and wild type at pH 7 and pH 5.5 (column 3) are represented. In the table the functions are also represented that for each analysis were found up and down-regulated.

	SAM 10% FDR (WILD TYPE pH 7 vs WILD TYPE pH 5.5)	SAM 10% FDR (WILD TYPE pH 7 vs OMPR pH 7)	SAM 10% FDR (WILD TYPE pH 5.5 vs OMPR pH 5.5)
UP_Regulation	GO_BP Transport	KEGG Fatty acid metabolism	GO_BP Response to stress
		KEGG Flagellar assembly	GO_BP Cell cycle
		KEGG ABC transporters	KEGG Pyruvate metabolism
DOWN-Regulation	GO_BP Translation	GO_BP Translation	GO_BP Transport
	KEGG Pyruvate metabolism	KEGG Ribosome	
	KEGG Ribosome	KEGG Pyrimidine metabolism	

NarP, the nitrate/nitrite response regulator, was connected to nucleotide and amino acid metabolism; the envelope stress response regulator BaeR was connected to genes involved in transport and metabolism. Most of the other two component system regulatory subunits were poorly connected to genes in the pathways I had previously identified. On the basis of these observations, I proposed the novel hypothesis that OmpR may be a key regulator of acid response in *E. coli* BW25113.

3.3.2 Model validation shows that OmpR is a key regulator of acid response

If my hypothesis is correct, I expect a mutant strain lacking *ompR* to display the following properties. First, those genes differentially regulated between Δo*mpR* and the wild type strain should significantly overlap with genes differentially regulated during acid adaptation in the wild type. Second, this overlap should be consistent with the regulation of *ompR* itself during acid adaptation in the wild type strain. For example, if *ompR* expression decreases on acid shock, then those genes that are normally repressed by OmpR

should increase after acid shock, and show significant overlap with genes that are up-regulated if *ompR* is deleted. Third, the Δ*ompR* mutant strain should be less able to initiate the normal transcriptional response to acid adaptation and, hence, should show a significant decrease in survival of acid shock relative to the wild-type strain. In order to test these predictions, I analysed the expression profile of a Δ*ompR* mutant strain at pH 7 and 5.5, and performed a phenotypic characterization of this strain by flow cytometry after a direct challenge at pH 2.5. The results of this analysis closely matched my predictions. Genes differentially regulated between wild type and the Δ*ompR* mutant strain showed significant overlaps with genes differentially expressed in the wild type strain during acid adaptation (**Figure 3.4**, panels **C** and **D**).

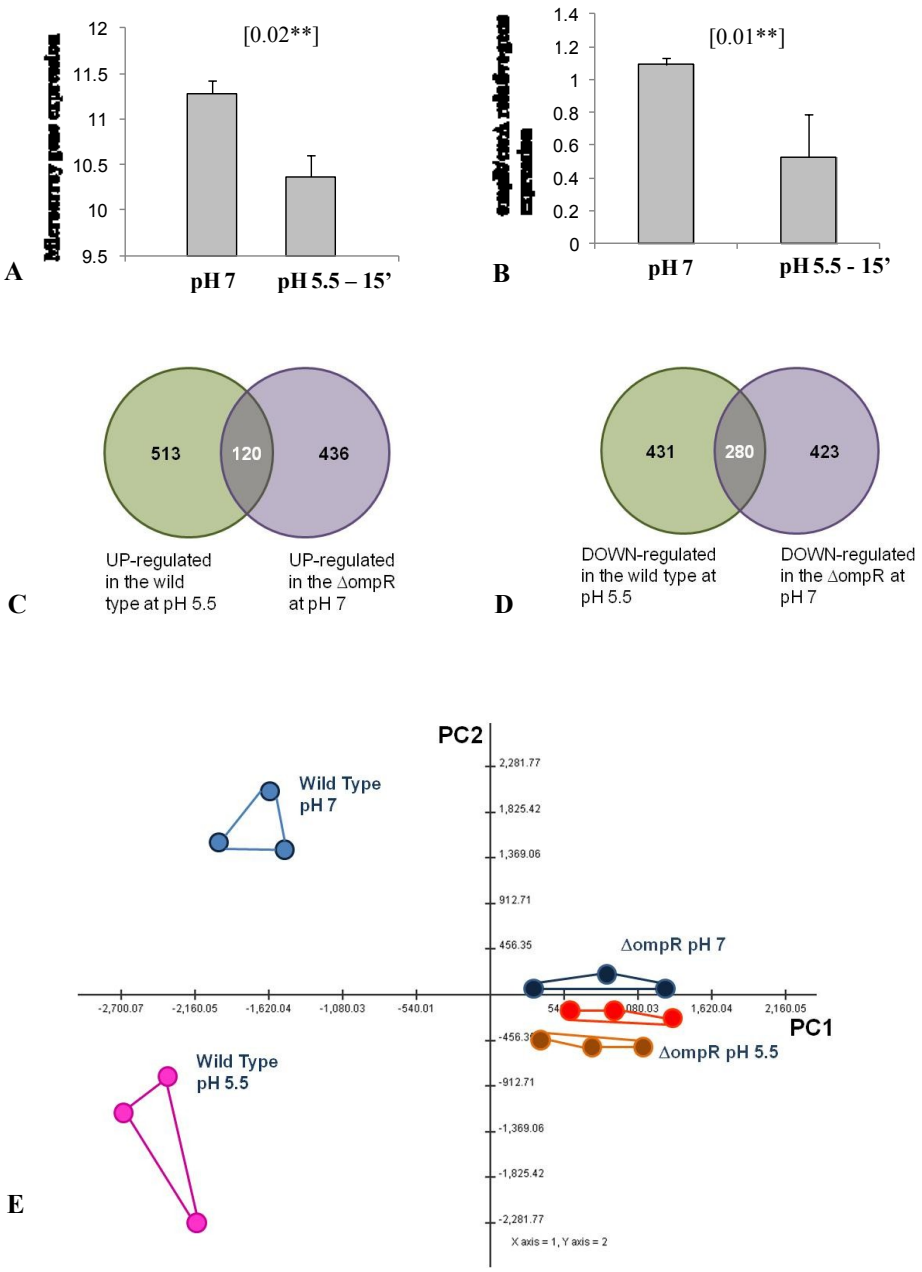


Figure 3.4 Figure 3.4: Comparison of transcriptional responses of ompR and wild-type to acid stress of pH 5.5.

Panels (A and B) show the expression levels of *ompR* at pH 7 and after 15 min exposure at pH 5.5, based on either microarray data (A) or qPCR data (B). Bars show standard deviations of three biological replicates. *ompR* is significantly down-regulated at pH 5.5 in both data sets (P-value 0.02 and 0.01, respectively, for microarrays and qPCR data, obtained with t-test). (C and D) show the extent of overlap between genes which are differentially up-regulated (C) or down-regulated (D) in the wild-type at pH 5.5, relative to expression at pH 7, and genes which are over-expressed (C) or under-expressed (D) in ompR at pH 7, relative to the wild-type at pH 7. (E) PCA plot of transcriptome changes in wild-type and the ompR mutant, analysed at pH 7 and 5.5. Blue dots, wild-type pH 7, pink dots: wild-type pH 5.5, black dots ompR at pH 7, red and orange dots ompR at pH 5.5 after 30 s and 15 min of exposure, respectively.

More specifically, 120 (23%) of the up-regulated genes in the wild type during acid adaptation were also expressed more highly in the ΔompR mutant than in the wild-type at pH7, while 280 genes (71%) that were down-regulated in the wild type during acid adaptation were expressed more weakly in the ΔompR mutant than the wild type at pH 7. The direction of the overlap was consistent with the observed down-regulation of *ompR* in response to acid exposure (**Figure 3.4**, panels **A** and **B**). qPCR showed the down-regulation of ompR at pH 5.5 with values of slope and r^2 for the efficiency and the accuracy of the measurements: csrA slope = -3.1 cycles/logdecade, $r^2 = 0.99$; ompR slope = -3.01 cycles/logdecade, $r^2 = 0.99$. Even more strikingly, there was no transcriptional response detectable to acid exposure in the Δ*ompR* mutant either using a univariate statistical analysis approach (no differentially expressed genes were identified up to a FDR<50% threshold) or using a multivariate exploratory analysis approach (**Figure 3.4**, panel **E**).

From my analysis the role of OmpR appeared to be crucial in order to allow the cells to adapt to mild acid conditions. The next step was to understand whether this mutation could also phenotypically affect the cells at pH 2.5

3.3.3 Comparison between effects of acid exposure in different mutant strains confirms the importance of OmpR

The phenotypic experiments showed that mutant cells showed increased resistance to acid shock if pre-adapted to mild acid conditions, suggesting a small residual ability to mount an effective response, although they were still significantly less resistant than the unadapted wild-type strain (**Figure 3.5**, panel A). Complementation of the Δ*ompR* mutant strain with a copy of the wild-type *ompR* gene expressed under its own promoter in the

low-copy-number plasmid pZC320 led to restoration of normal levels of acid resistance, whereas the vector alone had no effect (**Figure 3.5**, panel **B**).

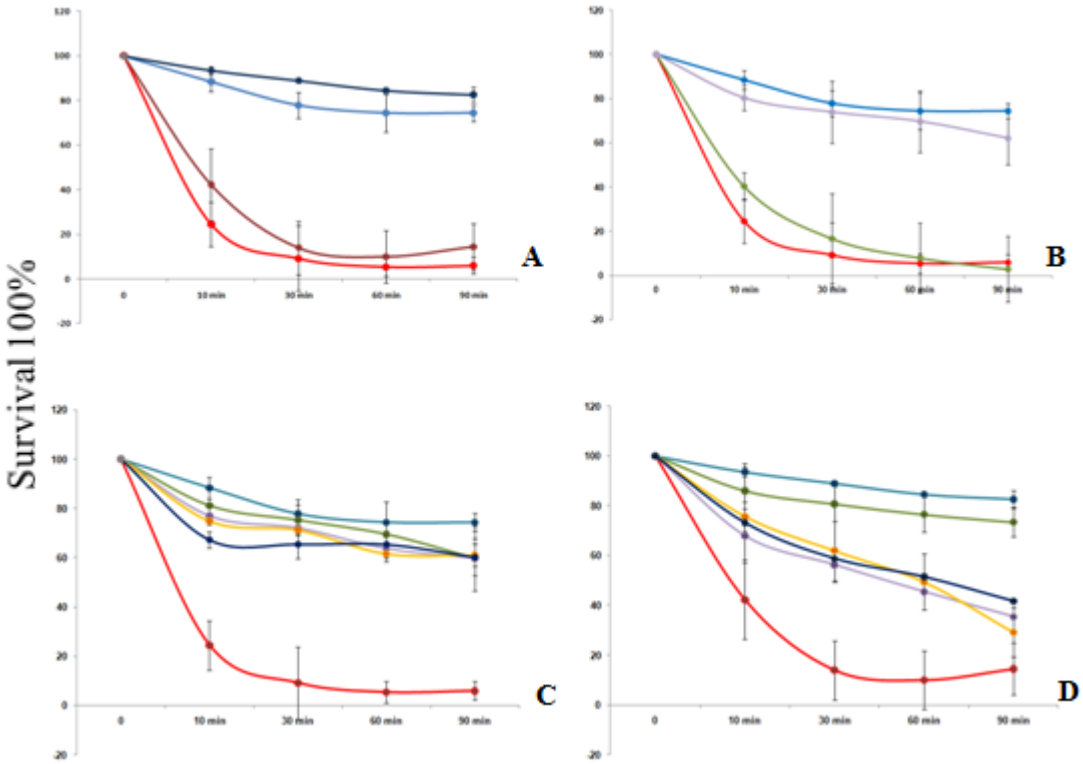


Figure 3.5: Comparison between phenotypes.

(A) Acid sensitive phenotype of the ΔompR strain. Survival of wild type and ΔompR strains in response to exposure to pH 2.5, either directly or after prior adaptation to pH 5.5, was measured by flow cytometry. Error bars show standard deviations of four independent biological replicates. The y axis shows % survival relative to viable cell numbers at t0. Wild type without adaptation, pale blue; wild-type with adaptation, dark blue; ΔompR without adaptation, red; ΔompR with adaptation, brown. (B) Complementation of the ompR strain restores acid resistance. The ΔompR mutant was complemented with the plasmid pZCompR, or as a control with the empty vector pZC320, and acid resistance measured at pH 2.5 without prior adaptation Error bars show standard deviation values of four biological replicates for WT and ompR and three biological replicates for the complemented strain. Wild type, pale blue; ΔompR without vector, red; ΔompR with empty vector, green, ompR with pZCompR mauve. (C, D) Comparison of the ompR acid resistance phenotype with those of other mutations in key acid response genes. Survival curves of the following mutants after exposure to pH 2.5, without (A) or with (B) prior adaptation at pH 5.5 for one hour: ΔompR (red), ΔgadE (yellow), ΔgadC (purple), ΔadiC (green) and ΔrpoS (dark blue) compared to the wild type (blue). Error bars show standard deviation values of four independent biological replicates for wild-type and ΔompR and three biological replicates for the other mutant strains.

Having demonstrated that OmpR truly regulates the shift in the expression of metabolic genes, I set to establish whether the process controlled by OmpR may be more important in survival to acid shock than the GAD system itself. Since I have shown that the expression of energy metabolism genes is the best predictor of survival our expectation was that a Δ*ompR* should have a stronger phenotype than mutants belonging to the AR systems.

I confirmed that the Δ*ompR* strain showed the strongest phenotype of all these mutants when cells were not adapted with a prior shock at pH 5.5 (**Figure 3.5**, panel **C**). After 10 minutes of exposure less than 30% of the Δ*ompR* cells were still alive, while in the other strains more than 70% of the cells were healthy at this stage. Strains carrying mutations in either of the two amino-acid antiporters showed a weak phenotype, in the culture conditions considered in the considered conditions. Different effects were seen when cells were induced at pH 5.5, before exposure to acid shock at pH 2.5 (**Figure 3.5**, panel **D**). All mutants in this case (apart from the *adiC* mutant) showed large reductions in viability compared to the wild-type strains; however, the Δ*ompR* strain still had the most severe phenotype.

3.3.4 Inferring a connection between different ARs: A cross talk between the GAD and F_OF_1 ATPase systems?

Having successfully applied ARACNE to identifying regulatory networks controlling the metabolic switch underlying acid response I set to explore whether the approach could be used to test the hypothesis that the classical ARs represent an integrated system. We reasoned that if this is the case we should be able to identify higher-level regulators controlling their differential usage.

We therefore applied ARACNE to study the neighbourhood of 36 genes belonging to the ARs (**Figure 3.6**). Among the strongest connections there were several genes representing GAD System components (e.g. *gadA*, *gadB*, *gadC*, *gadE*, *gadX*, *gadW*, *hdeA*, *hdeB*, *hdeD)* (**Figure 3.6**).

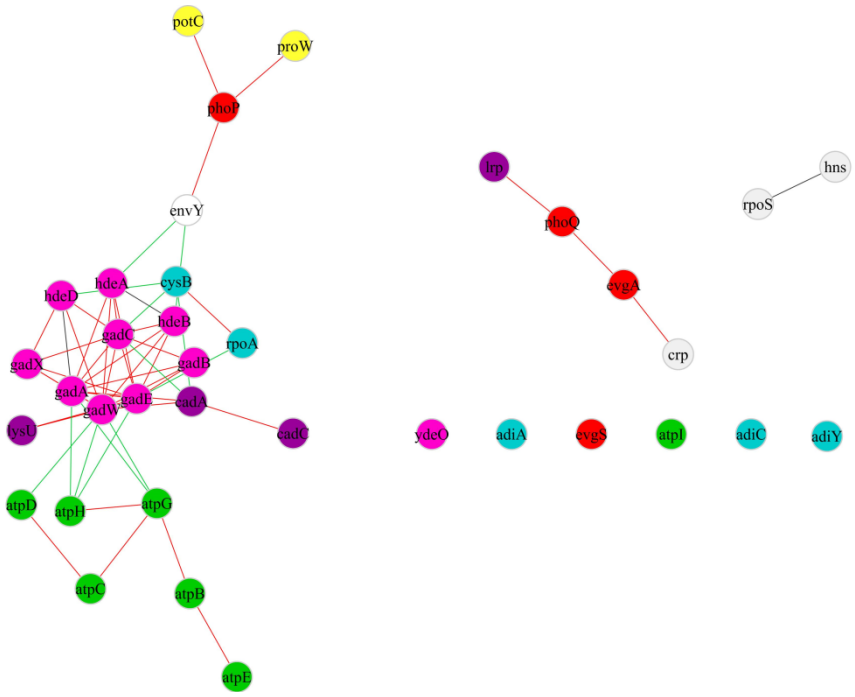


Figure 3.6 Gene network ARACNE of the genes belonging to the ARs.

After selecting the genes belonging to ARs, 36 genes were found in the network. The colour of the nodes connections is indicative of the correlation found between the genes: red for a positive correlation and green for a negative correlation. ARACNE was able to find most of the connections between ARs, which are already known in literature. The genes belonging to the GAD system (pink nodes) were highly connected between them; the same behaviour was found for the *atp* genes (green nodes). The gene gadW was also connected to the genes of the FoF_1 ATPase. Interestingly, another connection found by the software involves the TCSs PhoP/PhoQ and EvgA/EvgS (red nodes) which are known to interact through the SafA protein (Eguchi, 2007). In the network were also considered the genes belonging to the AR3 (light blue nodes) and AR4 (purple nodes). Master regulators are indicated in gray.

The network identified several known regulatory interactions, such as the connection between the regulators GadXW with the decarboxylase encoding genes gadA and gadB (Tramonti et al., 2002). More specifically, *gadW* was positively correlated with the general regulator of the GAD system, *gadE* (**Figure 3.7**, panel A). The connection between the Two-component systems PhoP/PhoQ and EvgA/EvgS (Eguchi et al., 2007) was also represented in the model.

The network also revealed a potentially interesting, previously unreported link with ARs. More specifically, I discovered that the expression of the *gadW* gene is negatively correlated to the expression of genes encoding for components of the F_OF_1 ATPase (*atpD*, *atpG*, *atpH (***Figure 3.7**, panels B, C and D*)*.

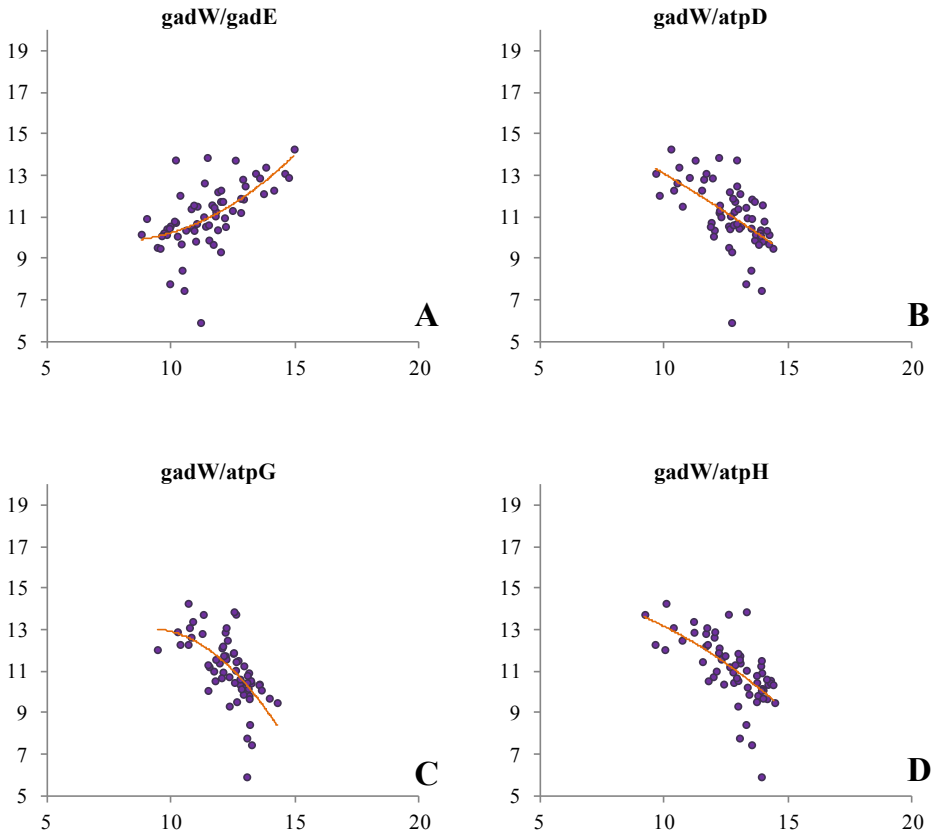


Figure 3.7 Gene correlation between the *gadW* and *gadE*, *atpD*, *atpG* and *atpH*.

The four panels represent gene expression correlations plots between gadW and the gadE (A), atpC (B), atpD (C) and atpH (D). While the correlation between gadW and gadE was clearly positive, the genes of the ATPase showed a negative trend of correlation.

The F_OF_1 ATPase is not known to be involved in the mechanism in mild acid conditions (Richard and Foster, 2004). Instead it plays a role at very low pH (< 2.5) where it works extruding protons at expense of ATP (Richard and Foster, 2004). The mechanism underlying the reversing action of the pump at a pH 2.5 is still unknown. However, it is important to consider that changes in enzymatic activities are dependent on thermodynamic conditions, which could determine reversibility in their action (Rottenberg, 1973). The negative correlation between GadW and the atp genes could therefore represent

a regulatory mechanism, mediated by the araC-like regulator GadW and aiming at switching off the expression of the F_OF_1 ATPase system when it is not needed.

I reasoned that it is possible that the negative correlation observed between GadW and the atp genes may be the result of a direct regulatory role of the gadW gene. If this hypothesis is correct, the expression of the *atp* genes, coding for FoF1, in a *gadW* mutant background should be higher than in the wild type strain.

I tested this hypothesis by using a gfp reporter system to monitor the expression of specific ATPase genes in different mutant strains. More specifically, I have used a gfp promoter reporter construct for the controller of the i-subunit of the ATPase (*atpI),* as a representative of the F0F1ATPase system (Matthies et al., 2011). I have then considered the gene *gadB*, coding for one subunit of the glutamate decarboxylase, as control reporter for the classical regulatory action of GadW/X (Tramonti et al., 2006, Ma et al., 2002). The plasmids carrying the two considered promoters were inserted in Δ*gadW* and Δ*gadX* strains and the experiments were performed in constant growth conditions at pH 5.5 for 2 hours and 30 minutes. I have considered 6 time points at pH 5.5 (one time point every 30 minutes) and one at pH 7. The fluorescence was then measured with flow cytometry. The results of the experiments are shown in **Figure 3.8**.

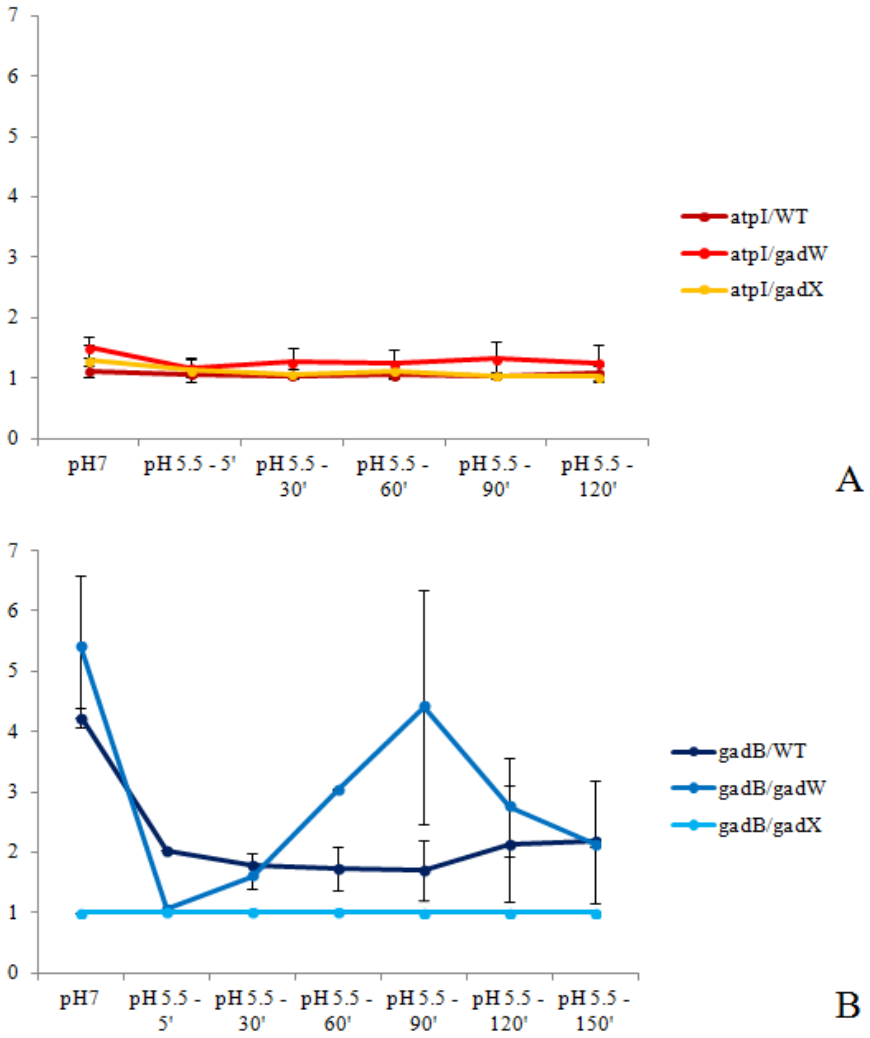


Figure 3.8 Acid adaptation assays performed with flow cytometry on gfp promoter constructs for *atpI* and *gadB*.

The gfp expression was analysed in both Δ*gadX* and Δ*gadW*. Panel A) The atpI gfp reporter did not show any induction in *gadW* (red) and *gadX* (orange) mutant when compared to the wild type (dark red). Panel B) The gadB promoter construct was highly induced in a Δ*gadW* background (blue), while in a Δ*gadX* (light blue) background when compared to the wild type (dark blue). The bars in both panels represent the standard deviations values of the three biological replicates for each time point. The *gadB* gfp reporter was highly induced in Δ*gadW*, but the induction was not observed in Δ*gadX*. The gfp expression of *atpI* promoter was not induced in the three considered conditions, therefore in the WT and the two mutant strains.

The inhibition exerted by GadW on *gadB* was confirmed by the results obtained during acid adaptation. In absence of GadW, the gfp expression of the *gadB* promoter construct was significantly higher when compared to WT (**Table 3.3**), although not in all the time points. The possible explanation for these results could be a problem due to the considered experimental conditions: I have encountered a big variability related to the *gadB*-gfp construct. In a Δ*gadX* background, the *gadB*-gfp expression was not induced, confirming the activating role of this regulator on the GAD genes (**Table 3.3**).

Table 3.3: T-test for detection of significant inductions of the gfp reporters *atpI* and *gadB*.

The T-test was performed for each time point for both the constructs. wt vs. Δ*gadW* is the comparison between the gfp expression found in the Wild Type and the Δ*gadW* condition. wt vs. Δ*gadX* is the comparison between the gfp expression found in the Wild Type and the Δ*gadX* condition. Δ*gadW* vs. Δ*gadX* is the comparison between the gfp expression found in the Δ*gadW* and the Δ*gadX* conditions. The *atpI* construct did not show any significant induction. The *gadB* construct was significantly induced in Δ*gadW* when compared to the wild type and to Δ*gadX*. Significant values are indicated by yellow cells and red font. p-values were corrected for multiple comparisons by using the Benjamini and Hochberg correction (Benjamini and Hochberg, 1995).

	atpI gfp reporter						*gadB* gfp reporter					
	wt vs ΔgadW		wt vs ΔgadX		ΔgadW vs ΔgadX		wt vs ΔgadW		wt vs ΔgadX		ΔgadW vs ΔgadX	
	p-value	Benjamini	p-value	Benjamini	p-value	Benjamini	p-value	Benjamini	p-value	Benjamini	p-value	Benjamini
pH 7	1.22E-01	5.02E-01	2.70E-01	7.29E-01	4.90E-01	6.86E-01	**9.00E-03**	**3.15E-02**	3.30E-01	3.30E-01	**2.00E-02**	**4.67E-02**
pH 5.5 - 5'	4.04E-01	5.02E-01	6.25E-01	7.29E-01	8.60E-01	8.60E-01	**4.65E-05**	**3.26E-04**	**1.08E-05**	**7.56E-05**	5.70E-02	7.98E-02
pH 5.5 - 30'	1.76E-01	5.02E-01	5.30E-01	7.29E-01	1.30E-01	6.86E-01	5.40E-01	6.30E-01	2.18E-02	7.63E-02	4.10E-02	7.18E-02
pH 5.5 - 60'	3.10E-01	5.02E-01	5.12E-01	7.29E-01	6.00E-01	7.00E-01	**2.70E-02**	**4.73E-02**	7.20E-02	1.68E-01	**3.25E-05**	**2.28E-04**
pH 5.5 - 90'	2.70E-01	5.02E-01	1.00E+00	1.00E+00	2.41E-01	6.86E-01	8.00E-02	1.12E-01	1.32E-01	2.11E-01	9.00E-02	1.05E-01
pH 5.5 - 120'	5.20E-01	5.20E-01	4.90E-01	7.29E-01	3.40E-01	6.86E-01	**1.70E-02**	**3.97E-02**	1.76E-01	2.11E-01	6.70E-01	6.70E-01
pH 5.5 - 150'	4.30E-01	5.02E-01	4.60E-01	7.29E-01	4.86E-01	6.86E-01	9.50E-01	9.50E-01	1.81E-01	2.11E-01	**9.40E-04**	**3.29E-03**

Moreover, no significant induction was observed for the atpI-gfp reporter construct in the three considered conditions (WT, Δ*gadW* and Δ*gadX*). Consequently I concluded that the negative correlation observed between the expression of GAD and the *atp* genes may not be the result of a direct regulatory role of GadW.

3.4 Discussion

3.4.2 OmpR: an important acid response regulator

Our study provides strong evidence for a role of OmpR as a regulator of the transcriptional response to acid adaptation in BW25113. Although OmpR has not been directly implicated hitherto in acid resistance in *E. coli* , it has been shown that an *ompR* UPEC mutant shows reduced survival in the mouse urinary tract, and that the growth defect seen in this mutant in high salt is enhanced at low pH (5.5); lethal pH was not tested in this experiment (Schwan, 2009). OmpR has been shown to regulate the stationary phase acid inducible response in *Salmonella Typhimurium* (Bang et al., 2000) potentially by counteracting H-NS-mediated repression (Bang et al., 2002), and an OmpR-like regulator (HP0166) has also been implicated in the acid response of *Helicobacter pylori* (Bury-Mone et al., 2004) suggesting a broader role for regulators of this type in acid stress responses. Gene expression in the *ompR* mutant partially mimics the response to acid of the parent strain (**Figure 3.5**, panels **A** and **B**) but mutant cells are unable to mount any response to acid, suggesting that modulation of OmpR is required for an effective response to acid. This result, coupled with the acid sensitivity of the *ompR* mutant, are consistent with a model where OmpR is required for the expression of cellular components, or for the establishment of a particular cellular state, which is needed for the cells to be able to respond to acid stress in a way that enhances their survival. In the absence of OmpR, this state no longer exists and so the *ompR* mutant fails to respond to acidification and shows enhanced acid sensitivity. Cells lacking this important regulator are even less able to survive in extreme acid conditions than those carrying mutations in other important genes

implicated in acid resistance. The key role of OmpR in regulating adaptation to low pH was unexpected and has not featured in an extensive literature on the EnvZ–OmpR two-component system. Key questions for future research to answer are whether EnvZ mediates the response and, if so, to what chemical signal EnvZ responds. In a previous study on *Shigella*, FNR, the dual transcriptional regulator of the switch between aerobic and anaerobic metabolism, was found to regulate the length of type III secretion system needles, required for secretion of the invasion plasmid antigen. Exposure to oxygen at the surface of the gastrointestinal mucosa inactivates FNR, reversing the block on invasion antigen secretion and hence priming the bacteria for the attack (George et al., 1998, Marteyn et al., 2010). Moreover, a deeper analysis of FNR targets revealed a down-regulation of the functions modulated by FNR, which I found important for my work (**Table 3.2**).

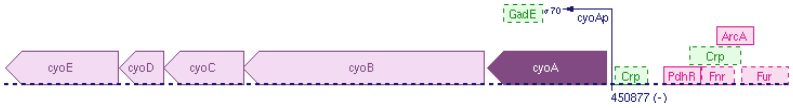
Whether OmpR exerts its function directly or by modulating the activity of other regulators remains an important question. I found that some of the genes downstream from OmpR (as defined by correlation and KO analysis) have OmpR-binding sites in their promoter regions (**Figure 3.9, A-B-C-D**), suggesting that at least in some cases OmpR may directly activate genes involved in acid response.

C-BOX

cyoA: cytochrome bo terminal oxidase subunit II

```
ACAGCTTCTTAAAATCAACCTGATATGTTTTGCAACATATGTGACCTGGCAGCCAAATCCAAGTAACAGG
                              TGAA-CAT-T
```

Start	End	Strand	PWM Score(s)	SEP Score	ATG-Distance	Location
450955	450964	+	12.18	-6.68	121	intergenic



argQ: tRNAargQ

```
AGTAATCCTCCCGGATGCACCATCTCTTACTTGATACGGCTTTAGTAGCG ... (130)...
TCGAATCCTCCCGGATGCACCATCTCTTACTTGATATGGCTTTAGTAGC
          TGAA-CAT-T
```

Start	End	Strand	PWM Score(s)	SEP Score	ATG-Distance	Location
2815943	2815952	+	12.04	-7.01	61	intergenic
2816217	2816226	+	12.04	-6.84	335	coding region

argZ: tRNAargZ

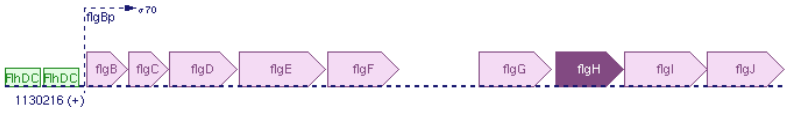
```
TTCGAATCCTCCCGGATGCACCATATTCTACGTACTTTCAGCGATGAAGG ... (124)...
          CCCGGATGCACCATCTCTTACTTGATACGGCTTTAGTAGCGGTATC
          TGAA-CAT-T
```

Start	End	Strand	PWM Score(s)	SEP Score	ATG-Distance	Location
2816217	2816226	+	12.04	-6.84	60	coding region
2816358	2816367	+	12.04	-6.96	201	coding region

flgH: flagellar L-ring protein

```
AACTGGTCAATATGATTCAGGTGCAACGCGCTTACGAAATCAACAGTAAA ... (50)...
CCACCGATCAGATGCTGCAAAAACTGACGCAACTCTAAGGCTTAACCGGT
          TGAA-CAT-T
```

Start	End	Strand	PWM Score(s)	SEP Score	ATG-Distance	Location
1134706	1134715	+	12.16	-7.2	72	coding region
1134646	1134655	+	12.55	-7.55	132	coding region



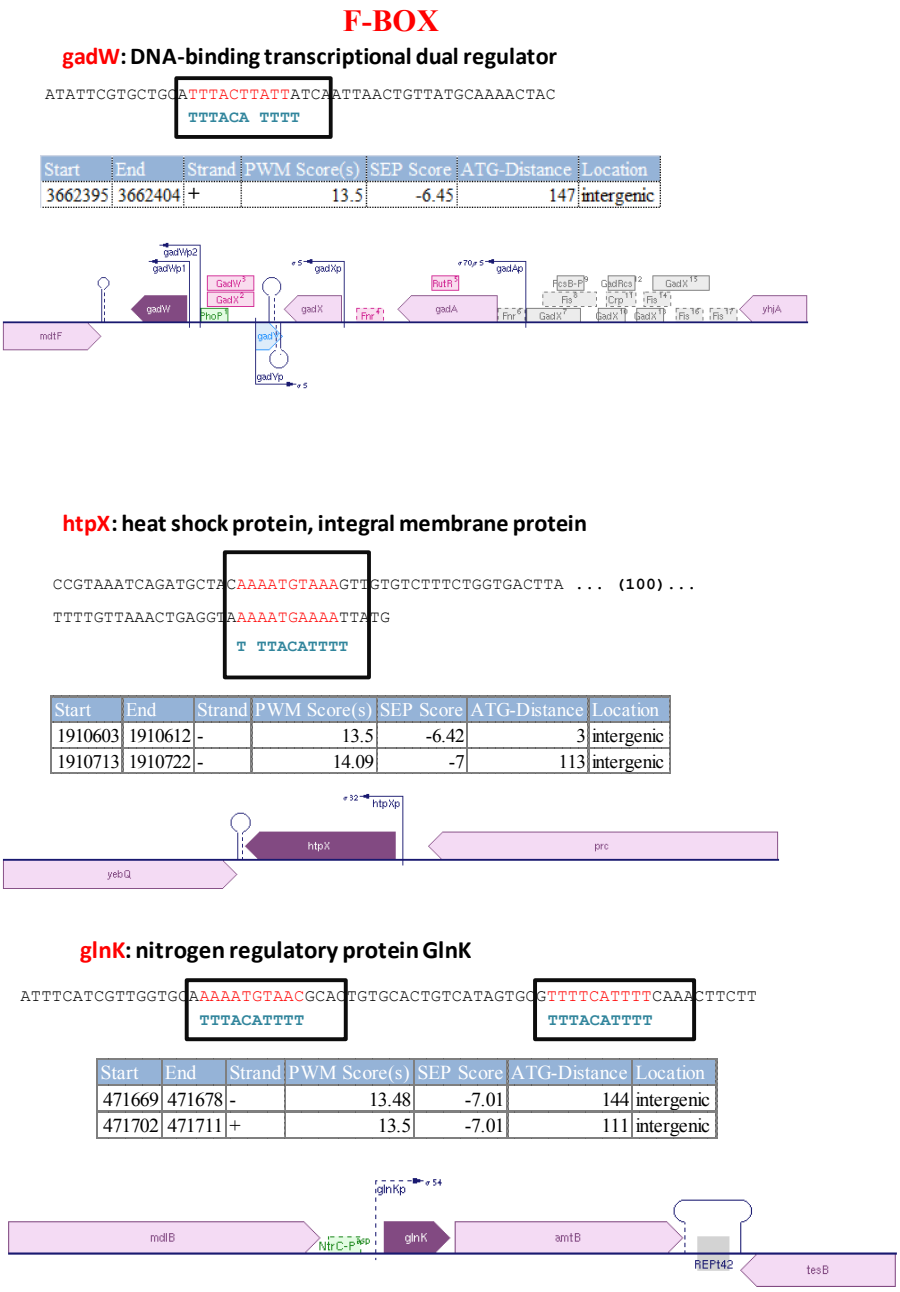


Figure 3.9 Binding site analysis of potential OmpR targets.

The potential binding sites of OmpR are shown, which were detected on the basis of the similarity with ompC-box and ompF-box. For each gene considered, the genome sequence, the statistical relevance of the similarity and the relative operon are shown. PWM scores and SEP scores were determined for the selection of the targets.

3.4.3 Role of OmpR in acid resistance and osmotic stress

The novel role I have found for OmpR suggests a link between osmotic and acid stress responses. A potential connection between the activation of anaerobic metabolism (an important feature of acid adaptation in this strain) and osmotic stress has been previously described (Ni Bhriain et al., 1989), where it was linked to changes in DNA topology. Many genes involved in the response to osmotic shock are also modulated in response to acid in my experiments, as previously seen (Chapter 2).

A major two-component system involved in regulating the acid stress response, the EvgAS system, did not emerge from this study. This is unsurprising as it is known that the expression of the evgAS operon is little affected by acid and so it would not have been detected as being important using the methods described here (Ma et al., 2004). Furthermore, the EvgAS system is mainly thought to be important in regulating the gad system, but this system has a limited role in the conditions described here, as can be seen by the fact that a *gadE* deletion only confers a mild phenotype. Different *E. coli* strains show a wide range of resistance to acid. It is not yet clear whether OmpR expression can explain some of the variation in this naturally occurring resistance. It may be that the relative importance of OmpR may depend on the strain being used, and that BW25133 is particularly dependent on OmpR. We are in the process of testing this hypothesis. Overall, three aspects of the physiological response of *E. coli* K-12 to mild acid stress require further explanation. First, the response regulator OmpR, which normally is thought of as responding to osmotic stress, is essential for adaptation to low pH and for priming survival of more severe acid stress. Second, many genes previously identified as components of OmpR regulon as well as additional previously unrecognized members, are essential for adaptation. Third, mild acid stress induces various metabolic switches, for example, from

glycolysis to gluconeogenesis and fatty acid synthesis, and from energy generation powered by the TCA cycle to expression of FNR-regulated genes associated with anaerobic respiration (**Table 3.4**).

3.4.1 GadW is a potential repressor of the F_OF_1 ATPase

Although the involvement of the F_OF_1 ATPase in the process of resistance to strong acid conditions is known, the mechanism of regulation is still unclear (Foster, 2004). The genes encoding the pump were all down-regulated in response to the acid exposure during the time course, as described in Chapter 2. These results, according to the general knowledge about acid resistance, confirmed that the ATPase is not expressed at pH 5.5 and in a rich growth medium, such LB. In the mutual information network I have identified a negative connection between GadW and *atpDGH*, encoding for 3 subunits of the F_OF_1 ATPase. GadW can be both activator and repressor of *gadA* and *gadBC* (Tramonti et al., 2008). The repression of the GAD system happens when cells enter the stationary phase, so under RpoS control (Burton et al., 2010). Therefore, the inhibition exerted by GadW during acid exposure involves not only the genes belonging to the GAD system but could also include the *atp* operon (**Figure 3.10**), perhaps under direct control of RpoS, since my experiments were performed at an early stationary phase of growth.

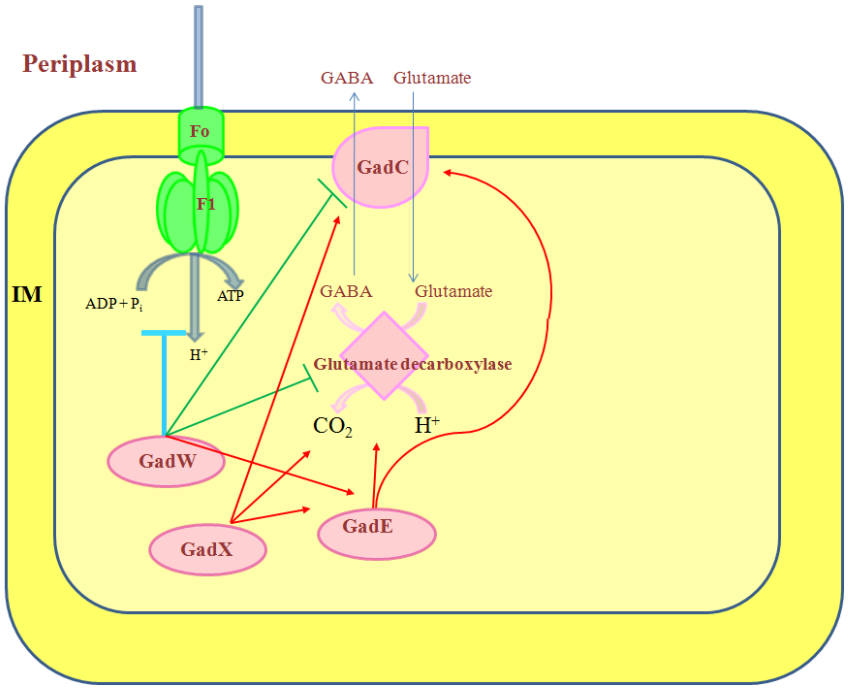


Figure 3.10 GadW repression of the *atp* operon hypothesis.

In figure are shown interactions from RegulonDB 9http://regulondb.ccg.unam.mx/), in red (activation) and green (repression), for GadW, GadX, and GadE towards the main component of the GAD system. In light blue, the interaction found in ARACNE, which describes the negative correlation between the two ARs.

The results of my experiments do not support the hypothesis that the negative correlation is a result of a direct regulatory role of GadW.

However, we cannot rule out the possibility that since the ATPase operon is composed of 9 genes and 3 transcription units (**Figure 3.11**), it is possible that the reporter gene I have chosen may not be relevant. Therefore further experiments would be required to completely rule out the regulatory role of GadW.

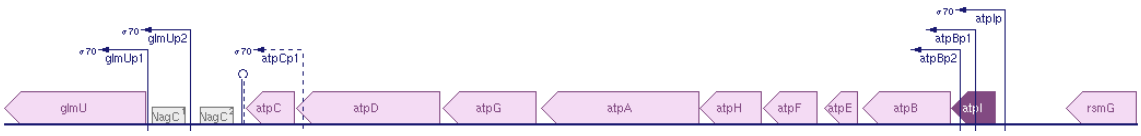


Figure 3.11 F_OF_1 ATPase gene operon (EcoCyc).

All the genes belonging to the ATPase are shown. *atpI,* even though belongs to the operon, does not seem to be essential for the functioning and the activation of the pump.

3.5 Conclusions.

In this chapter, a gene inference based model identified OmpR as a new regulator of acid response in *E. coli* BW25113, required for the switch from aerobic to anaerobic respiration. In the next chapter I will describe the mechanisms underlying acid response in another *E. coli* strain, the MG1655.

Chapter 4: Characterization of the genome-wide transcriptional response of the *E. coli* MG1655 strain to mild acid exposure

4.1 Introduction

In the previous chapters I have described a new acid response system and have shown that in the *E. coli* strain BW25113 this is playing a very important role, perhaps even more important than the classical ARs.

Although the result is important in its own right, the molecular response to acid adaptation is a complex phenomenon and involves the modulation of more genes than just the ARs and the novel OmpR dependent system I discovered. Moreover, it is possible that strain specific differences may play an important role. Indeed the analysis of BW25113 mutant strains suggested that there would be three different types of response to acid exposure, which may be indicative of the strain to strain variation in naturally occurring strains. For example, mutants belonging to an intermediate resistance group were showing up-regulation of aerobic and down-regulation of anaerobic pathways in response to pH 5.5, which is the opposite of what was observed in the wild type BW25113 strain.

Evidence exists for the involvement of other metabolic pathways in different stress responses. For example, *E. coli* W3110 also modulates the expression of genes encoding for enzymes involved in catabolism of sugar and amino acid in response to pH changes (Maurer et al., 2005). The increase of the amino acids levels is a typical feature of acid resistance, as previously described, but it also characterises other stress conditions (Jozefczuk et al.). In a recent study, based on the comparison of acid effects on the

MG1655 K-12 and SAKAI O157:H7 strains, it has been observed that the increase of the amino acid levels was in part due to the activation of de-novo biosynthesis pathways (King et al., 2010). In the same work it was also observed that the pathogenic *E. coli* strain was increasing the fermentative pathways and decreasing the aerobic metabolism (King et al., 2010), as seen for the BW25113 strain. Hence, the responses to stress conditions are also considerably variable between strains.

In order to fully understand the process underlying stress response, the comparison between strains becomes crucial. However, not many studies have been performed in this field by keeping the cultures of the analysed strains in identical conditions of growth. Moreover, it is unlikely that the analysis of transcriptomics data on its own can provide a full representation of the processes involved in acid adaptation, hence the need to use additional analytical techniques such as metabolomics.

The results reported in this chapter address these issues by providing the first characterization of the transcriptional and metabolic response of the *E. coli* MG1655 strain during acid adaptation. Since I wanted to make sure that a model built in the MG1655 strain could be compared with the results I previously obtained on the *E. coli* BW25113 strain I performed all culturing in exactly the same conditions.

My analysis revealed that, consistent with the model built in BW25113, the MG1655 intermediate acid resistant strain respond to acid exposure by up-regulating several enzymes involved in aerobic pathways and increasing the intracellular levels of amino acids. Moreover, the NMR metabolomics analysis showed an increase in amino acid levels, which could be caused by de-novo biosynthesis and/or proteolytic events.

4.2 Materials and Methods

4.2.1 Bacterial strains and culture conditions

MG1655 time course experiment. In order to perform a comparison with the BW25113 strain, the same experimental conditions were considered for the analysis of the MG1655 strain, as described in chapter 2. The experiments were performed by Dr Ayesha S. Rahman and Dr Sue Manzoor. The 1 hour time course experiment was performed, considering a total of fourteen time points every five minutes after exposure. Samples for transcriptomics and metabolomics analysis were collected every 5 minutes (14 time points in total), for 1 hour after the shift to pH 5.5. A time course experiments of 2 hours was carried out for comparison purposes with the Δ*lon* strain. The time course consisted of 8 time points, collected 30 seconds, 5', 15', 30', 45, 1 hour, 1hr 30' and 2 hours from exposure.

Lon mutant experiment. Δ*lon* strain was obtained from Dr P. Lund's laboratory. I have performed the time course experiment as described above for the wild type MG1655 strain 2 hours time course.

4.2.2 Expression profiling by microarray

The expression profiling microarray was performed as described in chapter 2, in materials and methods section.

4.2.3 Metabolomics experiments

The NMR metabolomics experiments for the characterization of the *E. coli* MG1655 strain were performed by Prof Mark Viant's laboratory, Dr A. Raman and Dr S.

Manzoor. Bacterial cultures were subjected to rapid filtration and washing prior metabolite extraction and NMR analysis. More precisely, a volume of culture (20ml) was applied to a 90 mm diameter Nylaflo membrane (Gelman) placed in a filtration assembly apparatus. The liquid culture was rapidly drawn through the filter using a Knf Neuberger Laboport vacuum pump. 20ml PBS (Lonza) was applied to the membrane to wash the bacterial cells and again this was rapidly drawn through the filter. The filter was then placed in a glass Petri dish containing 3 ml of 2:1 methanol water mix. This was then transferred onto dry ice to allow the contents to rapidly freeze thus preserving the metabolites present (this process was completed in less than 1 minute). After approximately 30 minutes on dry ice, the filter with the bacteria was transferred to a centrifuge tube and 1ml of 100% methanol was added to the Petri dish to remove any remaining metabolites/culture. The tube was then rotated and rocked on platform at room temperature for 5 minutes and then transferred to -20^{o}C. Bacterial cell debris was removed after the samples were defrosted by two centrifugation steps. The supernatant was then removed to a fresh 2 ml eppendorf and evaporated to dry in a speed vacuum rotor (Thermo Electron Corporation) attached to a pump (Knf laboport) and refrigerated vapour trap (Thermo Electron Corporation). Samples were then resuspended in 600 µl 200 mM 10% D2O buffer. Metabolite standards were analyzed on a DRX-500 NMR spectrometer (Bruker Biospin, Coventry, UK) equipped with a 5 mm TXI cryoprobe and BACS-60 automatic sample changer (Ludwig et al., 2012). Identification of metabolites in NMR spectra was performed through the Madison Metabolomics Consortium Database (MMCD) (Cui et al. 2008)

4.2.3 Data analysis

Microarrays Data processing and Functional annotation of the gene lists were performed as described in chapter 2, section materials and methods.

Metabolomics Data analysis: 1801 bins were identified.. After SAM analysis between the three replicates, 58 significant bins were found. The annotation was possible only for 12 metabolites.

4.3 Results

4.3.1 The transcriptional response of *E. coli* MG1655 to acid exposure is transient

The analysis of the time course microarray data identified 1831 genes differentially expressed (FDR<1%). These were used as an input of PCA to visualize the overall dynamics of transcriptional change during acid adaptation. Similarly to the BW25113 strain (chapter 2), the transcriptional profile of the MG1655 strain was characterised by a rapid shift in the first 30 seconds of exposure at pH 5.5, followed by a further shift on the first and second components in the 25 minutes post exposure (**Figure 4.1**). After 30 minutes, the transcriptional state of the cells gradually realigned to the transcriptional state seen at pH 7.

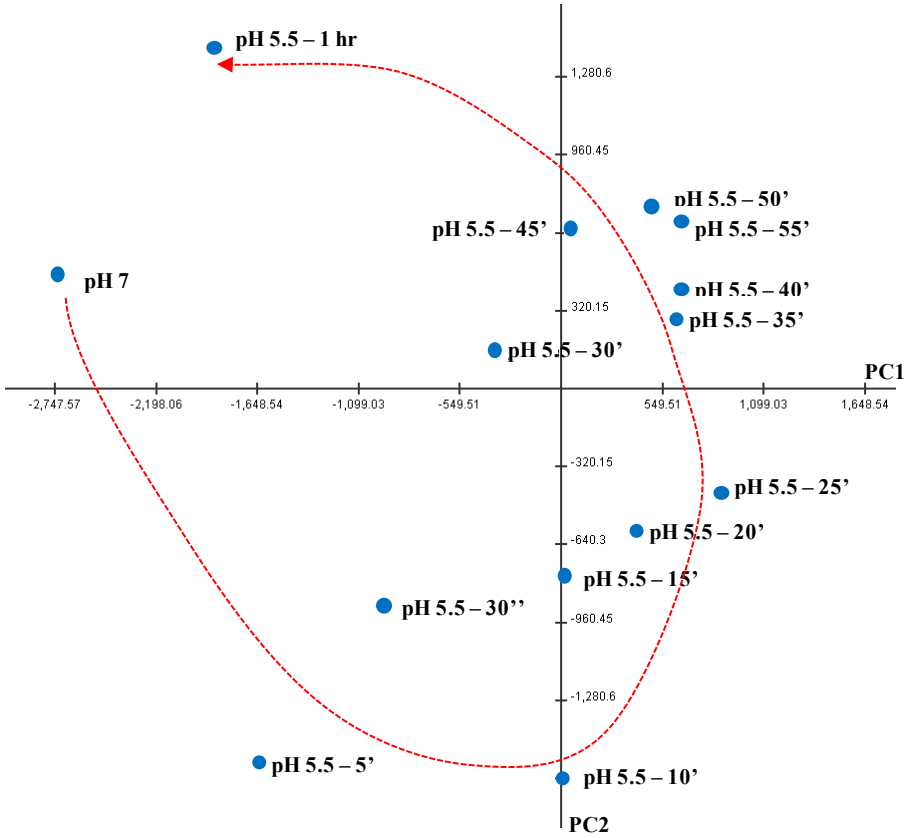
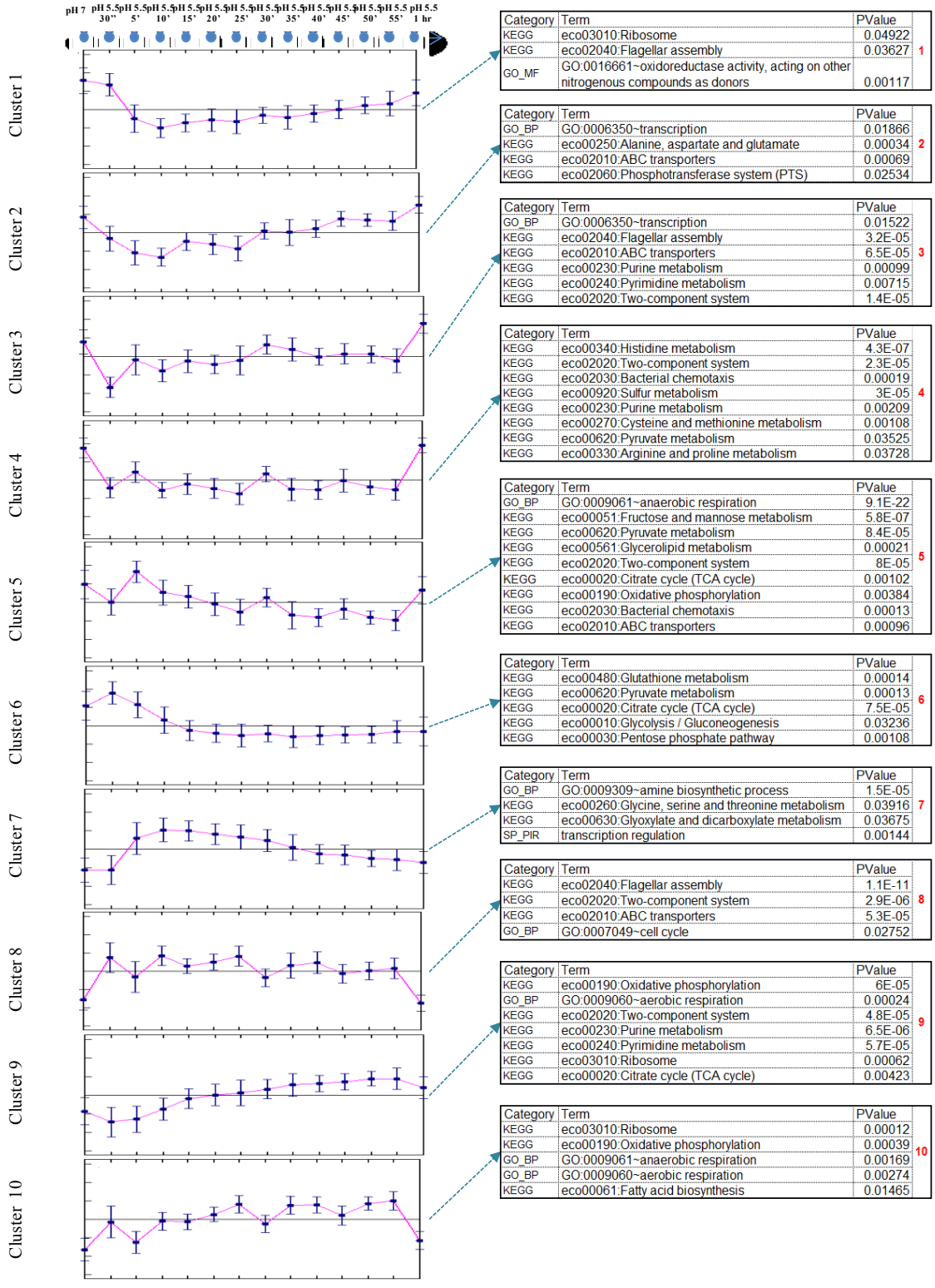


Figure 4.1 Principle component analysis of the transcriptional response of MG1655 to acid adaptation.

This PC plot shows the dynamics of change in the transcriptional response of MG1655 during the first hour of acid adaptation. The x and y axes represent the first and second principal components respectively. For the analysis were considered the 1831 differentially expressed genes.

In order to further characterize the dynamics of gene expression I performed cluster analysis. The analysis identified 10 clusters (**Figure 4.2**).



Category	Term	PValue	
KEGG	eco03010:Ribosome	0.04922	
KEGG	eco02040:Flagellar assembly	0.03627	1
GO_MF	GO:0016661~oxidoreductase activity, acting on other nitrogenous compounds as donors	0.00117	

Category	Term	PValue	
GO_BP	GO:0006350~transcription	0.01866	
KEGG	eco00250:Alanine, aspartate and glutamate	0.00034	2
KEGG	eco02010:ABC transporters	0.00069	
KEGG	eco02060:Phosphotransferase system (PTS)	0.02534	

Category	Term	PValue	
GO_BP	GO:0006350~transcription	0.01522	
KEGG	eco02040:Flagellar assembly	3.2E-05	
KEGG	eco02010:ABC transporters	6.5E-05	3
KEGG	eco00230:Purine metabolism	0.00099	
KEGG	eco00240:Pyrimidine metabolism	0.00715	
KEGG	eco02020:Two-component system	1.4E-05	

Category	Term	PValue	
KEGG	eco00340:Histidine metabolism	4.3E-07	
KEGG	eco02020:Two-component system	2.3E-05	
KEGG	eco02030:Bacterial chemotaxis	0.00019	
KEGG	eco00920:Sulfur metabolism	3E-05	4
KEGG	eco00230:Purine metabolism	0.00209	
KEGG	eco00270:Cysteine and methionine metabolism	0.00108	
KEGG	eco00620:Pyruvate metabolism	0.03525	
KEGG	eco00330:Arginine and proline metabolism	0.03728	

Category	Term	PValue	
GO_BP	GO:0009061~anaerobic respiration	9.1E-22	
KEGG	eco00051:Fructose and mannose metabolism	5.8E-07	
KEGG	eco00620:Pyruvate metabolism	8.4E-05	
KEGG	eco00561:Glycerolipid metabolism	0.00021	
KEGG	eco02020:Two-component system	8E-05	5
KEGG	eco00020:Citrate cycle (TCA cycle)	0.00102	
KEGG	eco00190:Oxidative phosphorylation	0.00384	
KEGG	eco02030:Bacterial chemotaxis	0.00013	
KEGG	eco02010:ABC transporters	0.00096	

Category	Term	PValue	
KEGG	eco00480:Glutathione metabolism	0.00014	
KEGG	eco00620:Pyruvate metabolism	0.00013	6
KEGG	eco00020:Citrate cycle (TCA cycle)	7.5E-05	
KEGG	eco00010:Glycolysis / Gluconeogenesis	0.03236	
KEGG	eco00030:Pentose phosphate pathway	0.00108	

Category	Term	PValue	
GO_BP	GO:0009309~amine biosynthetic process	1.5E-05	
KEGG	eco00260:Glycine, serine and threonine metabolism	0.03916	7
KEGG	eco00630:Glyoxylate and dicarboxylate metabolism	0.03675	
SP_PIR	transcription regulation	0.00144	

Category	Term	PValue	
KEGG	eco02040:Flagellar assembly	1.1E-11	
KEGG	eco02020:Two-component system	2.9E-06	8
KEGG	eco02010:ABC transporters	5.3E-05	
GO_BP	GO:0007049~cell cycle	0.02752	

Category	Term	PValue	
KEGG	eco00190:Oxidative phosphorylation	6E-05	
GO_BP	GO:0009060~aerobic respiration	0.00024	
KEGG	eco02020:Two-component system	4.8E-05	9
KEGG	eco00230:Purine metabolism	6.5E-06	
KEGG	eco00240:Pyrimidine metabolism	5.7E-05	
KEGG	eco03010:Ribosome	0.00062	
KEGG	eco00020:Citrate cycle (TCA cycle)	0.00423	

Category	Term	PValue	
KEGG	eco03010:Ribosome	0.00012	
KEGG	eco00190:Oxidative phosphorylation	0.00039	10
GO_BP	GO:0009061~anaerobic respiration	0.00169	
GO_BP	GO:0009060~aerobic respiration	0.00274	
KEGG	eco00061:Fatty acid biosynthesis	0.01465	

Figure 4.2 Hierarchical clustering of the genes differentially expressed in response to acid exposure in the strain MG1655.

The hierarchical clustering performed on the genes differentially expressed during the time course experiments identified 10 clusters. 5 clusters (1,2,3,4,5) were showing a trend of down-regulations,

while the rest of the clusters (6,7,8,9,10) were up-regulated in response to acid. The functional annotation analysis, performed with DAVID, defined a switch of regulation in favour of the genes encoding aerobic respiration pathways. The genes coding for TCA cycle, Glycolysis and oxidative phosphorylation were mainly found in clusters in 6, 9 and 10. For the other functions, such as flagellar assembly and ABC transporters, it was not possible to identify a specific trend.

Clusters 1-5 represented down-regulated gene expression profiles followed by a recovery phase with the exception of cluster 5 that showed a persistent down-regulation. Clusters 6-10 represented up-regulated gene expression profiles. Clusters 8 and 10 also showed a later recover phase. The fact that most of the clusters show a recovery phase is consistent with the results of the PCA. In order to allow biological interpretation, I performed a functional enrichment analysis. Clusters 4, 5, 8 and 10 were characterised by oscillations, which were not due to the normalization process but could be considered effects of acid exposure.

This analysis revealed a clear trend in the regulation of aerobic and anaerobic pathways. More precisely I observed an increase in the expression of the genes involved in aerobic metabolic processes (**Figures 4.4 and 4.5**) and the down-regulation of 42 genes coding for anaerobic and fermentative pathways (**Figure 4.3**). Therefore the two strains (BW25113 and MG1655) were showing a completely opposite profile of gene expression during acid exposure (**Table 4.1**). However, some of the genes encoding anaerobic respiration enzymes were found up-regulated in response to acid exposure (**Figure 4.3**)

A similar behaviour was found in the BW25113 mutant strains belonging to the intermediate phenotype group (described in chapter 2, Table 2.1).

Table 4.1 Comparison of aerobic and anaerobic functions in BW25113 and MG1655 strains.

The response to pH 5.5 was characterised by two opposite trend in BW25113 and MG1655 strains. While in the first strain the major aerobic functions encoding genes (Oxidative phosphorylation, TCA cycle, Pyruvate metabolism) are strongly down-regulated, an opposite trend was detected in the MG1655. The p-values for each function were calculated with DAVID cluster analysis, based on the groups of up and down-regulated genes for each strain.

	BW25113		MG1655	
Functions	**Regulation**	**p-value**	**Regulation**	**p-value**
GO - Aerobic respiration	down-regulated	5.70E-19	up-regulated	9.70E-22
GO - Anaerobic respiration	up-regulated	2.70E-24	down-regulated	1.80E-28
KEGG - Oxidative phosphorylation	down-regulated	2.80E-26	up-regulated	3.70E-13
KEGG - TCA cycle	down-regulated	2.70E-21	up-regulated	1.00E-13
KEGG - Glycolysis/Gluconeogenesis	down-regulated	1.20E-10	up-regulated	9.50E-04
KEGG - Pyruvate metabolism	down-regulated	2.40E-18	up-regulated	4.00E-10

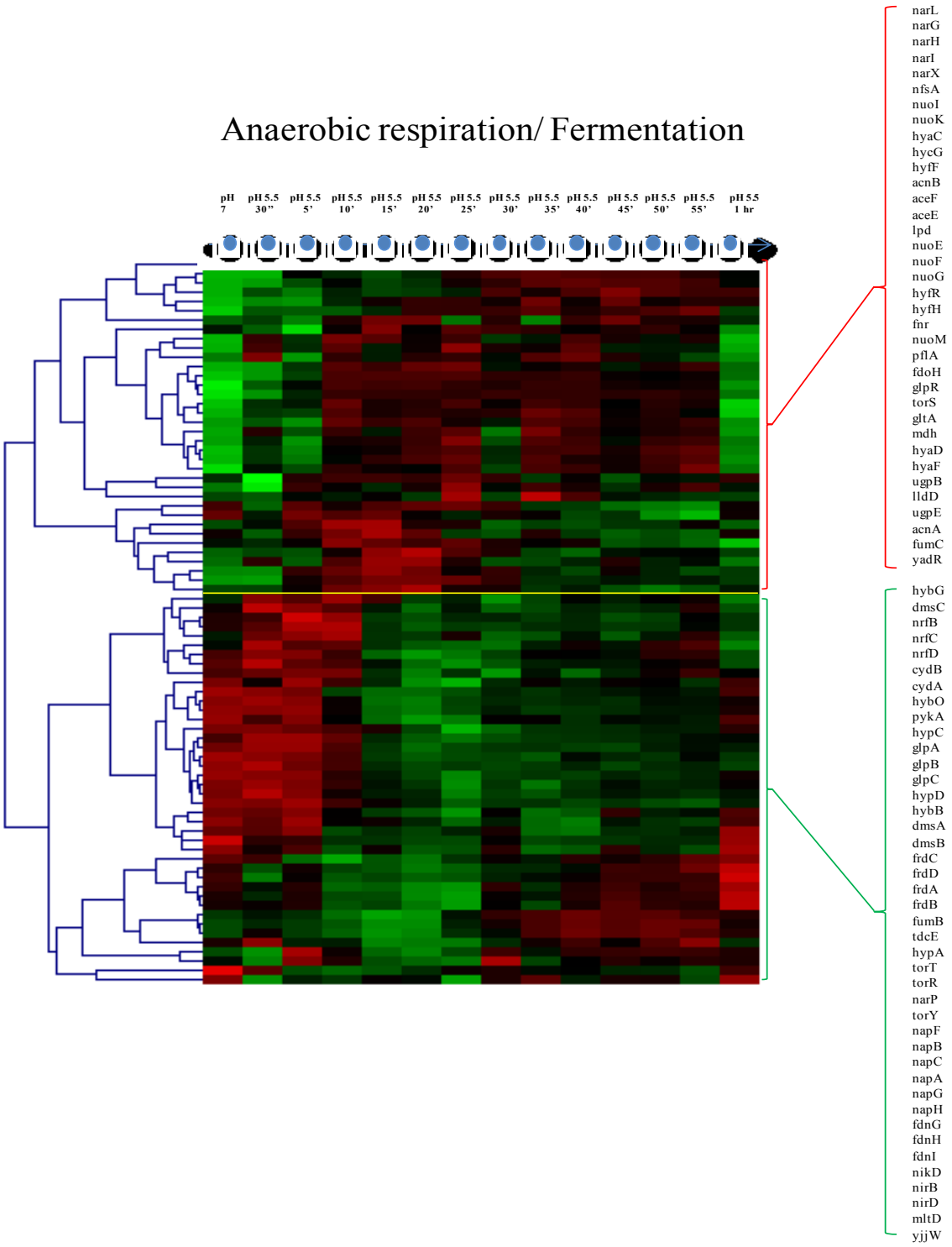


Figure 4.3 Heat map of the genes belonging to the anaerobic respiration pathway (GO:0009061).

In this figure are shown the genes encoding for the anaerobic respiration pathway differentially expressed in *E. coli* MG1655 strain. 43 genes over 79 were down-regulated in response to pH 5.5 (green line), while 36 were up-regulated (red line) on the cluster based analysis. Between the up-regulated genes, many of them encode for components (enzymes and transporters) involved in both aerobic and anaerobic mechanisms.

The cluster of the anaerobic respiration encoding genes showed that 36 genes of 79 were up-regulated over time (**Figure 4.3**). It is important to notice that this cluster included genes coding for enzymes that catalyse both aerobic and anaerobic reactions. For example, the NADH dehydrogenase enzyme, which was highly induced, is involved in the oxidative phosphorylation (Jaworowski et al., 1981) as well as in anaerobic processes (Matsushita et al., 1987). The genes *aceE* and *aceF* encode both components of the pyruvate dehydrogenase, which again is involved in aerobic (Abdel-Hamid et al., 2001) and anaerobic pathways (Clark, 1989). A total of 15 genes belonging to the cluster of the up-regulated anaerobic genes encode for components (enzymes and transporters) which entail both aerobic and anaerobic processes (**Table 4.2**).

Table 4.2 Anaerobic respiration genes up-regulated during acid exposure

The cluster analysis identified 36 genes coding for anaerobic respiration pathways induced at pH 5.5. 15 genes belonging to this cluster can be involved in both aerobic and anaerobic functions (yellow cells). Between them I have identified the NADH dehydrogenase, the pyruvate dehydrogenase, the malate dehydrogenase and the essential respiratory protein A encoding genes.

Gene	Function
aceE	Pyruvate dehydrogenase E1 component
aceF	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
acnA	Aconitate hydratase 1
acnB	Aconitate hydratase 2
fdoH	Formate dehydrogenase-O iron-sulfur subunit
fnr	Fumarate and nitrate reduction regulatory protein
fumC	Fumarate hydratase class II
glpR	Glycerol-3-phosphate regulon repressor
gltA	Citrate synthase
hyaC	Probable Ni/Fe-hydrogenase 1 B-type cytochrome subunit
hyaD	Hydrogenase 1 maturation protease
hyaF	Hydrogenase-1 operon protein hyaF
hycG	Formate hydrogenlyase subunit 7
hyfF	Hydrogenase-4 component F
hyfH	Hydrogenase-4 component H
hyfR	Hydrogenase-4 transcriptional activator
lldD	L-lactate dehydrogenase [cytochrome]
lpd	Dihydrolipoyl dehydrogenase
mdh	Malate dehydrogenase
narG	Respiratory nitrate reductase 1 alpha chain
narH	Respiratory nitrate reductase 1 beta chain
narI	Respiratory nitrate reductase 1 gamma chain
narL	Nitrate/nitrite response regulator protein narL
narX	Nitrate/nitrite sensor protein narX
nfsA	Oxygen-insensitive NADPH nitroreductase
nuoE	NADH-quinone oxidoreductase subunit E
nuoF	NADH-quinone oxidoreductase subunit F
nuoG	NADH-quinone oxidoreductase; NADH-quinone oxidoreductase subunit G
nuoI	NADH-quinone oxidoreductase subunit I
nuoK	NADH-quinone oxidoreductase subunit K
nuoM	NADH-quinone oxidoreductase subunit M
pflA	Pyruvate formate-lyase 1-activating enzyme
torS	Sensor protein torS
ugpB	sn-glycerol-3-phosphate-binding periplasmic protein ugpB
ugpE	sn-glycerol-3-phosphate transport system permease protein ugpE
yadR	essential respiratory protein A

Since MG1655 response is characterised by an aerobic switch, it is possible to consider the induction of these components as part of the aerobic response to pH 5.5.

Between the anaerobic processes genes showing up-regulation, I have identified *hyaD*, *hyaF*, *hyfF*, *hyfH* and *hyfR*, encoding for the hydrogenases 1 and 4 in *E. coli* (Andrews et al., 1997, Sargent et al., 1998). The genes coding for the nitrate reductase system (*narG*, *narH*, *narI*, *narL* and *narX*) were also up-regulated. This enzyme is required for respiration with nitrate as final electron acceptor (Guigliarelli et al., 1996). The gene expression is controlled by RpoS during stationary phase of growth (Chang et al., 1999), which might activate the enzyme during acid exposure in early stationary phase of growth.

In the oxidative phosphorylation pathway, I have observed the activation of the genes of the NADH dehydrogenase, as previously described, together with the succinate dehydrogenase and the cytochrome c-oxidase genes (**Figure 4.4**).The anaerobic components of this pathway, the *frdABCD* (fumarate reductase) and the *cydAB* (cytochrome bd complex), were instead down-regulated (**Figure 4.4**).

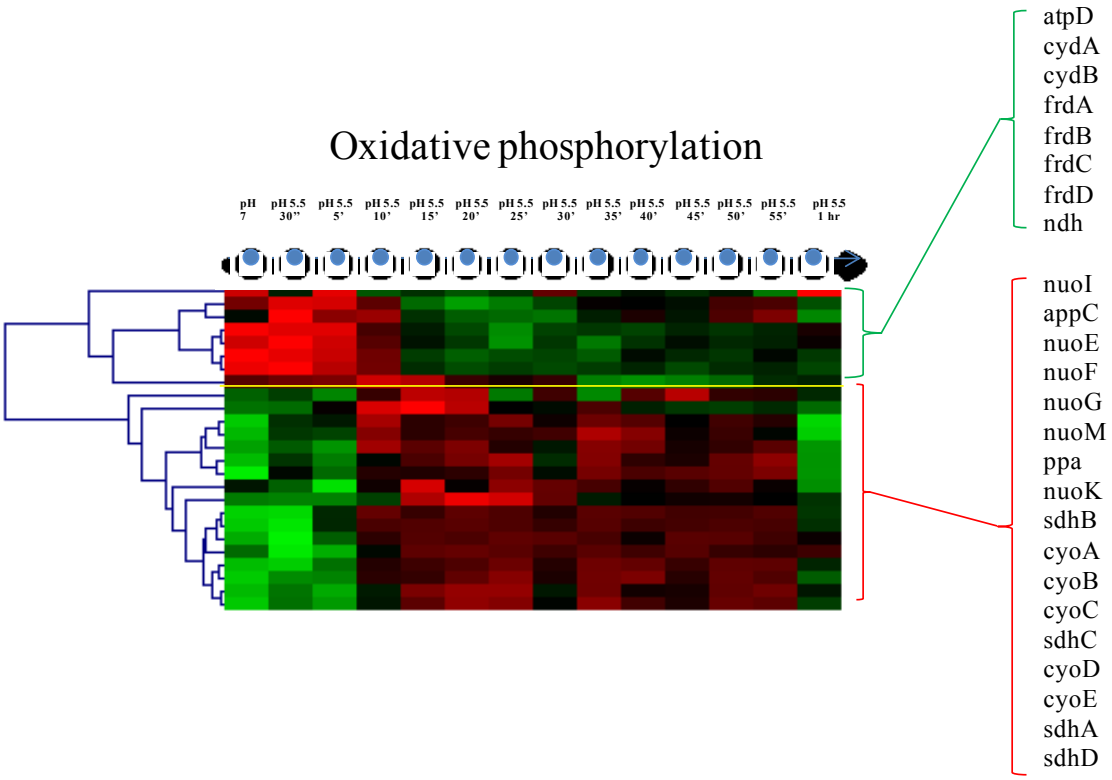


Figure 4.4 Gene regulation of oxidative phosphorylation pathway (KEGG:eco00190).

The increased expression of the NADH dehydrogenase, succinate dehydrogenase and the cytochrome c oxidase was a result of the aerobic respiration induction (in red), while the fumarate reductase and the cytochrome bd genes were down-regulated (in green). The genes encoding the F_OF_1ATPase were also down-regulated, as expected from literature on acid response.

TCA cycle pathway encoding genes were all up-regulated (**Figure 4.5**) with the exclusion of the phosphoenolpyruvate carboxykinase gene, *pck*.

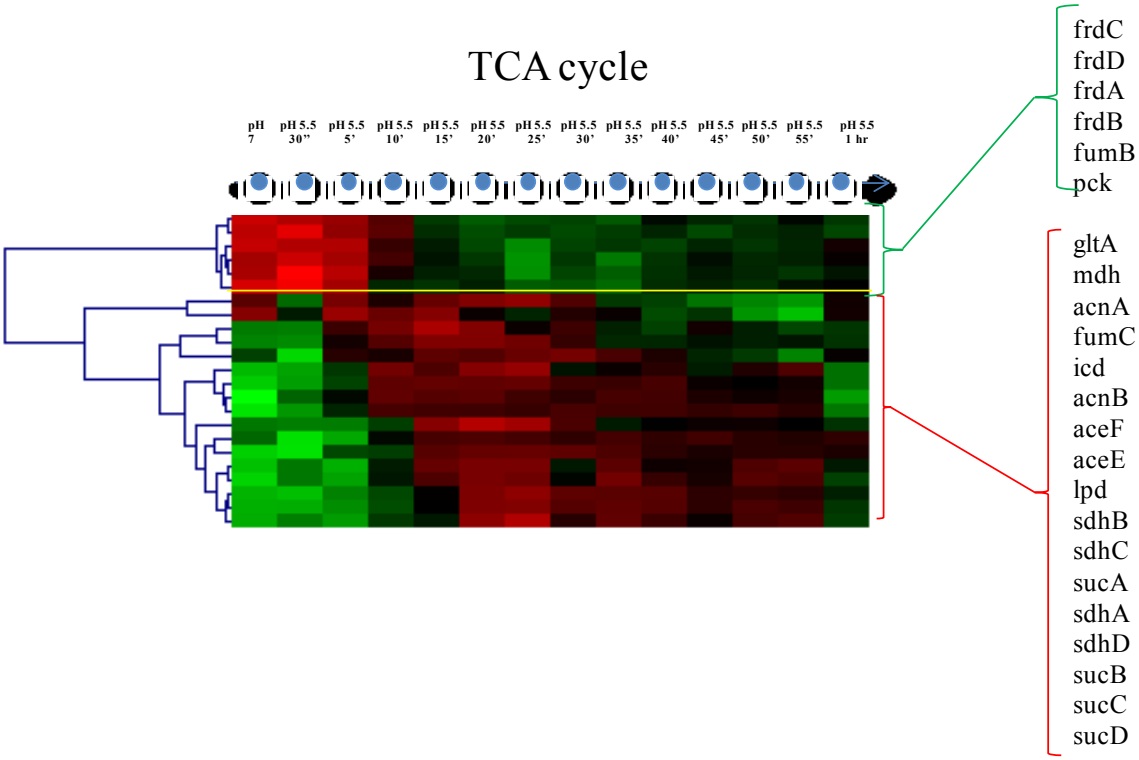


Figure 4.5 Regulation of the genes belonging to the TCA cycle pathway (KEGG:eco00020). The majority of the genes belonging to this pathway were up-regulated at pH 5.5 (in red). The only exception was the pck gene, encoding for the phosphoenolpyruvate carboxykinase (in green).

Furthermore, I found an increase in the expression of the gene *fumC*, one of the fumarase isozymes, which is not essential when oxygen concentrations are limited, whereas *fumB* is significantly down-regulated.

ABC transporters are also affected by acid exposure, but it is not possible to identify a clear trend of regulation for them (**Figure 4.6**).

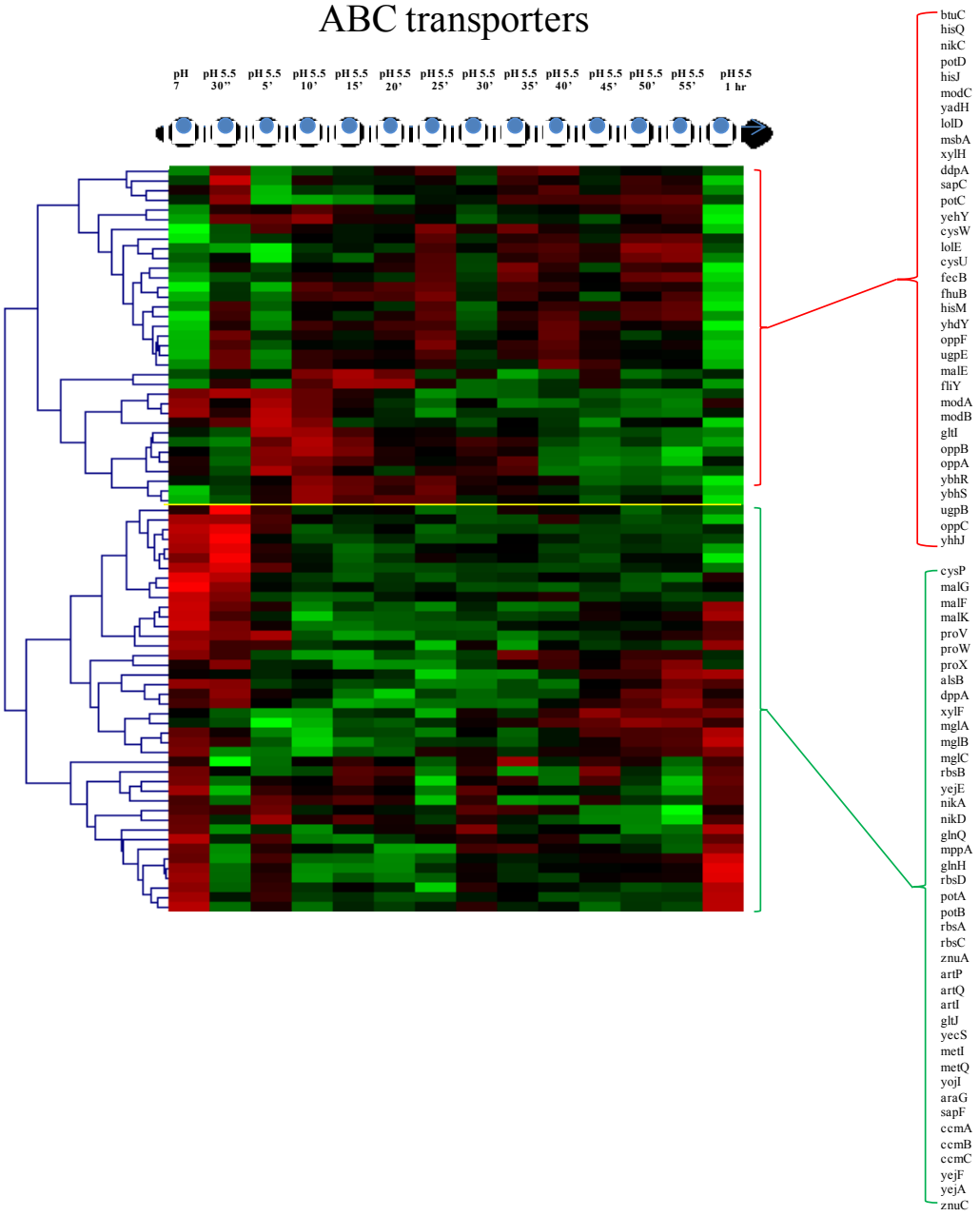


Figure 4.6 Heat map of the genes belonging to the ABC transporter family (eco02010).

The genes encoding for the ABC transporters were not characterised by a defined trend. However, many amino acid and sugar transporters were found to be down-regulated.

However, some of the amino acid transporters, such as arginine, valine, aspartate, and tyrosine transporters appeared down-regulated with most of the sugars transporters. The KEGG pathway for ribosome (**Figure 4.7**) was mostly up-regulated after 25 minutes of exposure.

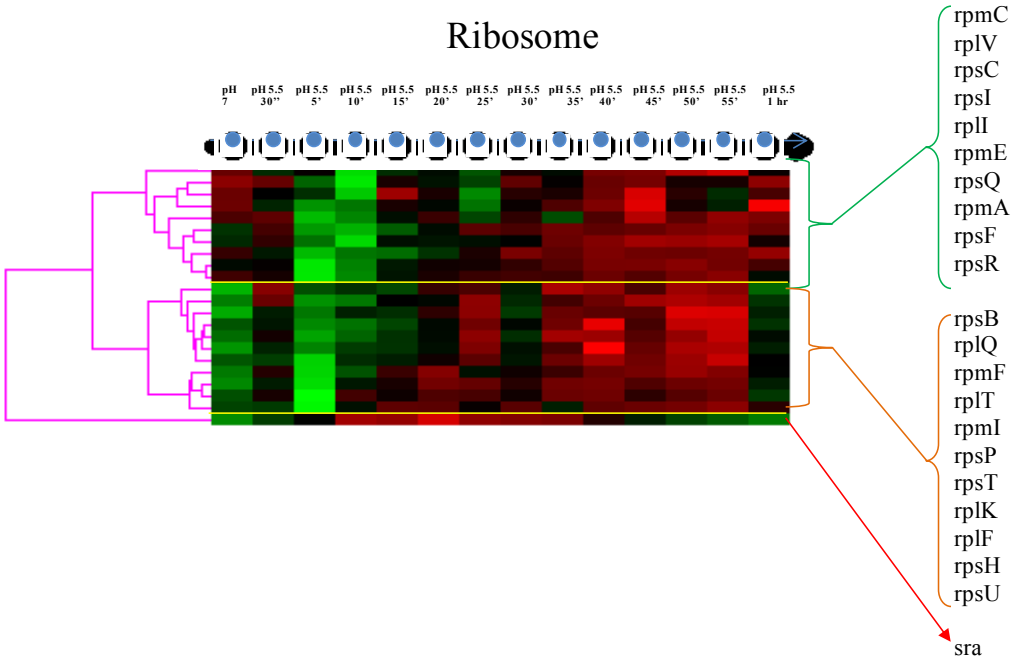


Figure 4.7 Heat map of the genes belonging to the ribosome pathway (eco03010).

The ribosome encoding genes were all highly induced after 35 minutes of exposure, denoting a potential need of the cells to re-activate translational pathways.

My data also showed another interesting pattern involving the genes coding for the flagellar assembly (**Figure 4.8**).

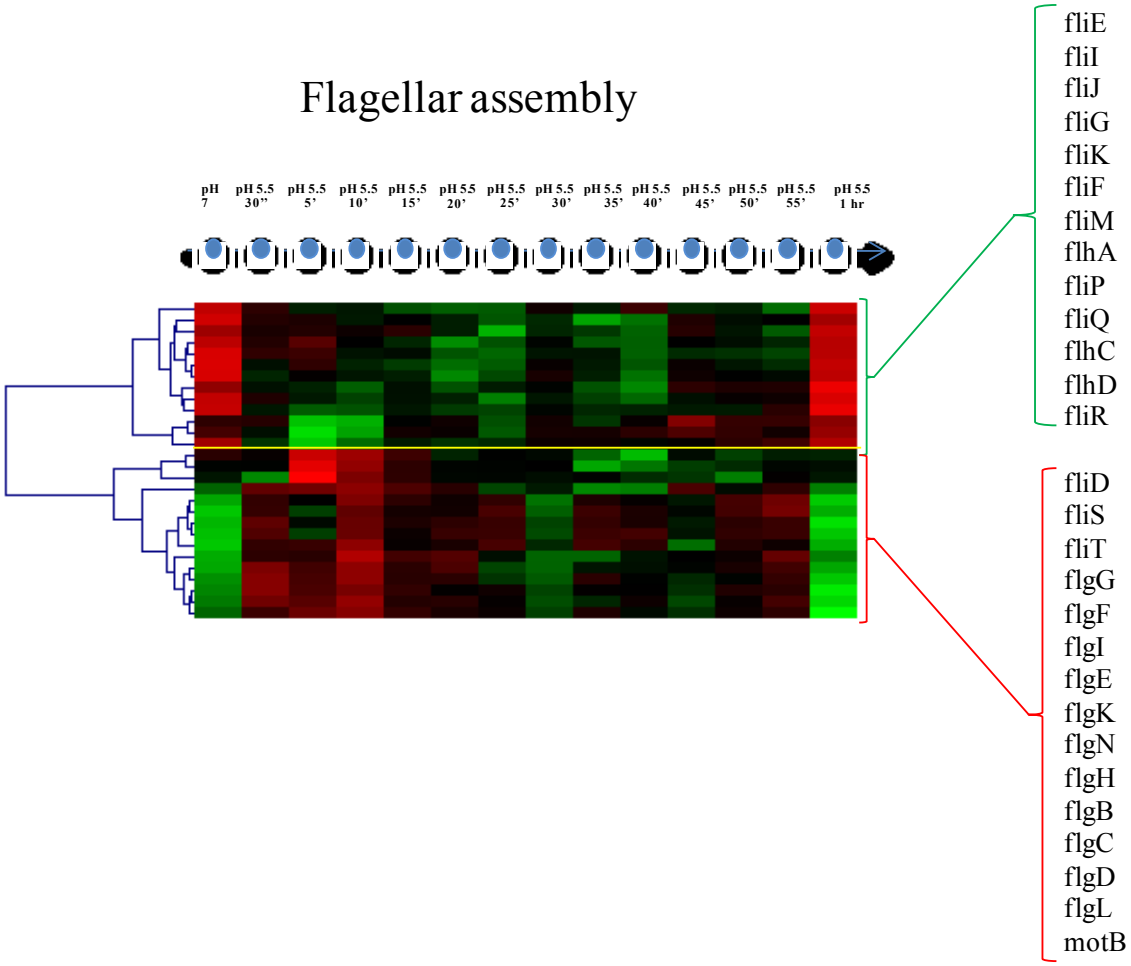


Figure 4.8 Regulation the genes belonging to the flagellar assembly pathway (KEGG:eco02040).

Several genes encoding for the filament part of the flagellum were significant induced in response to acid (in red), while the main components of the basal ring were repressed (in green).

The trend found was not typical for the general stress response in *E. coli*, in fact most of the times genes belonging to this pathway are usually down-regulated in response to mild acidic conditions (Maurer et al., 2005, Hayes et al., 2006). In my data, most of the genes coding for the filament part of the flagellum were immediately up-regulated after the switch to pH 5.5; vice versa, the genes for the flagellar motor was down-regulated after only 30 seconds of exposure. This could be a strategy of the bacterium to respond to acid condition, probably because of the damage of low pH exposure to the filament.

This preliminary analysis of the transcriptomic data for the MG16555 strain therefore showed a completely opposite behaviour when compared to the BW25113 strain. As a result of the acid exposure, it was seen that the cells immediately required the transcription of the aerobic pathways (**Table 4.3** and **Figure 4.9**).

Table 4.3 Regulation of the main components of Oxidative phosphorylation and TCA cycle in BW25113 and MG1655 strains

The differentially expressed genes encoding the enzymes of two main aerobic pathways were considered in this table. As previously described, the BW25113 showed a trend of down-regulation of all the considered enzymes. The aerobic components of oxidative phosphorylation and TCA cycle were all up-regulated in MG1655, with the exception of cytochrome - *bd*, the fumarate reductase and the phosphoenolpyruvate carboxykinase encoding genes. In both the strains, the gene *atpD*, coding for the β-subunit of the F_OF_1 ATPase was down-regulated in both strains. Up-regulated genes are indicated with red upward arrows; down-regulated genes are indicated with the green downward arrows.

			BW25113	MG1655
Oxidative phosphorylation	FoF1 ATPase	atpD	↓	↓
	Cytochrome - *bd*	cydA	↓	↓
		cydB	↓	↓
	Cytochrome - *bo*	cyoA	↓	↑
		cyoB	↓	↑
		cyoC	↓	↑
		cyoD	↓	↑
		cyoE	↓	↑
	Fumarate reductase	frdA	↓	↓
		frdB	↓	↓
		frdC	↓	↓
		frdD	↓	↓
	NADH Dehydrogenase	nuoE	↓	↑
		nuoF	↓	↑
		nuoG	↓	↑
		nuoK	↓	↑
	Succinate dehydrogenase	sdhA	↓	↑
		sdhB	↓	↑
		sdhC	↓	↑
		sdhD	↓	↑
TCA cycle	Pyruvate dehydrogenase	aceE	↓	↑
		aceF	↓	↑
	Aconitate hydratase 1	acnA	↓	↑
	Aconitate hydratase 2	acnB	↓	↑
	Fumarate reductase	fumC	↓	↑
	Cytrate synthase	gltA	↓	↑
	Isocitrate dehydrogenase	icd	↓	↑
	Lipoamide dehydrogenase	lpd	↓	↑
	Malate dehydrogenase	mdh	↓	↑
	Phosphoenolpyruvate carboxykinase	pck	↓	↓
	2-oxoglutarate decarboxylase	sucA	↓	↑
		sucB	↓	↑
	Succinyl-CoA synthetase	sucC	↓	↑
		sucD	↓	↑

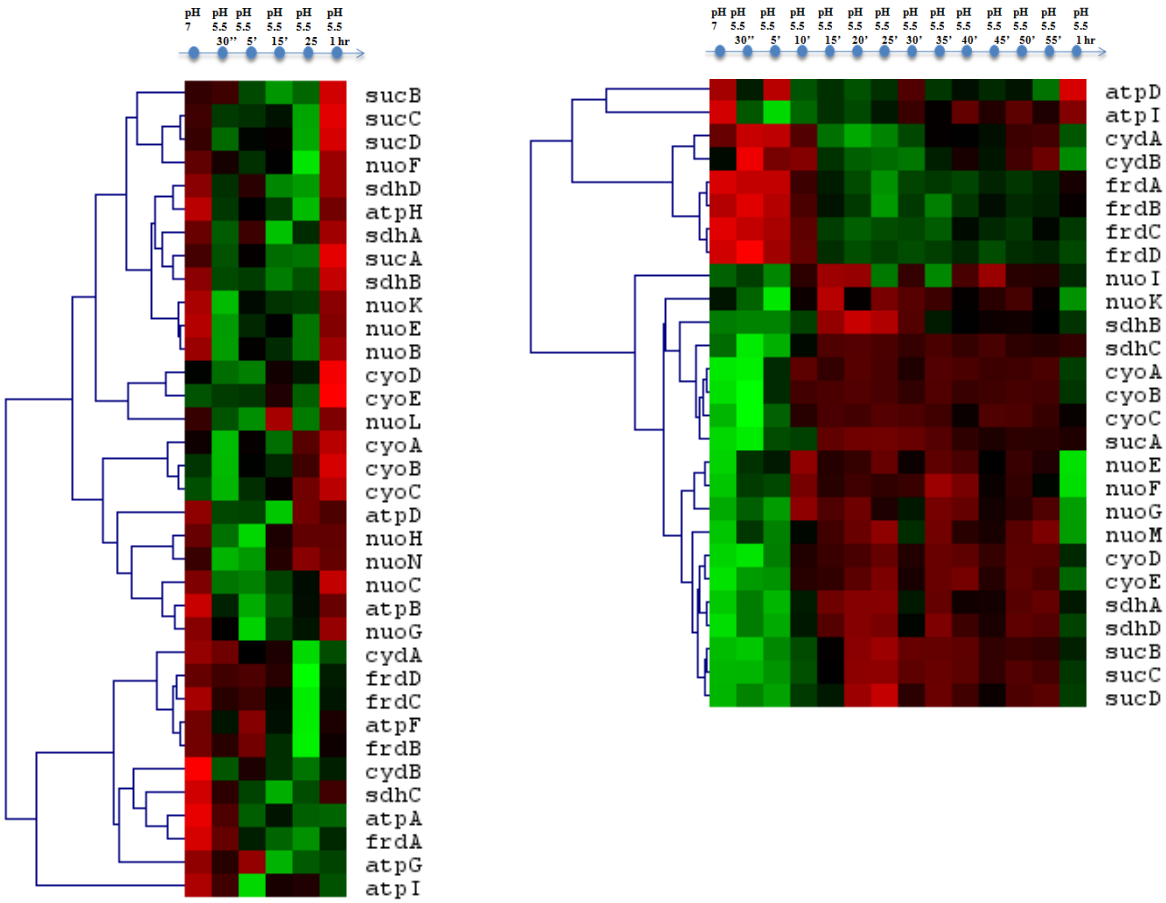


Figure 4.9 Selection of the genes encoding for important enzymes involved in aerobic pathways.

The genes encoding for the NADH dehydrogenase, succinate dehydrogenase, 2-oxoglutarate decarboxylase, cytochrome-bd and cytochrome-bo and the fumarate reductase were selected between the genes differentially expressed in the two strains. The BW25113 strain was characterised by the down-regulation of all these components during the time course, while MG1655 showed an increased expression of the NADH dehydrogenase, cytochrome-bo and the 2-oxoglutarate genes, and down-regulation for the fumarate reductase, the F_OF_1 ATPase and the cytochrome-bd encoding genes.

In order to understand these results, I have considered also a metabolomics screening, for identification of the metabolites affected by acid exposure.

4.3.2 Metabolomics analysis of the *E. coli* MG1655 strain during acid adaptation

Since the transcriptomic analysis revealed that genes involved in central metabolism were all differentially regulated in response to acid exposure I decided to use a metabolomics approach to integrate the analysis at the metabolite level (**Figure 4.10**).

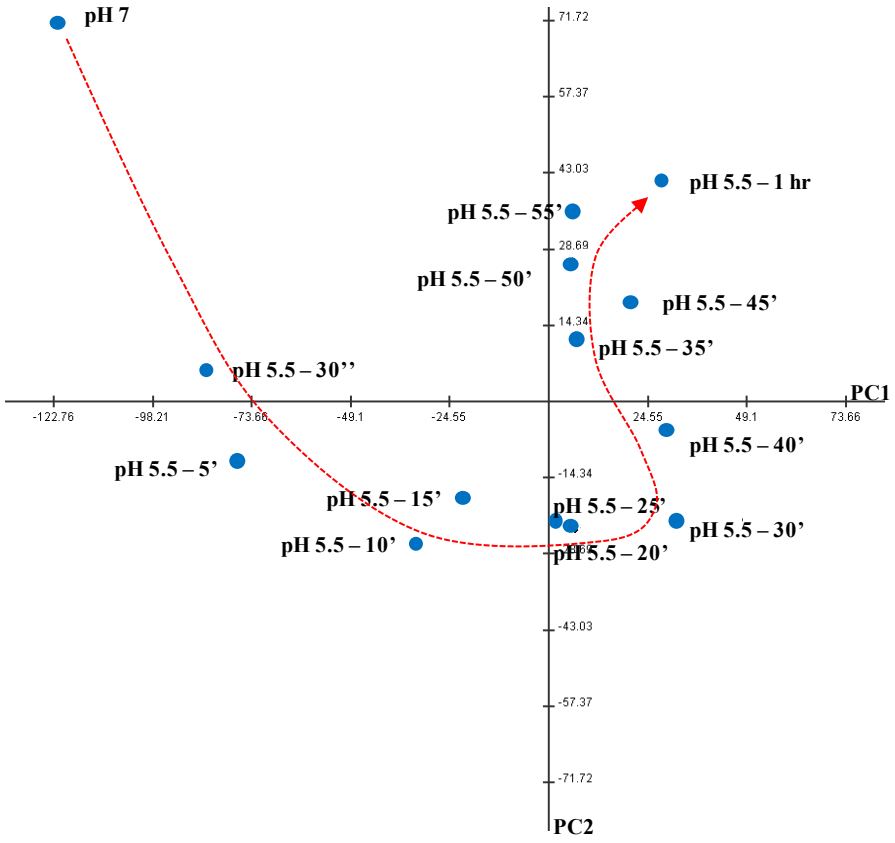


Figure 4.10 Principle component analysis of the metabolomics response of MG1655 to acid adaptation.

The PCA plot was performed on the significant metabolites changed after acid exposure. The trend is characterised by a rapid shift across the first component, denoting the change due to the acid perturbations. In contrast to the previous findings at the transcriptomic level, the metabolites did not restore their molecular state, seen at pH 7.

PCA plot performed on the significant metabolites confirmed the ability of acid exposure to affect the cells at different levels. However, in contrast to the transcriptional response, the metabolites did not restore the metabolic state seen at pH 7. After one hour at pH 5.5, it was not possible to detect further changes. The same experiment performed for 2 hours and 30' did not show further changes at metabolites level (data not shown).

Of the 1350 integrated peaks that represented the NMR spectra 58 were significantly different during the time course, (**Figure 4.11**). These were used as input of a hierarchical clustering procedure.

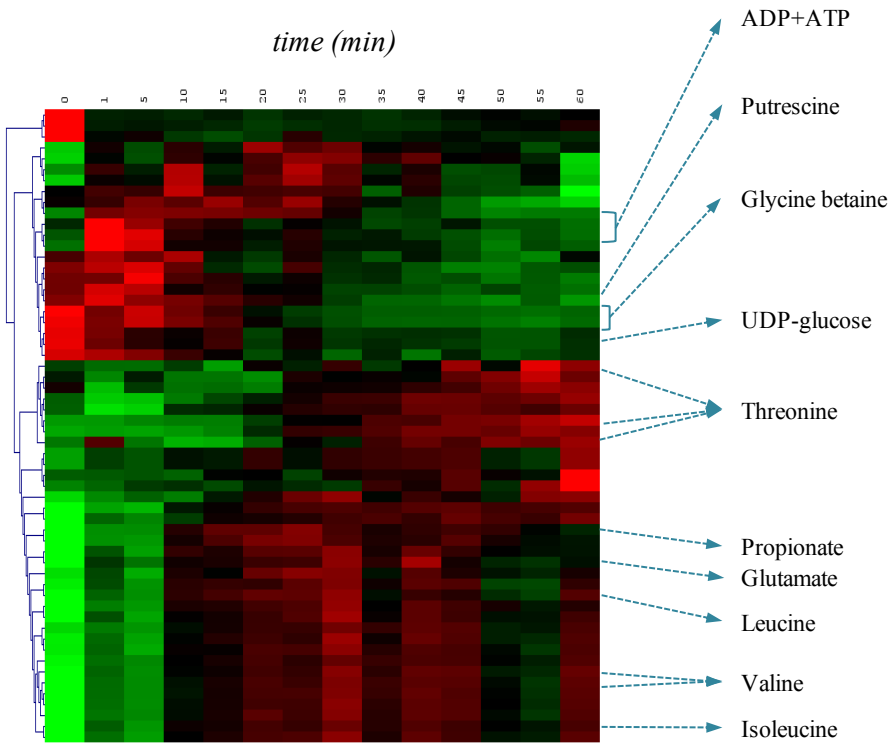


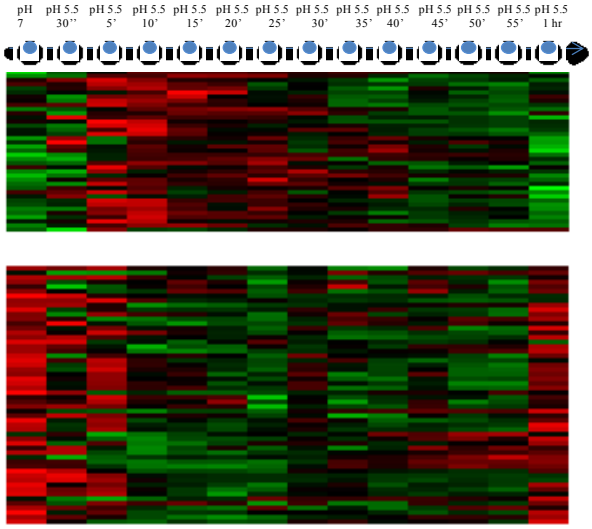
Figure 4.11 Hierarchical clustering of the significant metabolites.

The main result of this analysis was the rapid increase found at the amino acid levels. Threonine, glutamate, leucine, valine and isoleucine were rapidly increasing after pH 5.5 exposure, while putrescine and glycine betaine were decreasing. Between the identified bins, I have also seen the ADP+ATP. The technique did not allow distinguishing between the two species.

The concentration of Putrescine, a poly-amine formed either directly from L-ornithine by ornithine decarboxylase or indirectly from L-arginine by arginine decarboxylase, was found reduced in response to acid. Same effects were seen for the osmolyte glycine betaine, for the peaks considered for ATP/ADP and UDP glucose/sucrose. Finally, for the amino acids threonine, glutamate, leucine, valine and isoleucine and propionate I found a considerable increase in concentration in response to pH 5.5.

4.3.3 Increased concentration of amino acids may be the result of protein degradation

This increase of amino acids following acid exposure is interesting and could be the result of increased synthesis or import or be the result of protein degradation. I first decided to test the hypothesis that it may be dependent on increased transport. I tested this hypothesis by looking at the expression of genes encoding for amino acid co-transporters (**Figure 4.12**).



UP					
Category	**Term**	**Count**	**PValue**	**Genes**	**Benjamini**
GO_BP	GO:0009309~amine biosynthetic process	26	5E-30	ASD, ILVI, ILVE, METK, GDHA, ASPC, TYRB, SPEG, GLTI, METE, BETB, BETA, PROP, GLTD, META, SERA, GLTB, SERC, CYSK, ARGE, RHTC, TRPA, PROC, TRPB, HISQ, HISM	7.659E-28
KEGG	eum00260:Glycine, serine and threonine metabolism	9	5E-13	TRPB, ASD, GCVP, BETB, BETA, LPD, SERA, TRPA, SERC	2.509E-10
KEGG	ecg00330:Arginine and proline metabolism	8	3E-10	GDHA, ADIA, ARGE, ASPC, YBAS, SPEG, ASTC, PROC	1.124E-08
KEGG	ecw00270:Cysteine and methionine metabolism	8	2E-11	METK, CYSK, ASD, ASPC, LUXS, TYRB, METE, META	2.02E-09
KEGG	ecd00250:Alanine, aspartate and glutamate metabolism	7	1E-09	GDHA, GADB, GADC,ASPC, YBAS, GLTD, PURA, GLTB	4.309E-08
GO_BP	GO:0009067~aspartate family amino acid biosynthetic process	7	8E-08	ASD, ASPC, METE, HISQ, META, RHTC, HISM	1.652E-06
KEGG	ece00400:Phenylalanine, tyrosine and tryptophan biosynthesis	4	9E-05	TRPB, ASPC, TYRB, TRPA	0.0014603
KEGG	ece00290:Valine, leucine and isoleucine biosynthesis	3	0.0035	ACEE, ILVI, ILVE	0.0463703

DOWN					
Category	**Term**	**Count**	**PValue**	**Genes**	**Benjamini**
GO_BP	GO:0009309~amine biosynthetic process	38	1E-41	PUTP, TDCC, THRA, MMUM, THRB, THRC, ILVH, ILVC, ILVB, ASPA, SPEA, SPED, GLTJ, SPEE, METC, ASNA, ASNB, ARTI, MALY, GLNA, FLIY, ARTQ, PROB, PROA, TYRP, GLTP, ARTP, HISG, CARA, HISF, HISI, CARB, LYSP, HISH, HISC, PUTA, HISD, AVTA	2.152E-39
KEGG	eum00250:Alanine, aspartate and glutamate metabolism	14	7E-22	PURF, GLNA, ANSA, ANSB, PURB, YNEH, CARA, CARB, ASPA, GLMS, PYRB, PUTA, ASNA, ASNB	5.216E-19
KEGG	ecd00270:Cysteine and methionine metabolism	9	9E-12	THRA, MMUM, MALY, SPED, SPEE, SDAB, METC, SSEA, DCM	8.26E-10
KEGG	eum00260:Glycine, serine and threonine metabolism	8	1E-09	GARK, THRA, THRB, THRC, KBL, SDAB, GCVT, TDCB	5.346E-08
KEGG	ecq00340:Histidine metabolism	6	2E-09	HISG, HISF, HISI, HISH, HISC, HISD	6.566E-08
KEGG	ecg00330:Arginine and proline metabolism	8	7E-09	SPEA, PUTA, GLNA, SPED, SPEE, PROB, PROA, YNEH	2.476E-07
KEGG	ecq00290:Valine, leucine and isoleucine biosynthesis	5	1E-05	ILVB, AVTA, ILVH, TDCB, ILVC	0.0001924
SP_PIR	amino-acid transport	10	7E-15	LYSP, TDCC, PUTP, ARTI, GLTJ, ARTQ, YECS, TYRP, ARTP	4.121E-13

Figure 4.12 Genes coding for the amino acid biosynthetic and metabolic processes.

The genes encoding for important enzymes involved in amino acid metabolism and amino acid transporters were considered. After functional annotation of the up and down-regulated genes, it was not possible to identify a regulatory trend, since many amino acid metabolism pathways were found induced and repressed.

The increase of the significant amino acids could be caused by de-novo synthesis or external uptake. I looked at the genes of the main transporters: glutamate and the poly-amine putrescine transporters were up-regulated, valine and threonine transporter genes were down-regulated, while the genes for leucine and isoleucine transport were not differentially expressed.

I have then analysed in detail amino acid metabolism pathways linked to the metabolites detected by the NMR analysis (**Figures 4.13**, **4.14**, **4.15**, **4.16** and **4.17**).

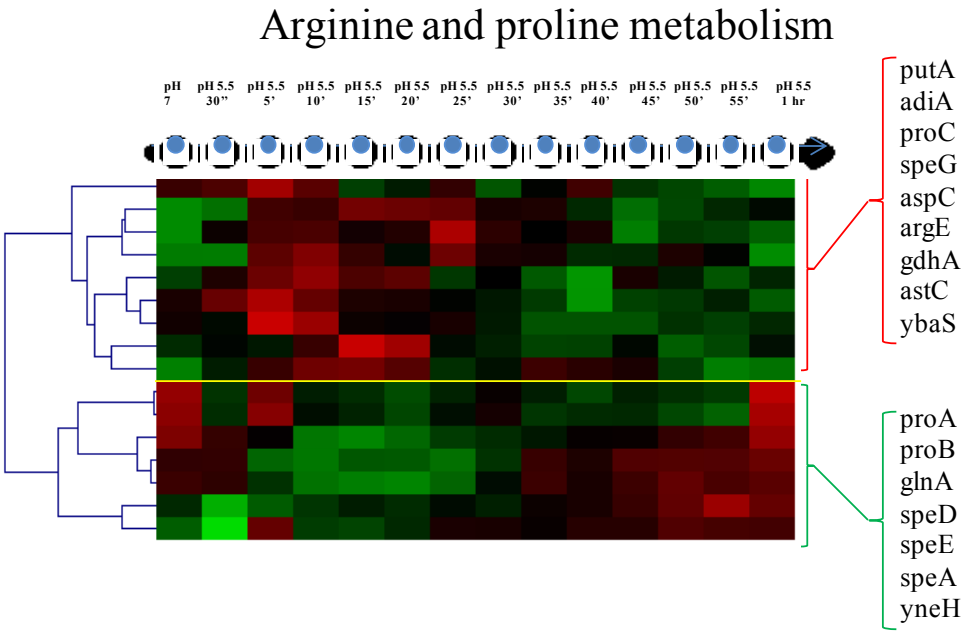


Figure 4.13 Regulation of the genes encoding for the arginine and proline metabolism.

The KEGG pathway in figure showed that a high percentage of the genes involved in arginine and proline metabolism were down-regulated (in green).

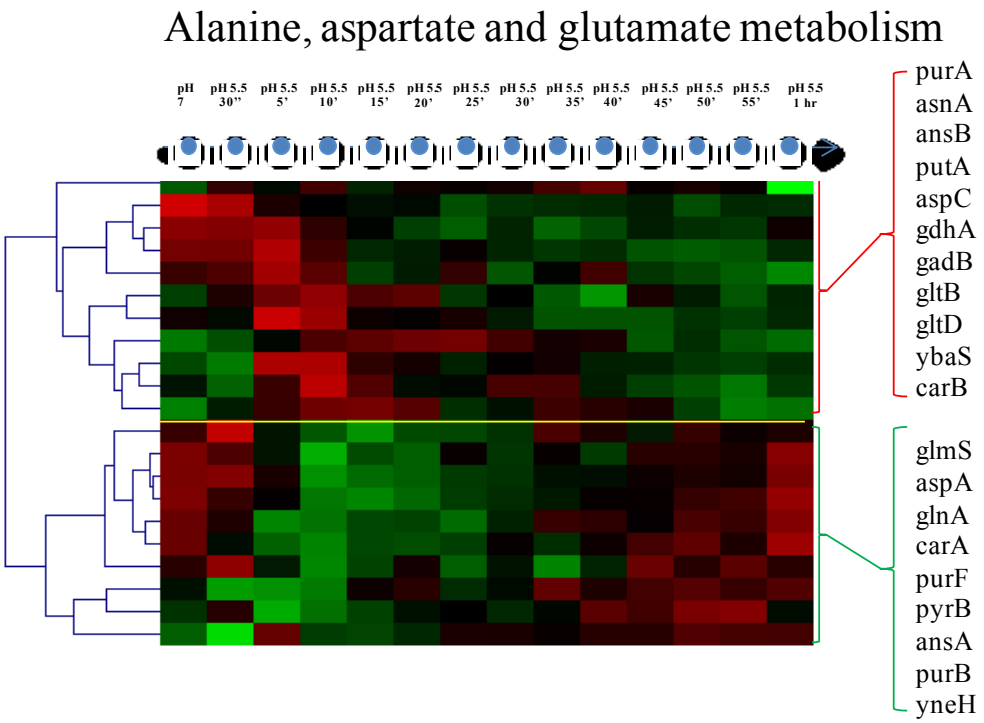


Figure 4.14 Regulation of the genes encoding for the alanine, aspartate and glutamate metabolism.

Alanine, aspartate and glutamate metabolism genes were mostly down-regulated (in green). However, the genes coding for the glutamate biosynthesis enzymes were found up-regulated (in red).

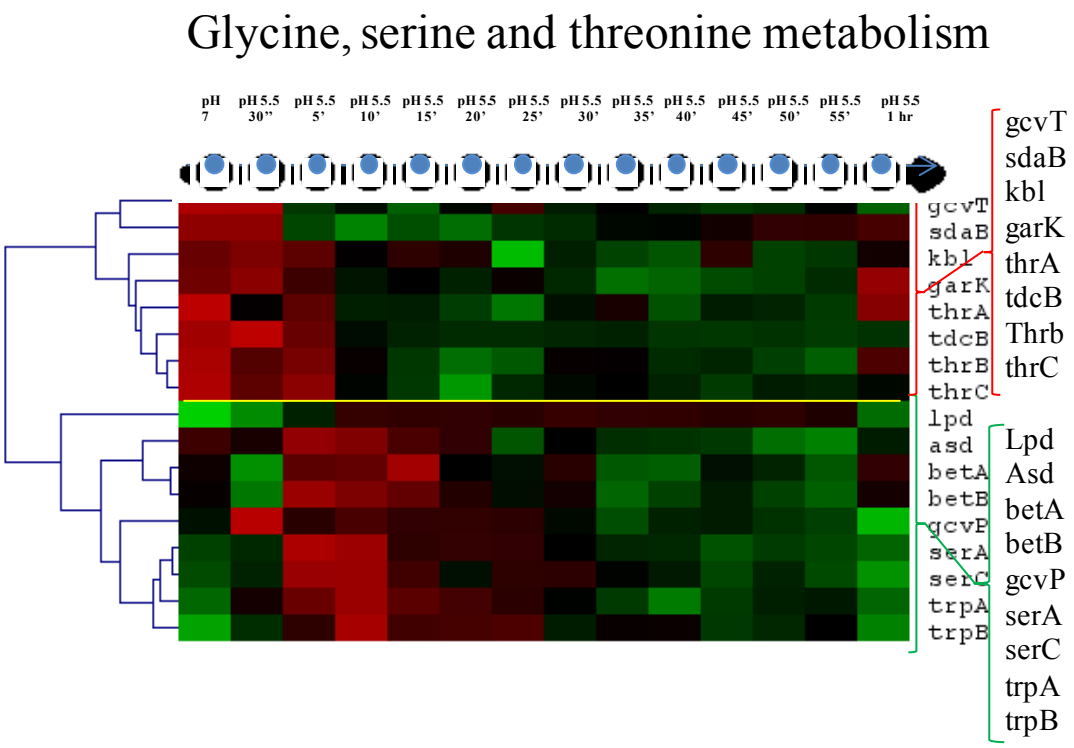


Figure 4.15 Regulation of the genes encoding for the glycine, serine and threonine metabolism.

Glycine, serine and threonine metabolism pathway was characterised by the down-regulation of a high percentage of genes (in green).

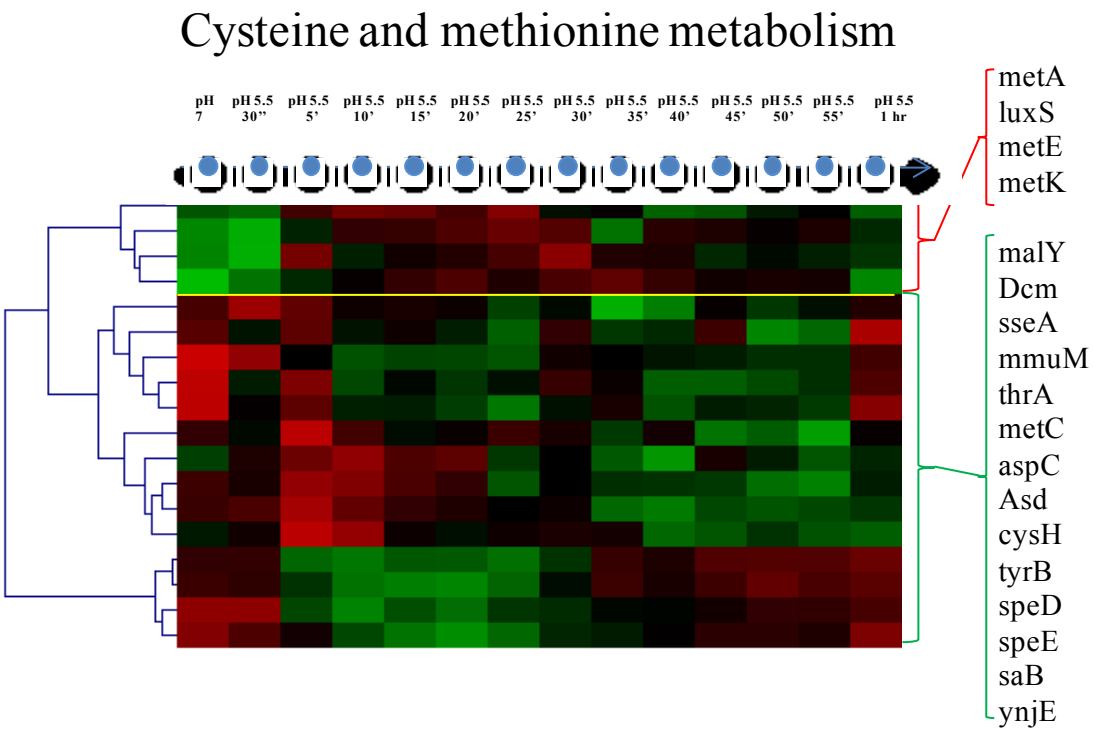


Figure 4.16 Regulation of the genes encoding for the cysteine and methionine metabolism
The genes involved in the cysteine and methionine pathway were mostly down-regulated in response to pH 5.5 (in green).

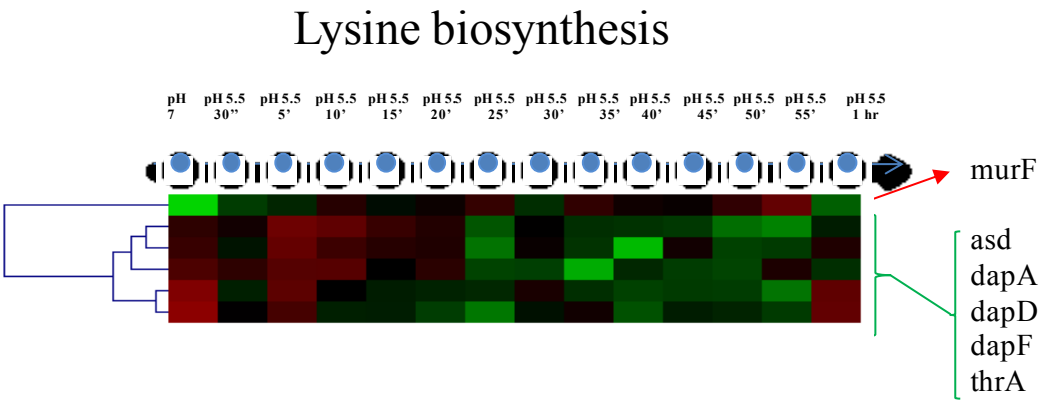


Figure 4.17 Regulation of the genes encoding for the lysine biosynthesis.

Only few genes were found differentially expressed in the lysine biosynthesis, however they were mostly repressed in response to acid exposure (in green).

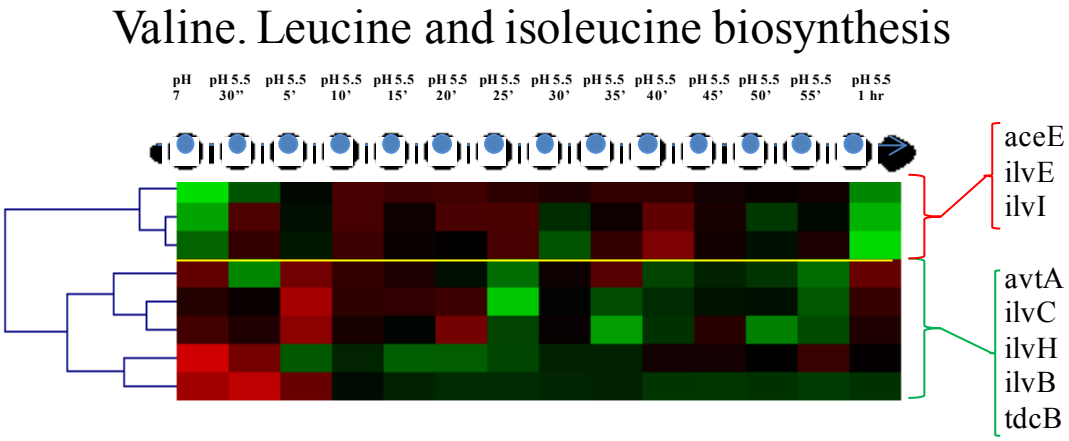


Figure 4.18 Regulation of the genes encoding for the valine, leucine and isoleucine metabolism.
The pathway of valine, leucine and isoleucine metabolism did not exhibit a clear trend since the genes differentially expressed belonging to this pathway were both up and down-regulated (in red and green respectively).

On the basis of these results, several trends of regulation of the amino acids pathways were found. For example genes in the cysteine and methionine metabolism pathway, the arginine and proline metabolism pathway and the lysine biosynthesis pathway were mostly down-regulated. The alanine, aspartate and glutamate pathway was also down-regulated, with the exceptions of the functions involved in the glutamate biosynthesis. The glutamate increase could be therefore considered as an effect of the transporters and biosynthesis encoding genes up-regulation. The glycine, serine and threonine metabolism pathway and the valine, leucine and isoleucine metabolism pathway were showing a mixed trend of regulation, in which several enzymes were both up and down-regulated. However, when looking at the expression of the genes involved in the degradation process of Valine, leucine and isoleucine, I found that most of them where induced during acid conditions (**Figure 4.19**).

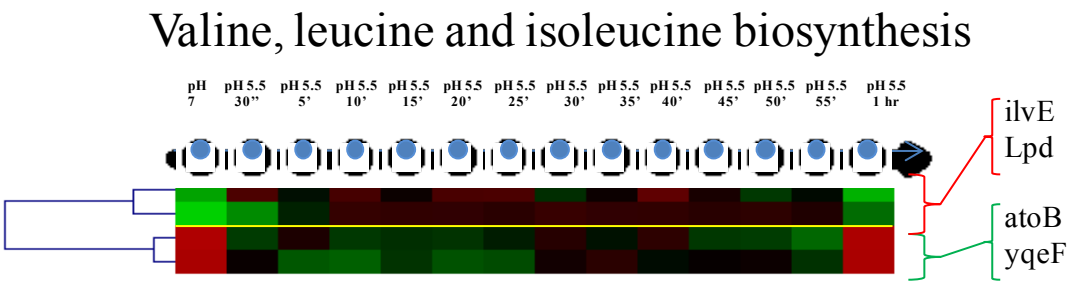


Figure 4.19 Regulation of the genes encoding for the valine, leucine and isoleucine degradation.
The pathway of valine, leucine and isoleucine degradation was instead characterised by the up-regulation of the majority of the genes differentially expressed (in red).

These results showed that, for some of the significant amino acids (threonine, leucine, valine and isoleucine), the increase could not be considered an effect of de-novo biosynthesis or external uptake, but could be considered a consequence of protein degradation. During stress conditions the probability of proteins misfolding is high (Mandelstam, 1958). Events such as proteolysis of damaged proteins to increase the amino acids pools is usually required, together with the chaperones involvement, in order to avoid misfolding events (Tomoyasu et al., 2001). Hence, I have focused my analysis on the understanding of the modulation of *E. coli* protease and chaperone genes during acid adaptation.

In order to address the question involving the increase of some amino acids, I have selected the genes coding for the major proteases in *E. coli*, all differentially expressed (**Figure 4.20**).

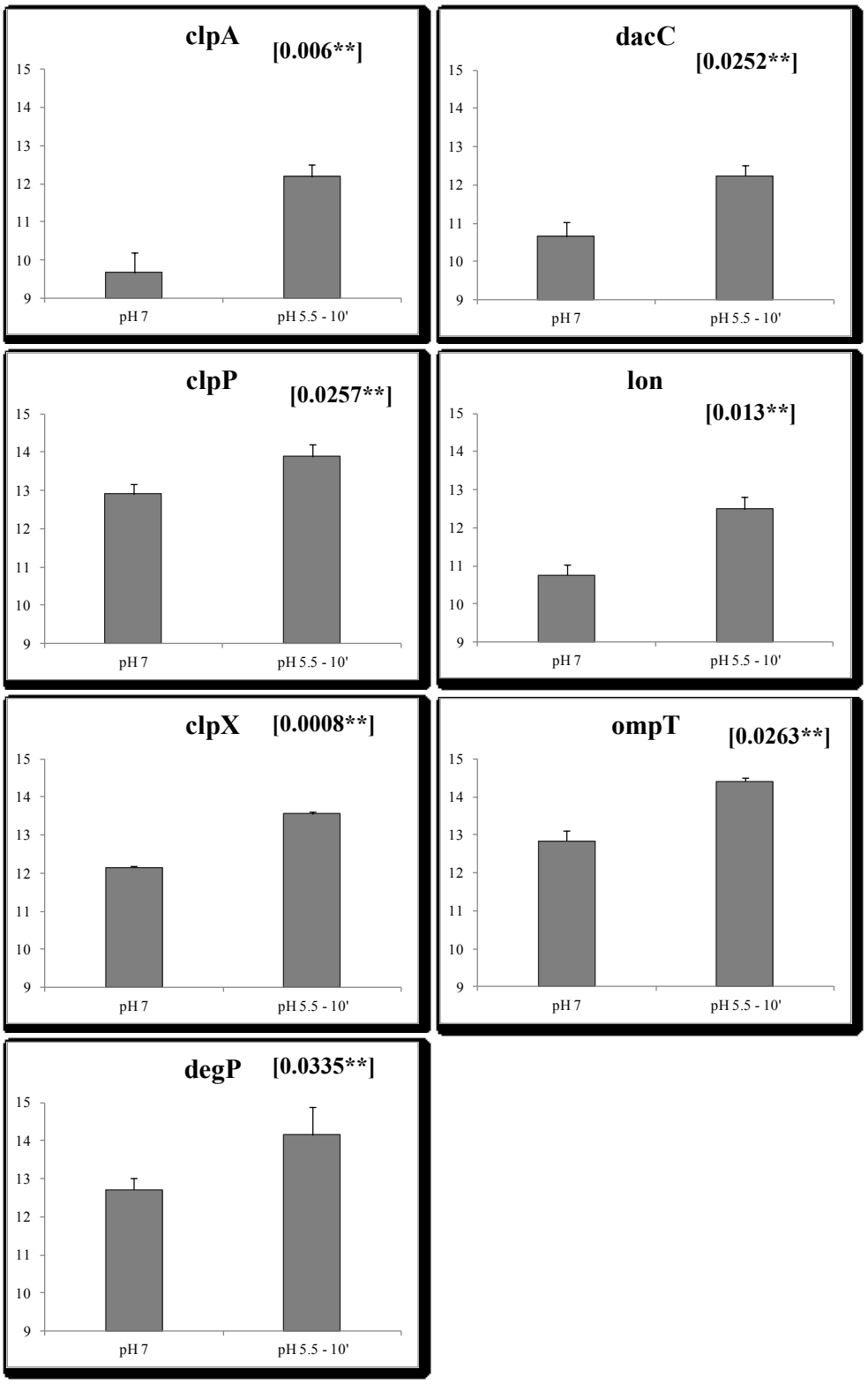


Figure 4.20 Genes coding for the major proteases differentially expressed.

The gene expression values are relative to pH 7 and pH 5.5 after 10 minutes. The error bars are related to standard deviation values on the three biological replicates.

To describe the modulation of the protease genes I have considered the time point 10 minutes from pH 5.5 exposure. At this stage, the selected genes were all significantly up-regulated in response to acid. The involvement of proteases in adverse conditions is already known; in acid stress response for example, a potential involvement of the Lon protease in the direct control of the GAD system was seen (Heuveling et al., 2008)

The protease pathway is not the only option for the cell to avoid the problem of misfolded proteins. In both prokaryotes and eukaryotes, chaperone molecules are involved in the process of protein folding, in order to reduce the possibility to get abnormal and unstable proteins (Liberek et al., 2008). In the time course data of the MG1655 strain, I have seen up-regulation of the genes of the major chaperones (**Figure 4.21**).

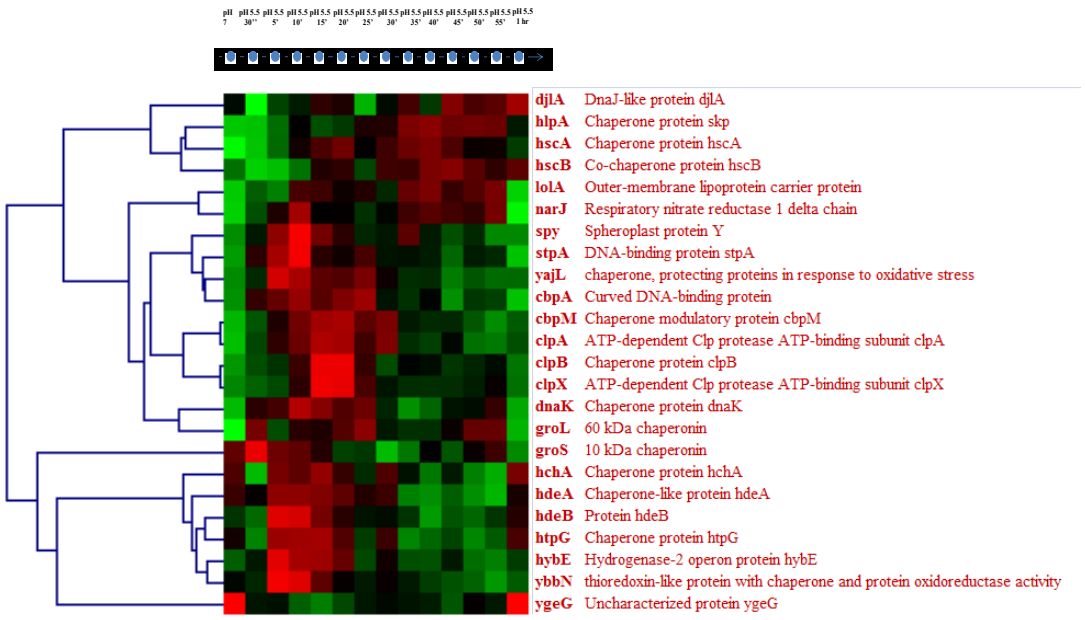


Figure 4.21 Genes coding for the major chaperones differentially expressed.

Most of the chaperone genes were positively regulated after 5-10 minutes exposure, few of them after 30 minutes. Only one gene was down-regulated, the uncharacterised YgeG protein.

4.3.3 A role for the protease LON in acid adaptation?

A recent study on Lon revealed that this protease has an important regulatory role in controlling the GAD AR system (Heuveling et al., 2008). More precisely, in this study it was observed that in a Δ*lon* mutant strain, several GAD genes (such *gadE*, *gadBC*, *gadA*) were up-regulated (Heuveling et al., 2008). In both wild type and mutant GadE was degraded in response to acid exposure (Heuveling et al., 2008); in the mutant the degradation was less evident when compared to the WT.

The observation that in a *lon* mutant background the GAD genes are up-regulated is interesting, but the effect of loss of this protease may be much wider than suggested by that paper and therefore changes in acid sensitivity may not be only dependent on the effects on the GAD system.

I set to test this hypothesis by performing a microarray analysis of the wild type MG1655 and *lon* mutant strains over 2 hours after shift to pH 5.5. Consistent with the original hypothesis I have found that the loss of Lon protease induced profound changes in the transcriptional state of *E. coli*. I could identify 2336 genes (corresponding to more than 40% of the *E. coli* genome) differentially expressed at 1% FDR. I first analyzed the overall response by using the principal component analysis (**Figure 4.22**)

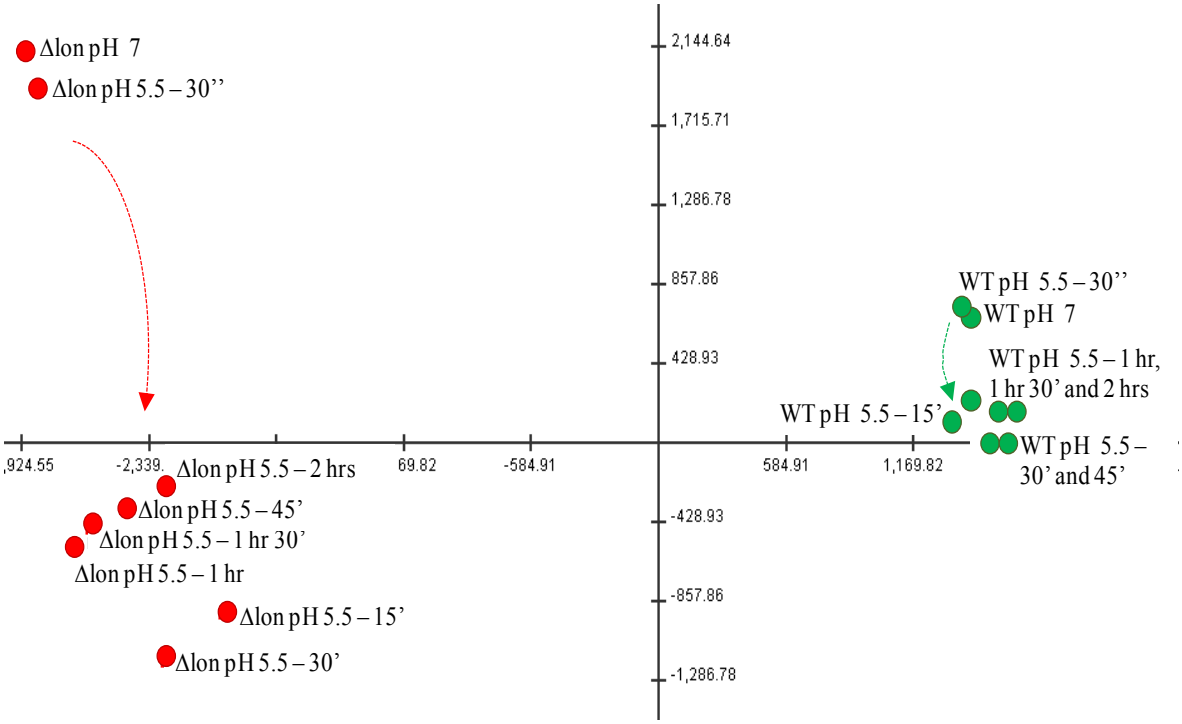


Figure 4.22 PCA plot of the MG1655 and Δ*lon* strains at pH 7 and in a 2 hours time course experiment at pH 5.5.

This PC plot shows the dynamics of change in the transcriptional response of Δ*lon* (red dots) and MG1655 (green dots) during 2 hours acid adaptation time course. The x and y axes represent respectively the first and second principal components. For the analysis were considered the 2336 differentially expressed genes. The difference between mutant and WT is the most evident aspect of these results, represented by the shift on the first component. Δ*lon* response to acid was more enhanced when compared to the wild type (shift across the second component).

The most evident aspect seen in the PCA plot was the difference between the two strains: the shift across the first component indicated differences between wild type and mutant, which is independent from the pH change. Moreover, changes across the second component of the plot represented the response to acid exposure. I have seen that the response to pH 5.5 in Δ*lon* was more enhanced when compared to the wild type.

We first set to repeat the observations published by Heuveling et al. (2008) and analysed the expression of genes involved in the ARs (**Figure 4.23**).

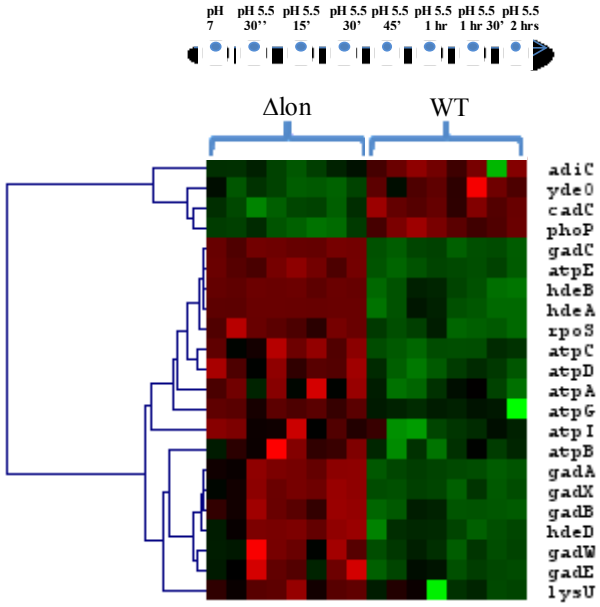


Figure 4.23 ARs gene selection from the time course experiments in the LON protease gene KO and the wild type.

After the normalization of both Δ*lon* and WT data I have selected the genes encoding the ARs (23 genes). The majority of the genes coding for the ARs were highly expressed in the LON compared to the wild type, with few exceptions.

Consistent with what was reported previously, *lon* mutant strain was characterized by increased expression of most of the ARs genes when compared to the wild type, with only few exceptions (**Figure 4.23**): *adiC*, *ydeO*, *cadC* and *phoP*.

Interestingly, most of the ARs genes in Δ*lon* were transcriptionally regulated in response to acid in the same direction than the wild type (**Figure 4.24**). The only exceptions were the genes *lysU*, down-regulated in the WT and induced in the mutant, and *phoP*, which vice versa was up-regulated in WT and oppositely modulated in Δ*lon* (**Figure 4.24**).

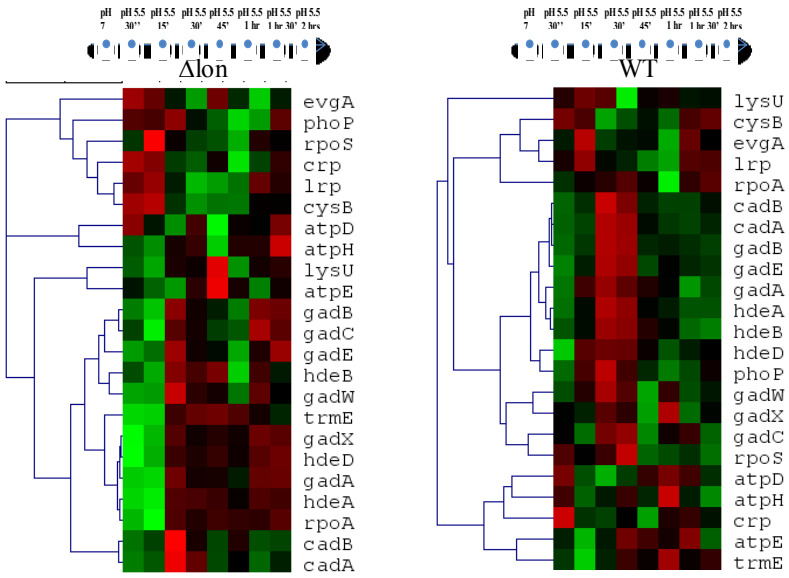
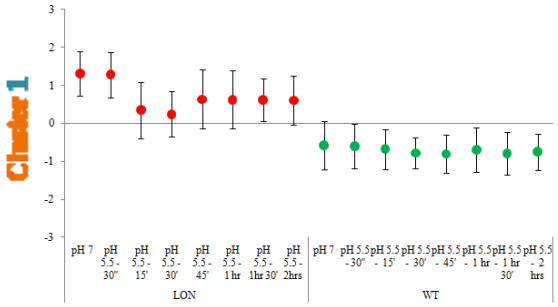


Figure 4.24 ARs gene selection from the time course experiments in the LON protease gene KO and the wild type.

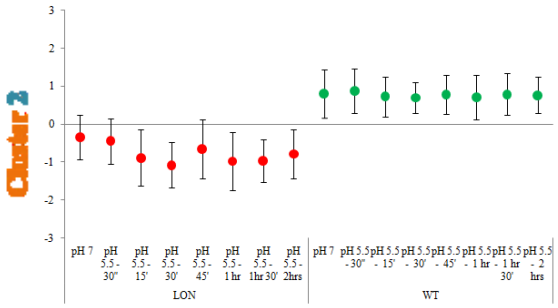
In this figure, the genes were selected from the datasets of both WT and mutant with single normalization. Most of the GAD genes were induced in both strains and the *atp* genes were instead down-regulated. The trend followed by the ARs genes was similar between the two strains, with few exceptions: *phoP* and *lysU*.

Having analysed the response of the genes involved in the ARs I then set to perform a broader analysis of the genome–wide transcriptional state of the wild type and mutant strains. Key drive for this analysis was to identify the molecular functions affected by the Lon mutation and to assess to what extent the general response to acid exposure may be similar in the two strains.

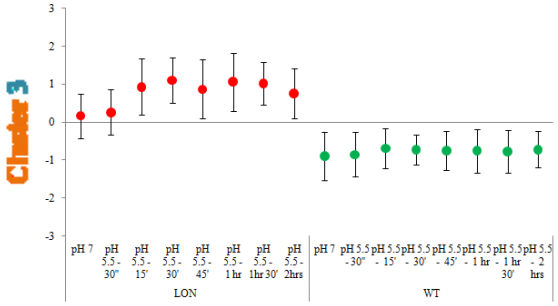
I addressed these questions by performing a cluster analysis on the differentially expressed genes, in order to identify the functions modulated in the two strains and in response to acid, obtaining four clusters (**Figure 4.25**).



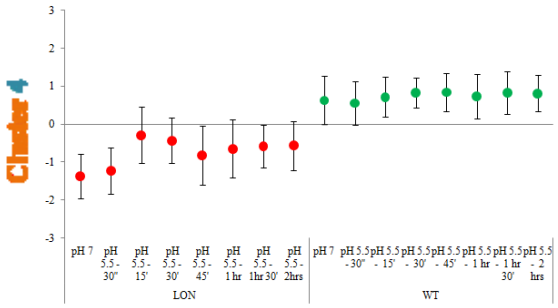
Category	Term	Count	PValue	Benjamini
GO_BP	GO:0009060~aerobic respiration	27	5.14E-15	1.30E-12
Kegg	eco00020:Citrate cycle (TCA cycle)	13	7.67E-11	1.12E-08
Kegg	eco02010:ABC transporters	32	3.13E-14	5.01E-11
Kegg	eco00620:Pyruvate metabolism	14	1.20E-09	1.06E-07
Kegg	eco00630:Glyoxylate and dicarboxylate metabolism	13	6.94E-10	6.93E-08
Kegg	eco00520:Amino sugar and nucleotide sugar metabolism	12	2.07E-07	6.62E-06
Kegg	eco02060:Phosphotransferase system (PTS)	13	8.03E-09	4.76E-07
Kegg	eco00052:Galactose metabolism	11	1.02E-07	3.70E-06
Kegg	eco00010:Glycolysis / Gluconeogenesis	12	2.08E-08	1.07E-06
Kegg	eco02020:Two-component system	20	1.41E-07	4.91E-06
Kegg	eco00030:Pentose phosphate pathway	9	5.90E-06	1.07E-04
Kegg	eum00480:Glutathione metabolism	6	1.58E-04	0.0021247
Kegg	eco00260:Glycine, serine and threonine metabolism	8	7.42E-05	0.0010592
Kegg	eco00561:Glycerolipid metabolism	5	2.57E-04	0.0033416
Kegg	eco00330:Arginine and proline metabolism	10	1.32E-05	2.23E-04
Kegg	eco00564:Glycerophospholipid metabolism	7	1.72E-04	0.002291
Kegg	eco00500:Starch and sucrose metabolism	6	0.00253	0.0246743



Category	Term	Count	PValue	Benjamini
Kegg	eco02040:Flagellar assembly	13	1.51E-13	1.49E-10
GO_BP	GO:0009061~anaerobic respiration	14	2.52E-06	2.50E-04
Kegg	eco02020:Two-component system	13	1.09E-06	1.08E-04
Kegg	eco02010:ABC transporters	14	8.31E-06	5.13E-04
Kegg	eco02030:Bacterial chemotaxis	5	2.62E-04	0.0066268
GO_BP	GO:0006350~transcription	46	0.00764	0.1190721
Kegg	eco00540:Lipopolysaccharide biosynthesis	6	8.19E-05	0.0032316
Kegg	eco00500:Starch and sucrose metabolism	5	0.00141	0.0259307
Kegg	eco00360:Phenylalanine metabolism	4	0.00156	0.0282408
Kegg	eco00190:Oxidative phosphorylation	6	5.59E-04	0.0116769



Category	Term	Count	PValue	Benjamini
Kegg	eco03010:Ribosome	37	4.83E-38	7.68E-35
Kegg	eco00970:Aminoacyl-tRNA biosynthesis	13	3.74E-11	3.96E-09
GO_BP	GO:0007049~cell cycle	15	4.95E-05	0.0019512
Kegg	eco00620:Pyruvate metabolism	12	3.38E-07	1.41E-05
Kegg	eco00230:Purine metabolism	20	2.47E-11	2.80E-09
Kegg	eco00240:Pyrimidine metabolism	8	0.00307	0.0382951
Kegg	eco02010:ABC transporters	20	2.78E-05	6.90E-04
Kegg	eco00030:Pentose phosphate pathway	8	5.86E-05	0.0013489
Kegg	eco00010:Glycolysis / Gluconeogenesis	11	8.82E-07	3.33E-05
Kegg	eco00540:Lipopolysaccharide biosynthesis	9	2.22E-06	7.65E-05
Kegg	eco00550:Peptidoglycan biosynthesis	6	0.00139	0.0204304



Category	Term	Count	PValue	Benjamini
Kegg	eco02010:ABC transporters	25	1.35E-11	1.81E-08
GO_BP	GO:0009061~anaerobic respiration	21	5.59E-10	3.59E-07
Kegg	eco02020:Two-component system	17	5.59E-08	7.51E-06
SP_PIR	lipopolysaccharide biosynthesis	16	6.78E-10	8.10E-09
GO_BP	GO:0016051~carbohydrate biosynthetic process	34	4.78E-06	3.42E-04
GO_BP	GO:0006350~transcription	48	0.26663	0.7329991
GO_BP	GO:0006023~aminoglycan biosynthetic process	13	3.54E-07	7.59E-05
GO_BP	GO:0006024~glycosaminoglycan biosynthetic process	13	3.54E-07	7.59E-05
GO_BP	GO:0009252~peptidoglycan biosynthetic process	13	3.54E-07	7.59E-05
Kegg	eco00550:Peptidoglycan biosynthesis	5	0.00163	0.0352678
Kegg	eco00240:Pyrimidine metabolism	9	3.88E-05	0.0013017
Kegg	eco00630:Glyoxylate and dicarboxylate metabolism	7	1.70E-04	0.0050647

Figure 4.25 Cluster analysis of the genes differentially expressed between the two strains.

The cluster analysis identified 4 clusters pointing out the main differences between the two strains. Clusters 1 and 3 were defined by the genes differentially expressed which were up-

regulated in the mutant. Clusters 2 and 4 were characterised by the Δ*lon* genes down-regulated when compared to the WT.

The 4 clusters showed the difference within the two strains; in addition, they showed the functions which significantly changed in Δ*lon* in response to acid exposure. Some of the aerobic pathways, such as pyruvate metabolism (KEGGeco00620), pentose phosphate pathway (KEGGeco00030) were up-regulated in the mutant when compared to the WT (**Figure 4.26**).

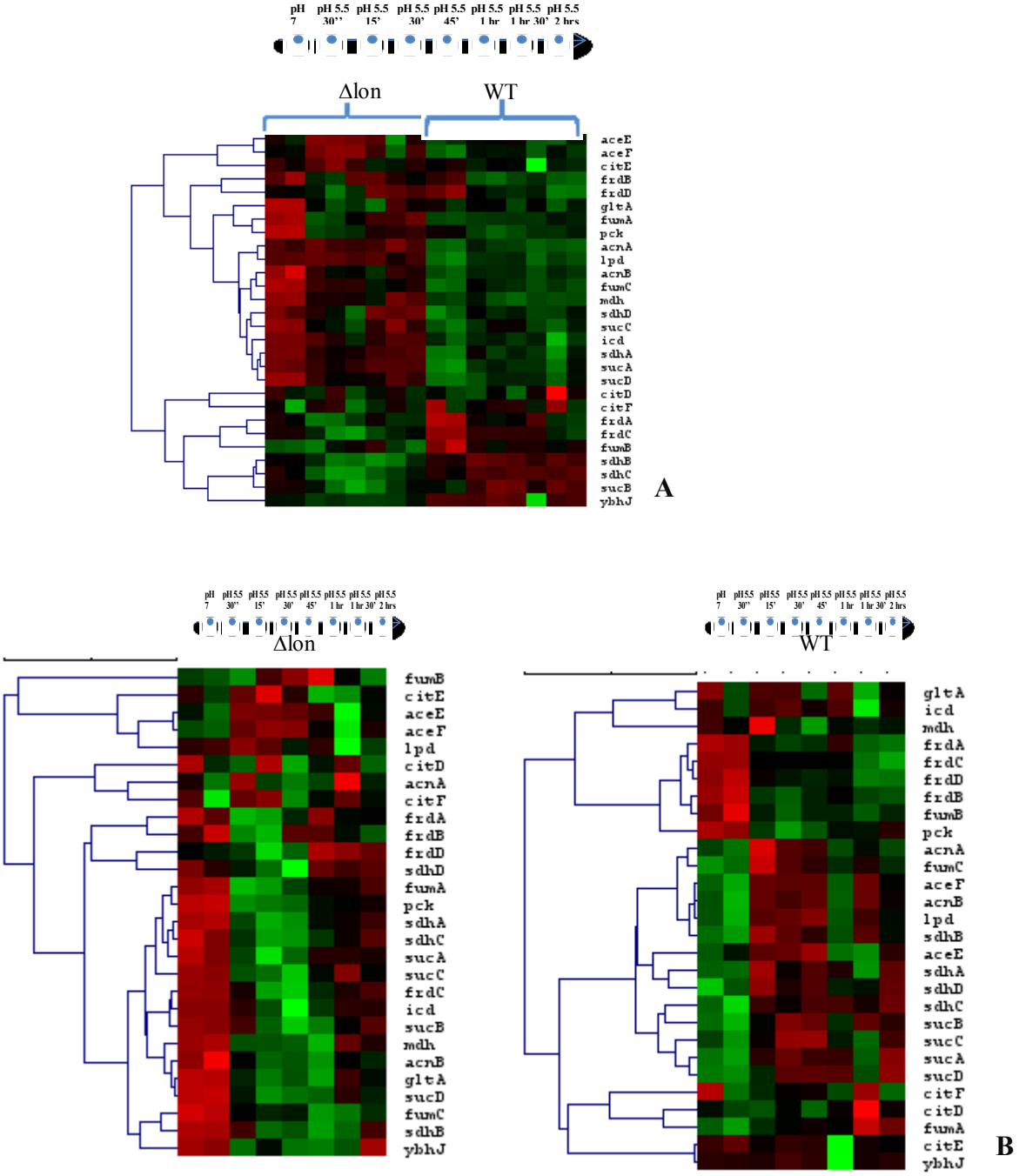


Figure 4.26 TCA cycle differentially expressed genes between the Lon mutant strain and the Wild Type

Many aerobic functions were up-regulated in the Lon mutant when compared to the WT. Δ*lon* was showing increased expression of most of the genes encoding for important enzymes of the TCA cycle (panel A), except for the genes encoding the fumarate reductase and some of the genes coding for the succinate dehydrogenase. Different behaviour was exhibited by the two strains in response to acid: many of the genes in Δ*lon* were down-regulated after exposure, while in the MG1655 strain they were mostly induced at pH 5.5 (Panel B).

The flagellar assembly (KEGGeco02040) was down-regulated in Δ*lon*, not only in response to acid but also when compared to the wild type; a different behaviour was instead found for the genes encoding the ribosome pathway (KEGGeco03010). Interestingly, all the genes coding for the anaerobic respiration pathways (GO:0009061) were significantly down-regulated in the mutant in relation to the wild type, however positively and negatively regulated in response to the pH change.

In order to evaluate the response of the *Δlon* strain to acid adaptation, a cluster analysis was performed, identifying 4 clusters which trends were defined by up and down-regulation of the genes in response to acid (**Figure 4.27**)

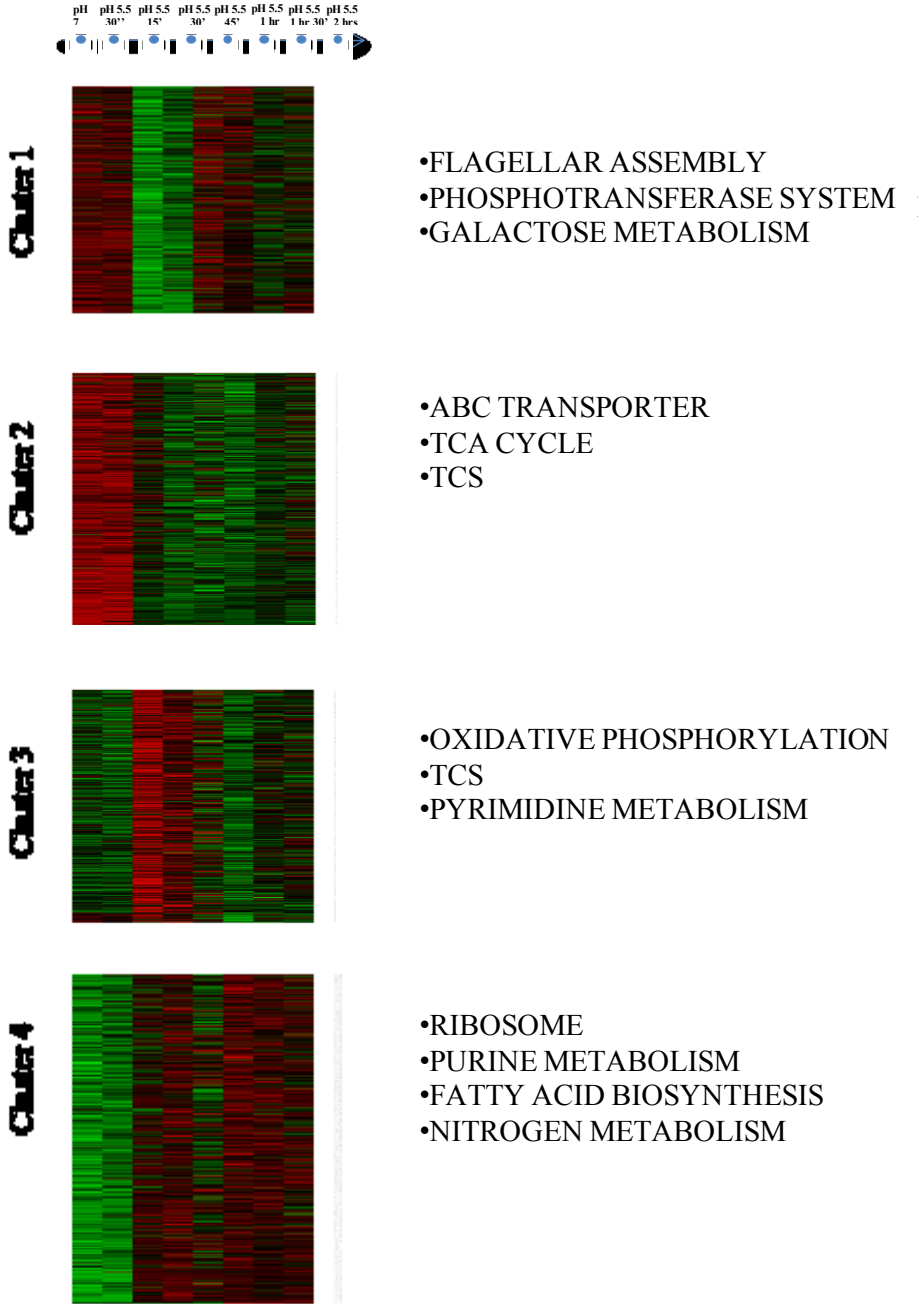


Figure 4.27 Cluster analysis of *Δlon* time course data.

This analysis was performed on the genes differentially expressed between the WT and Δ*lon*. The analysis revealed 4 clusters modulated in response to acid: cluster 1 and 2 were down-regulated, while cluster 3 and 4 were up-regulated after acid exposure. The functional annotation analysis revealed that flagellar assembly and ABC transporters were strongly repressed, while the ribosome encoding genes and the purine and pyrimidine metabolism were highly induced at pH 5.5. It was not possible to identify a trend of regulation for the metabolic pathways.

4.4 Discussion

4.4.1 Acid adaptation in MG1655 is characterized by up-regulation of several aerobic enzymes and a considerable increase of amino acids

In order to respond to stress conditions, all the microorganisms have developed adaptation mechanisms to facilitate survival. Although each environmental stress represents a specific challenge some aspects of stress response can be general. The ability to adjust the energy balance is one of these features. In contrast with the response seen in the BW25113 strain, MG1655 showed an increased expression of the aerobic functions at pH 5.5. The result is surprising, since the two strains, which derive from the same ancestor, differ for few sites in the genome sequence (Baba et al., 2006, Hayashi et al., 2006).

The results obtained from the time course data on the MG1655 strain did not show trends comparable to previous works in consideration of different environmental conditions. Carbon starvation and oxidative stress induced in *E. coli*, during logarithmic phase, the same profile of response: inactivation of the metabolic pathways involved in the process of the conservation of energy (Weber et al., 2005, Jozefczuk et al.). Most of the times, those stress conditions also involve the down-regulation of the aerobic functions genes, especially during oxidative stress (Chang et al., 1999) (Chang et al., 2002; Nystrom et al., 2005). In my data, I have seen up-regulation of TCA cycle, glycolysis and oxidative phosphorylation, and consequent inhibition of the anaerobic components. Some features of acid response could be compared to other mechanisms of regulation seen in previous works. For instance, I have seen the down-regulation of the genes coding for the ribosome for the first 30 minutes of acid exposure (**Figure 4.7**), therefore a decrease of de-novo synthesis of proteins, which is typical of the stress response (Weber et al., 2005).

Moreover, the analysis performed on the metabolomics data showed the increase of some of the essential amino acids, an aspect in common with other works; it has been reported that the increase of the amino acid levels could be associated with protein degradation (Mandelstam, 1958). Any cause of stress for the cells requires new proteins to promptly respond to the new condition; therefore the accumulation of those amino acids could be associated to the cells requirements. In fact, after 30 minutes of exposure at pH 5.5, the ribosome genes were gradually up-regulated, therefore increasing the translational pathway, probably using the amino acids accumulated during acid response. Additionally, the genes coding for important chaperones differentially expressed in my data, were mostly up-regulated. The cells which are not in their physiological state require the chaperones for the correct folding of the stress response proteins but also for the disaggregation and reactivation of aggregated proteins (Liberek et al., 2008). The increase in amino acids was also found in previous works based on metabolomics data, one of them in particular was focused on different stress conditions. In this work the analysis was performed in heat, cold and oxidative stress conditions and during carbon starvation. The researchers found an increase of several essential amino acids, such as arginine, threonine, leucine, valine (Jozefzuck et al., 2010). This increase was attributed to the stress conditions, as increased activity of the cellular proteolytic activity. The transporters, together with the biosynthesis pathways, are the easiest way to supply the cells with the amino acids; in my experiments *E. coli* MG1655 exhibited a decreased expression of the genes coding for the amino acid biosynthesis and transporters, except for the ones involved in mechanisms of acid resistance (i.e. GadC and AdiC). Furthermore, many amino acids metabolism pathways were not induced at pH 5.5. Consequently, most of the amino acids accumulated could probably be considered a result of protein degradation. The potential role of other amino

acids (such threonine, valine etc.), besides the known in acid response, could involve regulatory mechanisms of acid resistance. For example, it would be possible that other ARs, not described yet, could be involved. Moreover, the events of proteolysis happening inside the cells to increase the amino acid pools can require several proteases: one of those is known to have a role for the degradation of GadE (Heuveling et al., 2008).

The increased expression of the genes encoding the major proteases in *E. coli* proved that those molecules are essential for stress responses (Gottesman, 1996). Furthermore, RpoS is stabilised when the cell enter into the stationary phase (Hengge, 2009) by the complex CplP/X (Takayanagi et al., 1994; Schweder et al., 1996) because of its short life.. At the beginning of the stationary phase, the short life rapidly increases to 30 minutes, therefore becoming more stable (Schweder et al., 1996). It has been observed that the regulation at the beginning of the stationary phase of RpoS is exerted by the involvement of the RssB two-component system, which is likely to promote the activation of CplP/X (Hengge et al., 2009). For all those reasons, the involvement of the proteases in acid response (or in general, for stress responses), is probably essential for the amino acid supply and for the stabilization of the RpoS stress factor.

Another aspect to be considered is the reduction of the osmoprotectant glycine betaine. This molecule is known to be important for osmotic stress (Landfald and Strøm, 1986), from its precursor choline. In the previously considered strain, BW25113, I have seen that during acid adaptation the cells could experience situations similar to hypo-osmotic stress conditions. However, the contribution of this osmolite at acid pHs is not known yet, further studies will be required for the comprehension of the overlapping of the two stress conditions.

4.4.2 Lon in control of acid resistance: only a GadE protease?

Lon is one of the major proteases in *E. coli* and many microorganisms (Heuveling et al., 2008), but is also an important mitochondrial protease in humans (Fu and Morkovitz, 1998). Lon can degrade many transcription factors (Tsilibaris et al., 2006), but is not directly involved in the process of the stress σ factor degradation; however it has been seen that some of the RpoS downstream factors (i.e. GAD system regulators) could be degraded by Lon (Heuveling et al., 2008). The results obtained until now, with a microarray based approach, showed that in a Δ*lon* context, *E. coli* MG1655 not only has a different transcriptomic state, but exhibited a stronger response to acid exposure (**Figure 4.15**). The changes caused by the *lon* mutation are extremely wide and involve a broad spectrum of variation in many pathways when compared to the wild type. Lon can directly degrade the GadE regulator; hence its role in acid response could be extremely important. Nevertheless, I have seen that the role of Lon is not restricted to a simple degradation of the transcription factors essential for acid response. Possibly, the protease could affect more regulators and more stress proteins, thus changing completely the molecular structure in the bacteria. Most of the changes happening in Δ*lon* (when compared to the WT) involved all the energetic and translational pathways analysed until now (**Figure 4.28**).

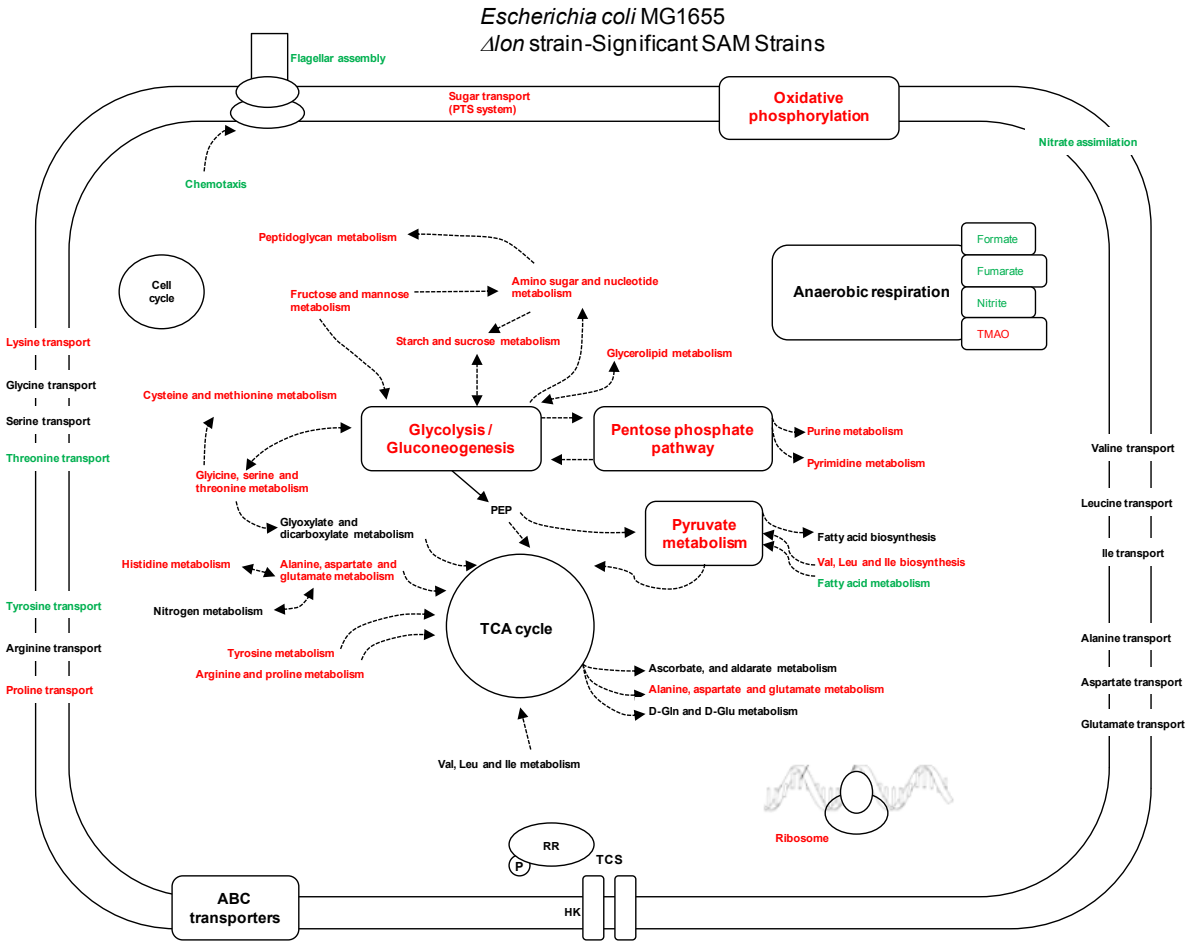


Figure 4.28 Differences between the MG1655 strain and Δ*lon*.

Most of the metabolic pathways were up-regulated in Δ*lon* when compared to the WT strain (red). However, many genes encoding anaerobic functions, such fumarate, formate and nitrate involving processes, were down-regulated in the mutant (green).

The figure was obtained considering the most important metabolic and energetic pathways in *E. coli* and their regulation during acid adaptation in the mutant *lon* compared to the WT strain. It was unmistakable that most of the functions are up-regulated in the gene KO, except for the anaerobic respiration pathways.

The aerobic pathways together with amino acid metabolism and biosynthesis were highly up-regulated when compared to the wild type, except for the function involved in the anaerobic respiration, which employs formate, fumarate and nitrate as final electron

acceptors. Since the expression of the ARs significantly changed in response to the mutation, in fact was strongly acid induced, probably it is not enough to acknowledge Lon as a protease of the Gad system. My hypothesis is that Lon (with other proteases) has probably more targets than the ones known until now and probably some of those targets are essential regulators in *E. coli*. Thus, the control of the acid resistance genes could be not the only aspect which defines Lon as an acid resistance factor.

4.5 Conclusions.

This chapter studied MG1655 during acid adaptation. The regulation of the important cellular functions was much different when compared to the BW25113 strain. In the next chapter I will use two gene models for the identification of regulators potentially important for acid adaptation in this strain.

Chapter 5: Dynamical modelling of *E. coli* MG1655 acid response by using a reverse engineering approach

5.1 Introduction

In Chapter 3, I have described the application of ARACNE to unravel the regulatory interactions underlying transcriptional response of the *E. coli* strain BW25113 during acid adaptation. The approach relied on a compendium of microarray data generated using strains mutated in genes modulated during acid response. Although the strategy has been indeed very successful, it is based on observed gene-gene correlations across several steady state snapshots and therefore cannot reveal the dynamics of the regulatory events (Hecker et al., 2009). However, dynamical modelling of gene regulatory networks can be achieved with a number of methodologies (Ortega et al., 2008) and naturally it relies on the availability of high resolution time course datasets. Moreover, even though reverse engineering gene regulatory networks from gene expression data is effective, these methodologies may also be used to integrate data from different sources (i.e. transcriptomics, proteomics, metabolomics, physiology measurements and published literature) with better results in terms of understanding biologically relevant regulatory processes(Hecker et al., 2009, Gupta et al., 2011).

Several approaches are currently used for inferring dynamical models of gene regulatory networks. These are information-theoretic approaches such as time-delay ARACNE (Zoppoli et al., 2010), Bayesian Networks (Perrin et al., 2003) and differential equations-based methods, such as ODEs (De Jong et al., 2004). Even though “omics” technologies

allow measuring the expressions of tens of thousands of genes, proteins and metabolites in a single experiment, they cannot measure all the factors that contribute to the regulatory interactions. In order to overcome this problem other formalisms, such as state-space models (SSMs) can be used. These consider hidden variables, which model biological processes that have not been measured (Rangel et al., 2004). Unfortunately, these methods are not suitable to model a large number of variables, at least not with the number of time points that can be realistically measured with current technologies.

In order to build a genome-level model for acid response there was a need of a methodology that could reliably infer networks with thousands of molecular components from a relatively low resolution time course. This was achieved by using a simplification of the methods originally developed in my group (Rangel et al., 2004), which was proposed by Hirose et al. (Hirose et al., 2008). This method can analyse time course data considering the aggregation of gene expression profiles and the temporal gene networks, at the module level. SSM have been used for statistical inference of transcriptional module-based gene networks, applied on replicates of time course gene expression profiles (Hirose et al., 2008). The length of time course data is usually a limiting factor to infer gene regulatory networks. Hirose et al. proposed to explore genetic networks of transcriptional modules; the modules are sets of genes involved in the same pathway (Hirose et al., 2008).

The model can identify potential transcriptional modules and map them into gene-level networks (Hirose et al., 2008). The assumption of the SSM is that the dynamic of observed data is regulated by the time evolution of few latent factors. The model is built on two matrices; one of them takes into account the relationships of genes, while the second one considers the hidden state vectors.

Here I describe the analysis of a model based on the SSM framework developed by Hirose (Hirose et al., 2008)integrating both transcriptional and metabolic response to acid adaptation. The conceptual integration of the dynamical model with an ARACNE network inferred from a large compendium of gene perturbation experiments (Faith et al., 2007) identified TCS regulators as the most upstream regulatory components. This hypothesis was successfully validated against EcID (Leon et al., 2009), a database for *E. coli* gene interactions.

5.2 Materials and Methods

5.2.1 Datasets for the analysis

SSM model: the dataset considered was obtained from the time course experiment during acid conditions in the *E. coli* MG1655 strain, as described in Chapter 4. The time course consisted in 14 time points collected every 5 minutes during pH 5.5 exposure. Transcriptomic and metabolomics experiments were considered. Microarrays data were quantile normalised and de-noised by a) removing data where SD/mean was more than 0.9 and b) removing data where maximum–minimum was less than 1.5. The noise level for each metabolomics NMR spectrum was estimated by dividing the spectrum into 32 regions and calculating the smallest bin SD for each region and multiplying this by 3. These results were used to de-noise the data. Data from metabolomics and transcriptomics were then combined. Normalised combined microarray and metabolomics data were input to Genespring GX 7.3.1 (Agilent Technologies, Santa Clara, CA, USA). Statistically significantly changing genes were found by 1-way ANOVA with a multiple testing correction for a false discovery rate (FDR) <0.05, and with Welch T-tests employing the same FDR. Fold change cut offs of 1.5-fold were additionally applied. A classification

algorithm was used to compare previous data with the current data; this employed the Support Vector Machines algorithm within Genespring with the Kernel Function Polynomial Dot Product (Order 3), a Diagonal Scaling Factor of 0, for all genes passing QC cut-offs in both experiments. Gene ontology (GO) analyses were carried out within Blast2GO employing the GOSSIP package (Williams et al., 2011).

ARACNE model: the dataset consisted of 907 new and previously published *E. coli* Affymetrix Antisense2 microarray expression profiles already normalised, collected under various conditions including pH changes, growth phases, antibiotics, heat shock, different media, varying oxygen concentrations, and numerous genetic perturbations (Faith et al., 2008). The dataset was previously used for assessing the genome scale performance of the context likelihood of relatedness (CLR) algorithm (Faith et al., 2007).

5.2.2 ARACNE gene inference

The gene network from steady state data was obtained by using the software application ARACNE (Margolin et al., 2006 - b), considering a compendium of 907 Affymetrix arrays data representing (Faith et al., 2007). ARACNE is an algorithm used for the identification of molecular networks from observational data and is based on the mutual information (Margolin et al., 2006 - a). The algorithm is based on the assumption that the expression level of a gene can be considered as a random variable. The edges identified by the algorithm are described as irreducible statistical dependencies between gene expression profiles. The candidate interactions are identified by the estimation of pair wise gene expression profile mutual information, using the Gaussian Kernel estimator. The algorithm performs ranking of the gene expression profiles before estimating mutual information values. This step is required to decrease the arbitrary transformations in the data

processing. In order to eliminate non direct interactions the data processing inequality principle (DPI) was applied. When considering a triplet of interaction, the DPI can remove the edge with the smallest mutual information value.

Significant interactions were defined by a P-value threshold of $P<10^{-18}$. In order to eliminate non-direct interactions, I used the inequality principle as implemented in ARACNE with a DPI of 0.1 (Margolin et al., 2006 - b). The resulting network was visualized using the Cytoscape software application (Shannon et al., 2002). The network consisted of 4118 nodes and 17275 edges, with MI values between 0.064 and 0.675. The network was then further tresholded, considering MI values greater than 0.15, therefore obtaining a new matrix of 3113 nodes and 8503 edges. Selected hubs were validated by using the web based tool Hubba (Lin et al., 2008), with the Degree topology-based algorithm scoring method (Jeong et al., 2001).

5.2.3 State Space Model

In order to develop a genome-wide dynamical model of acid response a computational method was used, originally developed by Hirose and colleagues (Hirose et al., 2008). This method is able to infer genome-wide molecular networks underlying complex biological processes. In order to overcome the complexity of reconstructing the interaction between thousands of molecular components this algorithm incorporate the concept of network modularization, where each module represents a number of genes with similar expression profiles.

A general SSM is supported by the concept that the dynamical behaviour of the observed data is based on the time evolution of few hidden transcription factors. Previous proposed SSM were input-driven based, such as:

$$y_n = Hx_n + Ay_n + w_n$$

$$x_n = Fx_n + Bx_n + v_n$$

in which A captures the causal relationships between genes and B captures the influences of the genes on the hidden state vectors (Rangel et al., 2004; Beal et al., 2005). The input-driven model not only considers gene expression profiles, but also concentrations of proteins or other factors. Therefore the model is more specific compared to a basic SSM. The derived network was estimated on the basis of this equation:

$$y_n = (HB+A)y_{n-1} + HFx_{n-1} + w_n + Hv_n$$

However, since the temporal regulation is based on (HB+A), which is a likelihood estimation, as the Vector Autoregressive Model (VAR), when the number of the genes is much larger then time course data, A and B cannot be efficiently estimated. Therefore the Hirose method was not based on the input-driven model, but on a standard SSM:

$$Y_n = Hx_n + w_n \qquad n \in N_{obs}$$

moreover the evolving time course is modelled following this equation

$$X_n = Fx_{n-1} + v_n \qquad n \in N$$

The model considers a parameter estimation process for the evaluation of the state dimension, hence of the *k* modules of genes relevant to the temporal structure of gene expression (Hirose et al., 2008). The temporal structure of the modules was defined by the autoregressive coefficient matrix

$$\psi \equiv D^T \wedge FD$$

which allows controlling the complexity of the system by selecting the dimensions of the state vector *k*. Since the method was based on replicated time-course data, it was possible to achieve highly efficient estimation of gene networks, by using permutation tests on the autoregressive coefficient matrix. The method can identify the temporal aggregations of the gene expression profiles and assemble them into large scale gene networks (Hirose et al., 2008).

The model was developed by Dr Francesco Falciani and was performed on an integrated dataset of transcriptomic and metabolomics experiments collected during acid exposure time courses in *E. coli* MG1655. 10 modules were identified after fitting the model under the likelihood estimation parameter $k = 5$. The definition of the positive and negative modules is made on the basis of the ranking of the highest and lowest 100 genes in the estimated projection matrix. The temporal structure of the network was determined by the estimated coefficient matrix ψ, with 5% value for significance. The network consisted of 1672 nodes and 438665 edges. The matrix ψ was then tresholded by selecting transcription factors of TCS and seven global regulators of *E. coli* (Martinez-Antonio and Collado-Vides, 2003). First neighbours of the selected hubs were considered, therefore obtaining a network of 1174 nodes in total (**Figure 5.1** and **Table 5.3**). Ten regulators were identified with the largest number of connections: *creB*, *narL*, *dcuR*, *fis*, *ihfB*, *lrp*, *ompR*, *phoP*, *rstA* and *yehT*.

5.2.4 SSM Model validation

EcID (Escherichia coli Interaction Database) was used for validation of the gene-gene interactions found in the SSM. It provides interactions based on the following sources: (metabolic pathways, protein complexes and regulatory information), KEGG (metabolic

pathways), MINT and IntAct (protein interactions), high-throughput pull-down experiments and potential interactions extracted from the literature using the web services associated with the iHOP text-mining system and various prediction methods based on genomic sequences, such phylogenetic profiles (Andres Leon et al., 2009). The database consists of 1384359 interactions, each of them annotated on the basis of the source of origin and classified by an assigned score, the AODE score, which gives a reliability measure associated with each connection.

5.3 Results

5.3.1 A SSM defines a temporal hierarchy of events linking transcriptional and metabolic response during acid adaptation

The SSM model we developed summarizes the transcriptional and metabolic response of the *E. coli* MG1655 strain in five modules, each one with a positive and negative component, hence obtaining a total of 10 modules equal to the number of clusters previously seen in chapter 4 (**Figure 5.1 and Figure 5.4.**).

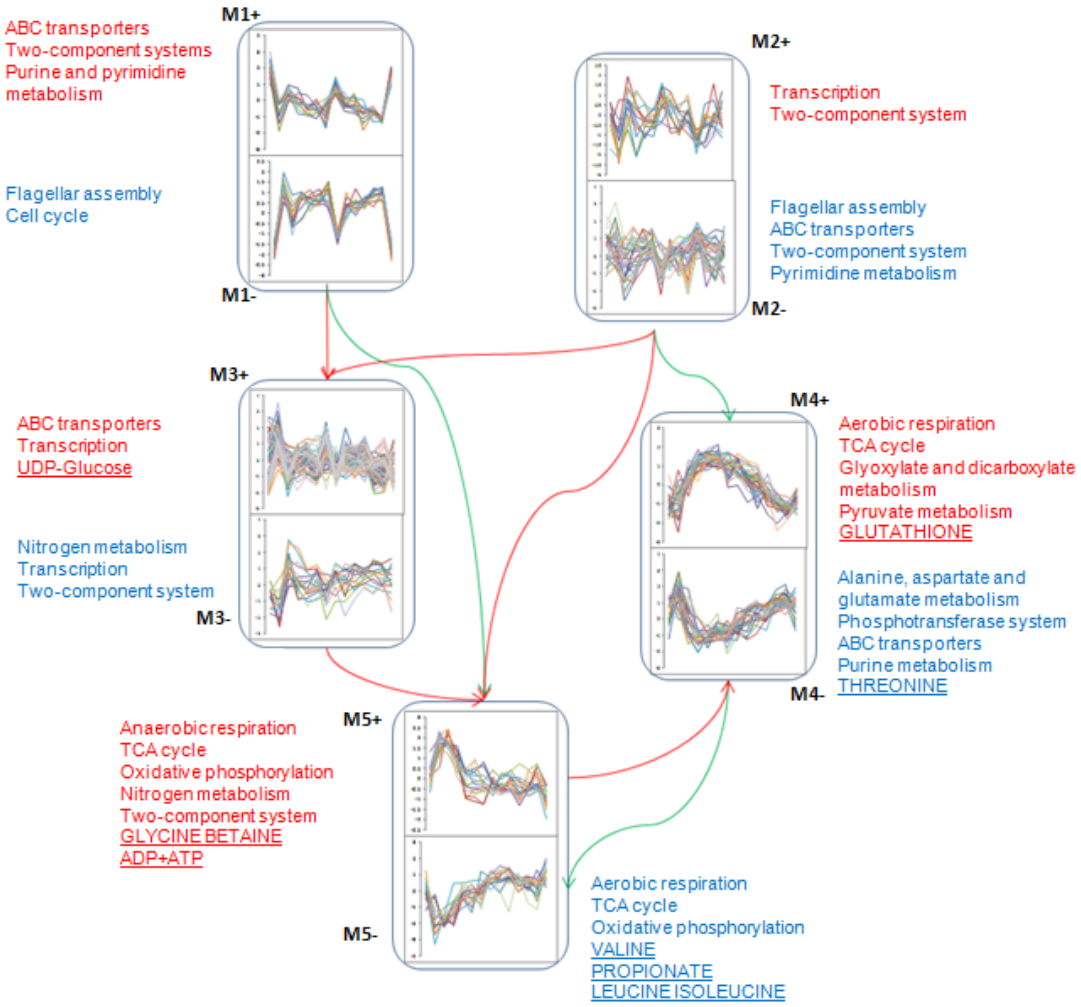


Figure 5.1 obtained from the SSM with the relative temporal structure.

The SSM allowed the identification of 5 modules, each of them was characterised by two symmetrical transcriptional profiles: + modules in red, - modules in blue. The structure shows that the TCS were located in the top module, the cellular metabolism functions were instead in the bottom modules. Most of the ABC transporters were found in the module 1+ and 3+, amino acid metabolism in module 4-. The metabolites (underlined uppercase words) were not identified in the first two modules of the model. Sucrose and glutathione were located in modules 3+ and 4+ respectively, glycine betaine and ADP+ATP in module 5+. Module 5- exhibited most of the amino acids (valine, leucine and isoleucine), while threonine was observed in module 4-.

The model shows that transporters and regulatory molecules such as TFs and TCSs sit on the highest level of the hierarchy alongside genes encoding for flagellum and cell cycle proteins (module M1 and M2). The lower level modules (modules M3, M4 and M5) were all characterised by the presence of energy metabolism pathways and by the metabolites measured with NMR. More specifically, glutathione and threonine were in M3 and M4 respectively and the majority of the amino acids, alongside with glycine betaine and ATP/ADP were represented in M5. This representation of the model somehow recapitulates a paradigm for response to external stimuli. TCSs can sense and transmit the signal to specific transcription factors, and these control the modulation of effector pathways such as energy metabolism and protein translation. Interestingly sucrose was found in the third module, which represents a rapidly decreasing profile. This is consistent with an earlier decrease in the expression of most of the ABC transporters (present in M1 and M2), as well as glucose specific transporters (Chapter 4). We could observe a similar relationship between early transcriptional activity followed by late metabolite changes in the amino acid metabolism pathways (M4 and M5 respectively).

5.3.2 Network inference reveals that TCS regulators coordinate the genome-wide response to environmental perturbations

The SSM described in the previous paragraph represents a high level model of acid response in the *E. coli* MG1655 strain. Since our ultimate aim is to identify the most important regulatory components it is necessary to select genes and metabolites in order to generate a more detailed model.

In order to support the selection of key regulatory genes in stress response, I developed an ARACNE network model from a compendium of 907 Affymetrix arrays, representing the same response of the strain to a wide range of stresses (Faith et al., 2008) (**Figure 5.2**). In order to identify key regulators in this network, I identified the most highly connected genes (network hubs) (**Table 5.1**).

Table 5.1 Hubs identification in the MI based network with the web based tool Hubba-Hubba (Lin et al., 2008).

Name	Score	Rank	Name	Score	Rank	Name	Score	Rank
uvrY	**1013**	**1**	yebR	9	24	stfR	8	36
cpxR	**990**	**2**	ycdY	9	24	pdxB	8	36
yehT	**708**	**3**	yajC	9	24	yibN	8	36
rssB	**623**	**4**	wbbL	9	24	lolB	8	36
yfjR	**550**	**5**	ycgL	9	24	chbG	8	36
evgA	**524**	**6**	rzpR	9	24	dnaB	8	36
rcsB	**480**	**7**	pdxJ	9	24	yrhC	8	36
ompR	**454**	**8**	yifL	9	24	ompG	8	36
dpiB	**399**	**9**	grxC	9	24	minC	8	36
rstA	**368**	**10**	glf	9	24	yqiB	8	36
torR	**366**	**11**	uhpA	8	36	dapF	8	36
cusR	**297**	**12**	trkH	8	36	ydaW	8	36
phoB	**295**	**13**	dsbB	8	36	ydjA	7	63
fimZ	**281**	**14**	ychE	8	36	rhlB	7	63
phoP	**228**	**15**	yiaM	8	36	ydhP	7	63
narL	**220**	**16**	narV	8	36	slyA	7	63
creB	**217**	**17**	yniB	8	36	selD	7	63
dcuR	**143**	**18**	ycgM	8	36	prmB	7	63
narP	**142**	**19**	rfaD	8	36	rfbC	7	63
basR	**81**	**20**	pphA	8	36	ycbV	7	63
baeR	**70**	**21**	tatB	8	36	ynfB	7	63
qseB	**46**	**22**	trpE	8	36	ycgJ	7	63
tehB	10	23	hemD	8	36	hyfR	7	63
ftsY	9	24	yeaX	8	36	ddpD	7	63
yiaA	9	24	minD	8	36	gcvR	7	63

The network was also graphically represented using a force driven layout and the top most highly connected hubs were labelled in red (**Figure 5.2**).

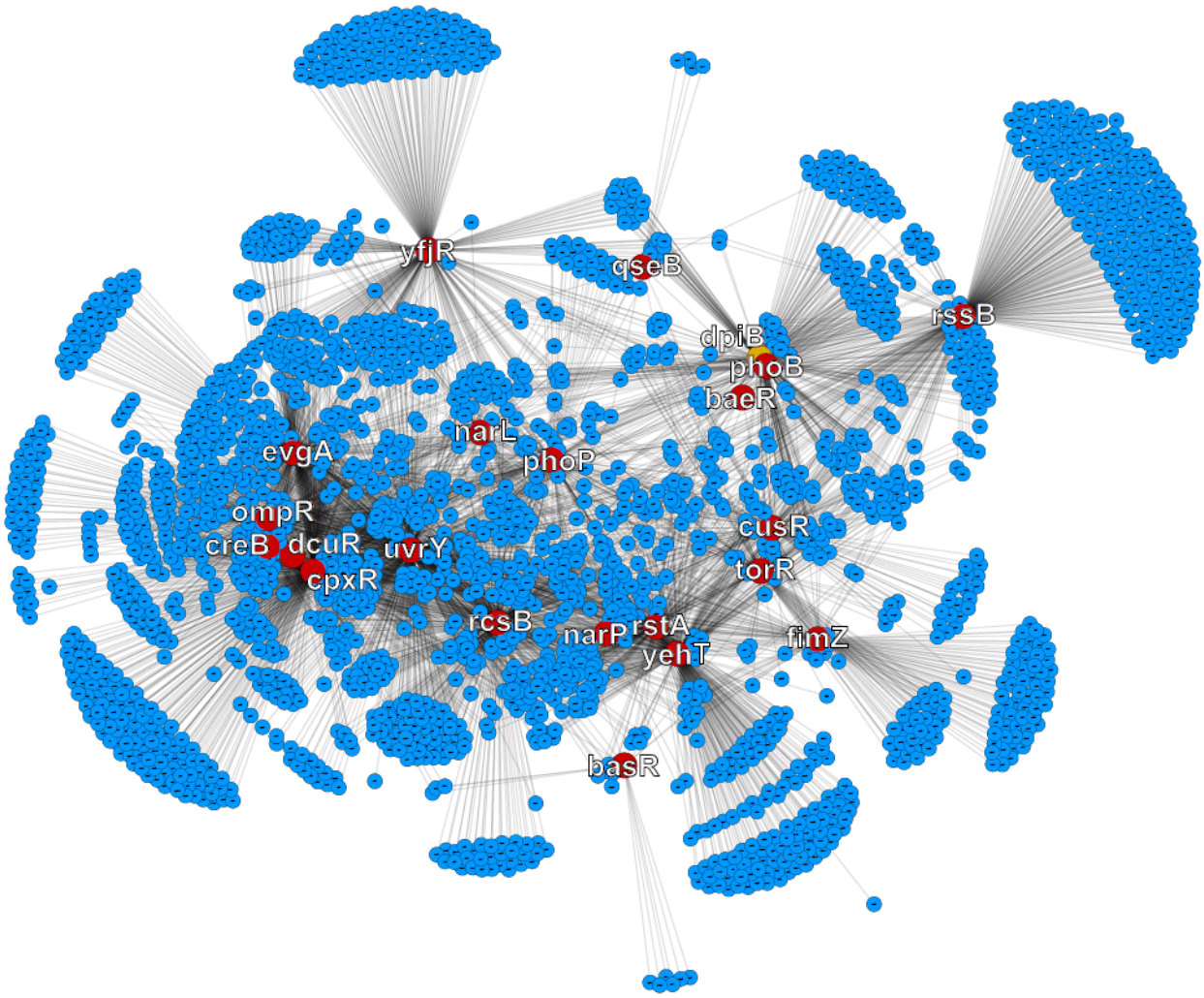


Figure 5.2 Mutual Information based network.

The network consisted of 3118 nodes, after performing a spring embed layout, I obtained the structure in figure. The nodes in red were the most highly connected genes, which also happened to be TCS regulators, except for the sensor of the DpiB (in orange).

Remarkably, I discovered that the top 22 gene hubs were either genes encoding for TCS regulators (21/22) or in case of *dpiB* a TCS sensor. Some of the nodes with the largest number of connections (*uvrY* and *yehT*) were putative TCS whose function in *E. coli* is yet largely uncharacterized.

In order to generate hypotheses on the biological processes controlled by each TCS, I performed a functional annotation analysis for the neighbours of each hub. Interestingly most of the energy metabolism pathways (both aerobic and anaerobic) were highly connected to *ompR*, *rcsB*, *cpxR*, *uvrY* and *yehT*. Consistent with what was reported in the literature we also found that a regulator of central carbon metabolism, *uvrY*, was characterized by the largest number of connections (1013), with genes involved in energy metabolism, ribosome and cell cycle (**Table 5.2**).

Table 5.2 Functional annotation analysis of ARACNE hub neighbours.

Functional annotation analysis of *cpxR* neighbours				
Category	Term	Count	p-value	Benjamini
KEGG	Oxidative phosphorylation	28	1.10E-26	1.80E-23
GO_BP	aerobic respiration	22	8.80E-08	2.40E-06
GO_BP	carbohydrate biosynthetic process	79	8.40E-20	3.20E-17
KEGG	Aminoacyl-tRNA biosynthesis	19	1.90E-18	3.90E-16
GO_BP	anaerobic respiration	35	7.30E-16	9.80E-14
GO_BP	cell cycle	26	5.00E-10	2.00E-08
GO_BP	cell division	25	3.10E-07	6.90E-06
KEGG	Purine metabolism	31	4.80E-19	1.30E-16
KEGG	Pyrimidine metabolism	15	1.20E-07	3.00E-06
KEGG	Ribosome	17	3.10E-09	1.20E-07
KEGG	Glycolysis / Gluconeogenesis	18	4.30E-13	2.40E-11
KEGG	Fatty acid biosynthesis	9	1.30E-08	4.40E-07
KEGG	Glutathione metabolism	9	5.10E-07	1.00E-05
KEGG	Pentose phosphate pathway	14	4.30E-10	1.80E-08
KEGG	Two-component system	25	1.90E-08	5.90E-07
KEGG	Amino sugar and nucleotide sugar metabolism	14	1.00E-07	2.60E-06
KEGG	Fructose and mannose metabolism	12	3.50E-07	7.50E-06
KEGG	Protein export	10	2.60E-08	7.90E-07
KEGG	Bacterial secretion system	10	2.20E-06	3.60E-05
KEGG	Pyruvate metabolism	13	3.50E-07	7.40E-06
KEGG	RNA degradation	8	3.00E-06	4.60E-05
Functional annotation analysis of *ompR* neighbours				
Category	Term	Count	p-value	Benjamini
KEGG	Ribosome	31	1.40E-32	1.80E-29
KEGG	Protein export	13	1.80E-15	1.80E-13
KEGG	Bacterial secretion system	14	4.30E-13	3.00E-11
KEGG	Pyrimidine metabolism	14	2.20E-09	1.10E-07
KEGG	Purine metabolism	14	4.50E-07	1.60E-05
GO_BP	cell cycle	12	1.80E-04	4.10E-03
GO_BP	cell division	13	3.90E-04	7.60E-03
KEGG	Two-component system	16	7.50E-06	1.90E-04
KEGG	Pentose phosphate pathway	7	1.40E-04	2.40E-03
KEGG	DNA replication	5	1.00E-03	1.30E-02
KEGG	ABC transporters	19	6.10E-06	1.60E-04
GO_BP	anaerobic respiration	13	1.10E-04	2.90E-03
KEGG	Lipopolysaccharide biosynthesis	7	1.10E-04	1.90E-03
KEGG	Amino sugar and nucleotide sugar metabolism	10	4.30E-06	1.20E-04
Functional annotation analysis of *rcsB* neighbours				
Category	Term	Count	p-value	Benjamini

GO_BP	carboxylic acid biosynthetic process	41	5.20E-10	1.50E-07
GO_BP	cell cycle	15	7.80E-07	3.40E-05
KEGG	Amino sugar and nucleotide sugar metabolism	12	7.10E-09	9.90E-07
KEGG	Glycolysis / Gluconeogenesis	12	6.60E-10	2.10E-07
KEGG	Glycine, serine and threonine metabolism	10	4.20E-08	3.70E-06
KEGG	Vitamin B6 metabolism	6	1.10E-06	4.40E-05
KEGG	Fructose and mannose metabolism	9	1.10E-06	4.30E-05
KEGG	Bacterial secretion system	12	1.70E-11	2.20E-08
KEGG	Protein export	7	2.30E-06	7.80E-05
KEGG	Phenylalanine, tyrosine and tryptophan biosynthesis	8	3.40E-07	1.60E-05
KEGG	Oxidative phosphorylation	8	2.10E-05	5.80E-04
GO_BP	anaerobic respiration	12	2.90E-04	4.90E-03
KEGG	Lysine biosynthesis	5	3.20E-04	5.90E-03
KEGG	Glycerophospholipid metabolism	6	1.80E-04	3.60E-03
KEGG	Cysteine and methionine metabolism	6	4.90E-04	8.10E-03
KEGG	Two-component system	13	8.40E-05	1.90E-03
KEGG	Nitrogen metabolism	7	2.10E-04	4.20E-03
KEGG	Purine metabolism	9	6.10E-04	9.30E-03
KEGG	Pyrimidine metabolism	7	1.80E-03	2.30E-02
KEGG	Pentose phosphate pathway	6	8.20E-04	1.20E-02
KEGG	Alanine, aspartate and glutamate metabolism	5	4.30E-03	4.70E-02
Functional annotation analysis of *uvrY* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
KEGG	Ribosome	51	9.50E-57	1.40E-53
SP_PIR	cell cycle	30	3.20E-22	4.80E-21
KEGG	Lipopolysaccharide biosynthesis	18	5.50E-15	2.50E-13
KEGG	Purine metabolism	32	1.20E-17	1.10E-15
KEGG	Pyrimidine metabolism	26	1.20E-17	1.10E-15
KEGG	Aminoacyl-tRNA biosynthesis	16	6.40E-13	2.20E-11
KEGG	Fatty acid biosynthesis	11	3.10E-11	8.80E-10
KEGG	Bacterial secretion system	18	3.00E-14	1.30E-12
KEGG	Protein export	14	1.90E-13	7.30E-12
KEGG	Oxidative phosphorylation	23	2.70E-17	2.20E-15
GO_BP	anaerobic respiration	28	7.40E-10	2.90E-08
GO_BP	aerobic respiration	17	1.90E-04	2.60E-03
KEGG	Homologous recombination	13	1.30E-08	3.30E-07
KEGG	Peptidoglycan biosynthesis	7	1.10E-03	9.40E-03
KEGG	Pentose and glucuronate interconversions	11	3.70E-06	6.60E-05
KEGG	Arginine and proline metabolism	15	2.20E-08	5.30E-07
KEGG	Pentose phosphate pathway	11	5.20E-06	8.90E-05
KEGG	Glyoxylate and dicarboxylate metabolism	10	2.00E-04	2.10E-03
KEGG	ABC transporters	28	4.20E-06	7.20E-05
KEGG	Glycolysis / Gluconeogenesis	12	4.60E-06	8.00E-05

KEGG	Pyruvate metabolism	12	3.80E-05	5.20E-04
KEGG	Two-component system	21	9.60E-05	1.10E-03
KEGG	Amino sugar and nucleotide sugar metabolism	12	3.80E-05	5.20E-04
Functional annotation analysis of *yehT* neighbours				
Category	Term	Count	p-value	Benjamini
KEGG	Oxidative phosphorylation	18	1.10E-15	1.80E-12
GO_BP	aerobic respiration	27	1.10E-14	8.40E-12
GO_BP	anaerobic respiration	26	5.20E-12	7.70E-10
KEGG	Amino sugar and nucleotide sugar metabolism	14	4.90E-09	3.10E-07
KEGG	ABC transporters	32	5.80E-13	1.90E-10
KEGG	Flagellar assembly	14	2.90E-10	3.50E-08
KEGG	Phenylalanine, tyrosine and tryptophan biosynthesis	9	5.30E-07	2.20E-05
KEGG	Alanine, aspartate and glutamate metabolism	10	5.50E-07	2.20E-05
KEGG	Nitrogen metabolism	9	3.10E-05	7.70E-04
KEGG	Pyrimidine metabolism	13	3.90E-07	1.70E-05
KEGG	Glycine, serine and threonine metabolism	10	1.90E-06	6.80E-05
KEGG	Aminoacyl-tRNA biosynthesis	7	1.70E-04	3.40E-03
KEGG	Arginine and proline metabolism	10	1.80E-05	5.00E-04
KEGG	Two-component system	16	8.10E-05	1.80E-03
KEGG	Fatty acid biosynthesis	5	8.00E-04	1.20E-02
KEGG	Fructose and mannose metabolism	10	9.20E-06	2.70E-04
KEGG	Phosphotransferase system (PTS)	8	8.00E-04	1.20E-02
KEGG	Purine metabolism	12	1.80E-04	3.60E-03
KEGG	Pyruvate metabolism	10	3.30E-05	8.10E-04

Table 5.3 TCS regulators and connected sensors.
Connections between TCS regulators and sensors (in red) were detected in ARACNE. The nodes *basR*, *dcuR*, *narP*, *ompR*, *phoP*, *qseB*, *rcsB* and *yehT* were all presenting interactions with their corresponding sensors. *yehT* showed the highest number of interactions with TCS sensors.

Regulator	Sensors neighbours
baeR	*rcsC, rcsD*
basR	*basS*
cpxR	*phoR, arcB, glnL, cheA, cpxA*
creB	*dcuS, arcB*
cusR	*rcsC*
dcuR	*dcuS*
dpiB	*cheY, narX, narQ*
evgA	*ypdA, dcuS, arcB, yedW*
fimZ	*qseC*
kdpE	
narL	*narX*
narP	*narX*
ompR	*envZ, dcuS, torS*
phoB	*phoR, uhpB*
phoP	*phoQ, rcsD*
qseB	*qseC*
rcsB	*rcsC, qseC, arcB, rcsD*
rssB	*yfhK, rcsC, kdpD*
rstA	*rstB, phoQ, rcsC*
torR	*cheY, cheB, kdpD, cheA*
uhpA	
uvrY	*narX, yedV, yedW*
yehT	*cheA, narX, rcsC, yehU, dcuS, narQ*
yfjR	*arcB, narQ*

Half of the TCS regulators (BasR, CpxR, DcuR, NarP, OmpR, PhoB, PhoP, QseB, RcsB, RstA and YehT) were also directly connected to genes encoding for their partner sensor proteins. This observation is consistent with the literature where several studies (Wulf et al., 2000; Groisman, 2001) report that regulators are able to modulate the expression of their sensor components. Interestingly, the uncharacterized *yehT* gene was directly connected to 6 TCS sensors suggesting that it could be involved in the coordination of other TCS (**Table 5.3**).

5.3.3 A TCS-centric model of acid response

Since the ARACNE model supported the hypothesis that TCS regulators are likely to be the highest level of the hierarchy, I extracted from the SSM model a sub-graph that represented the neighbourhood of the TCS regulators. I obtained a network of 1174 nodes and 4946 edges (**Figure 5.3**).

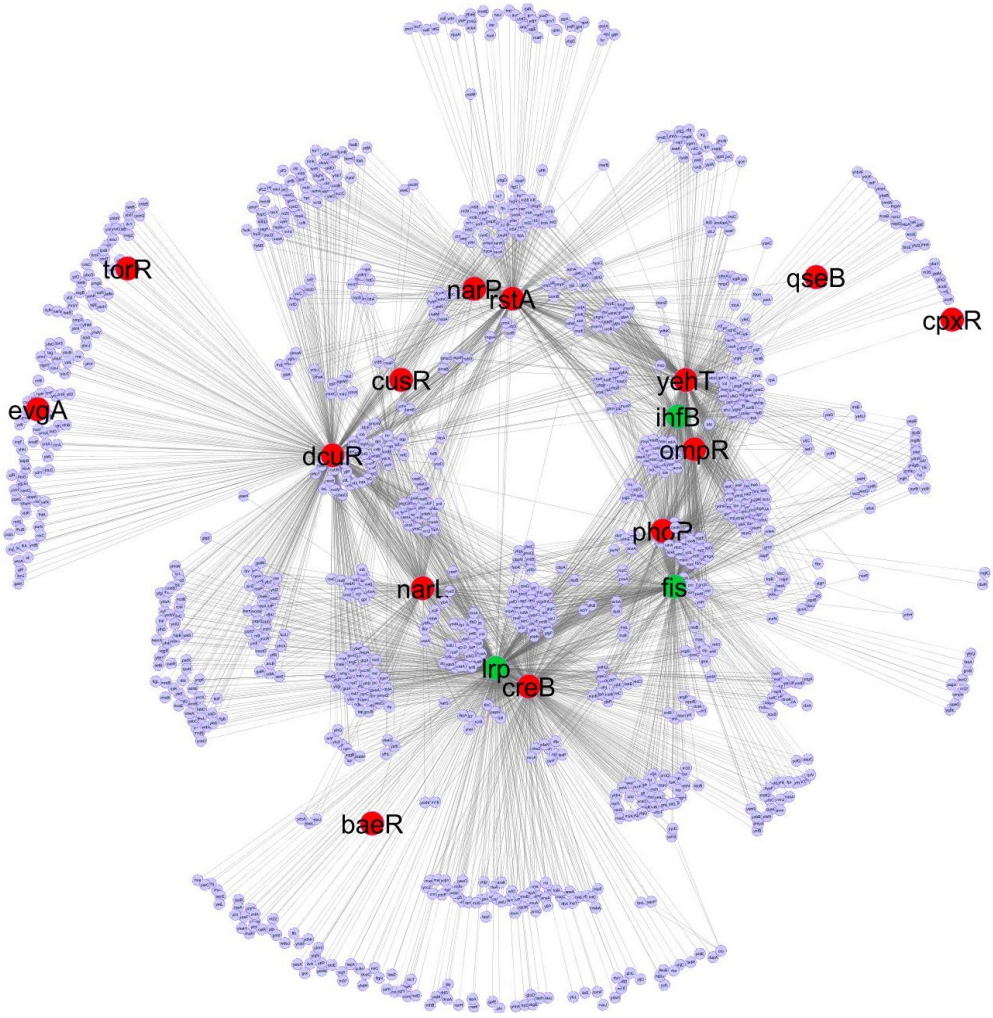


Figure 5.3 Network obtained from the selection of the TCS regulators and master regulators.
The network was obtained from the estimated coefficient matrix ψ after selection of the main regulators of TCS (in red) and the master regulators in *E. coli* (in green).

In order to identify the regulators that are more likely to control the main functions we knew are modulated during acid adaptation, I have performed a functional annotation analysis on the neighbours of each seed hub in the network (**Table 5.4**). The analysis revealed that TCS regulators *ompR*, *yehT*, *dcuR*, *rstA* and *phoP* have the highest number of connections to the genes coding for aerobic and anaerobic respiration pathways, amino acid metabolism, sugar metabolism and TCS.

Table 5.4 Functional annotation of the main hubs of the SSM network.

Functional annotation *rstA* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
GOTERM_BP_FAT	anaerobic respiration	42	1.70E-36	3.30E-34
GOTERM_BP_FAT	oxidation reduction	78	2.80E-11	1.80E-09
KEGG_PATHWAY	Citrate cycle (TCA cycle)	19	1.80E-23	2.30E-20
KEGG_PATHWAY	Oxidative phosphorylation	16	3.20E-15	3.10E-13
KEGG_PATHWAY	Phosphotransferase system (PTS)	12	8.20E-10	4.30E-08
KEGG_PATHWAY	Fructose and mannose metabolism	10	2.60E-08	8.70E-07
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	7	2.60E-04	2.80E-03
KEGG_PATHWAY	Pyruvate metabolism	11	1.40E-08	5.20E-07
KEGG_PATHWAY	Nitrogen metabolism	8	7.60E-06	1.30E-04
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	7	2.20E-05	3.40E-04
KEGG_PATHWAY	Two-component system	14	3.20E-06	6.40E-05
KEGG_PATHWAY	ABC transporters	15	3.10E-05	4.50E-04
KEGG_PATHWAY	Glycerolipid metabolism	4	1.20E-03	1.10E-02
Functional annotation *dcuR* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
GOTERM_BP_FAT	anaerobic respiration	43	8.00E-32	1.80E-29
KEGG_PATHWAY	Two-component system	29	1.60E-17	2.40E-14
KEGG_PATHWAY	Ribosome	16	2.80E-11	1.50E-09
KEGG_PATHWAY	Citrate cycle (TCA cycle)	15	1.70E-14	2.00E-12
GOTERM_BP_FAT	fermentation	16	8.60E-14	1.20E-11
KEGG_PATHWAY	Oxidative phosphorylation	14	9.70E-11	4.40E-09
KEGG_PATHWAY	Fructose and mannose metabolism	11	2.50E-08	7.80E-07
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	10	4.90E-06	8.90E-05
KEGG_PATHWAY	Phosphotransferase system (PTS)	10	8.70E-06	1.40E-04
KEGG_PATHWAY	Purine metabolism	13	2.30E-06	4.80E-05
KEGG_PATHWAY	Flagellar assembly	10	8.00E-07	1.80E-05
KEGG_PATHWAY	ABC transporters	18	2.30E-05	3.40E-04
KEGG_PATHWAY	Glycerolipid metabolism	5	1.50E-04	1.80E-03
KEGG_PATHWAY	Glycine, serine and threonine metabolism	7	2.70E-04	3.20E-03
KEGG_PATHWAY	Bacterial chemotaxis	6	2.50E-04	3.10E-03

KEGG_PATHWAY	Pyruvate metabolism	9	2.40E-05	3.60E-04
KEGG_PATHWAY	Nitrogen metabolism	9	2.60E-06	5.20E-05
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	4	4.60E-02	2.90E-01
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	7	6.40E-04	6.50E-03
KEGG_PATHWAY	Galactose metabolism	6	2.10E-03	2.00E-02
KEGG_PATHWAY	Glycerophospholipid metabolism	5	3.80E-03	3.30E-02

Functional annotation *yehT* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
GOTERM_BP_FAT	anaerobic respiration	33	6.70E-26	1.80E-23
GOTERM_BP_FAT	carbohydrate catabolic process	53	9.90E-23	1.10E-20
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	11	1.80E-09	1.00E-07
KEGG_PATHWAY	Citrate cycle (TCA cycle)	14	2.30E-15	2.50E-12
GOTERM_BP_FAT	aerobic respiration	22	5.10E-15	4.60E-13
KEGG_PATHWAY	Oxidative phosphorylation	10	9.10E-08	2.20E-06
KEGG_PATHWAY	ABC transporters	18	1.40E-07	3.10E-06
KEGG_PATHWAY	Phosphotransferase system (PTS)	13	2.00E-11	2.10E-09
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	8	1.80E-05	2.50E-04
KEGG_PATHWAY	Two-component system	16	3.40E-08	1.10E-06
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	10	2.00E-09	1.10E-07
KEGG_PATHWAY	Nitrogen metabolism	6	4.00E-04	4.00E-03
KEGG_PATHWAY	Pyruvate metabolism	10	1.10E-07	2.70E-06
KEGG_PATHWAY	Fructose and mannose metabolism	7	5.10E-05	6.20E-04
KEGG_PATHWAY	Cysteine and methionine metabolism	6	2.00E-04	2.30E-03
KEGG_PATHWAY	Glutathione metabolism	5	2.70E-04	2.90E-03
KEGG_PATHWAY	Pentose phosphate pathway	7	1.90E-05	2.60E-04
KEGG_PATHWAY	Lysine degradation	4	6.80E-04	6.40E-03
KEGG_PATHWAY	Arginine and proline metabolism	6	1.20E-03	1.10E-02
KEGG_PATHWAY	Glycine, serine and threonine metabolism	5	3.70E-03	2.90E-02
Functional annotation *ompR* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
GOTERM_BP_FAT	carbohydrate catabolic process	32	4.70E-15	1.90E-12
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	7	3.00E-07	2.00E-05
GOTERM_BP_FAT	anaerobic respiration	12	2.00E-07	2.00E-05
KEGG_PATHWAY	Citrate cycle (TCA cycle)	5	2.00E-04	3.70E-03
KEGG_PATHWAY	Phosphotransferase system (PTS)	9	8.40E-09	6.60E-06
KEGG_PATHWAY	Fructose and mannose metabolism	6	2.30E-05	7.20E-04
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	6	3.60E-05	9.90E-04
KEGG_PATHWAY	ABC transporters	10	9.30E-05	2.20E-03
GOTERM_BP_FAT	aerobic respiration	11	3.50E-07	2.10E-05
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	7	4.10E-06	1.70E-04
KEGG_PATHWAY	Fructose and mannose metabolism	6	2.30E-05	7.20E-04

KEGG_PATHWAY	Arginine and proline metabolism	5	7.50E-04	1.00E-02
KEGG_PATHWAY	Pentose phosphate pathway	5	2.00E-04	3.70E-03
KEGG_PATHWAY	Glyoxylate and dicarboxylate metabolism	4	3.90E-03	4.10E-02

Functional annotation *phoP* neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
KEGG_PATHWAY	Two-component system	16	4.10E-10	8.60E-08
KEGG_PATHWAY	Glycine, serine and threonine metabolism	8	3.20E-07	1.30E-05
GOTERM_BP_FAT	carbohydrate catabolic process	34	7.00E-13	3.30E-10
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	10	1.30E-10	1.10E-07
KEGG_PATHWAY	Nitrogen metabolism	6	9.60E-05	1.70E-03
KEGG_PATHWAY	Glyoxylate and dicarboxylate metabolism	8	5.10E-07	1.80E-05
KEGG_PATHWAY	Starch and sucrose metabolism	8	4.00E-07	1.50E-05
KEGG_PATHWAY	Citrate cycle (TCA cycle)	5	8.10E-04	9.60E-03
KEGG_PATHWAY	Arginine and proline metabolism	6	2.90E-04	4.20E-03
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	7	1.50E-05	3.60E-04
KEGG_PATHWAY	Pentose phosphate pathway	6	5.70E-05	1.20E-03
KEGG_PATHWAY	Pyruvate metabolism	6	4.70E-04	6.20E-03
KEGG_PATHWAY	ABC transporters	12	7.00E-05	1.40E-03
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	6	3.70E-04	5.10E-03

I therefore set to further simplify the networks focussing on the neighbourhoods of these genes. **Figure 5.4** shows the hierarchical view of such a network.

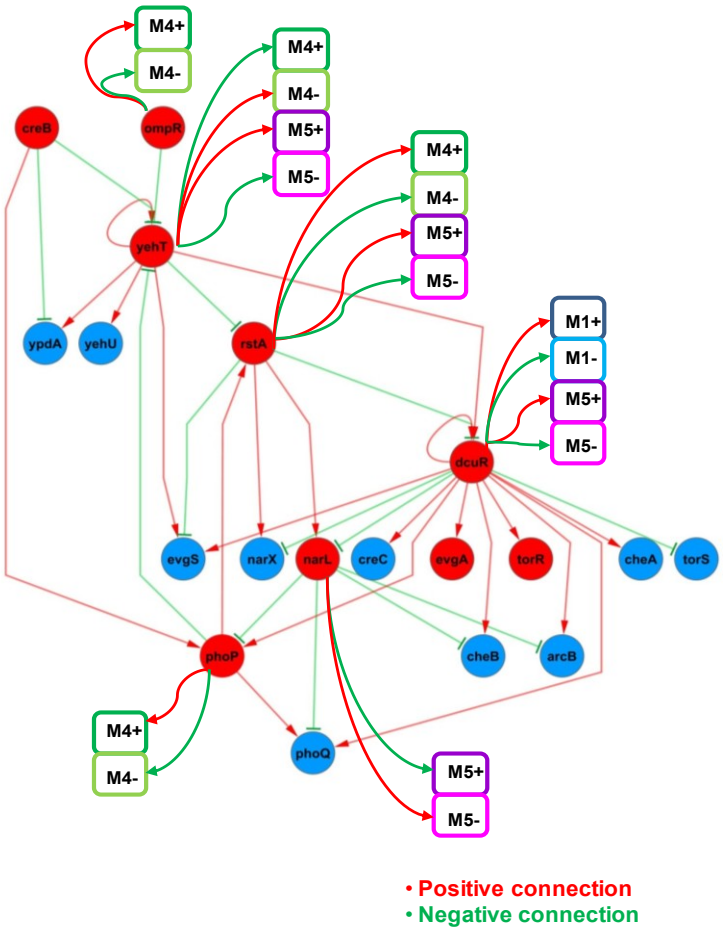


Figure 5.4 Hierarchical structure obtained by the selection of TCS regulators and sensors. OmpR and CreB are on the top of the hierarchy controlling the TCS regulatory structure. The central role of DcuR is also shown, on the basis of the high number of connections with most of the TCS found in the network. The edges are based on the estimated coefficient: in green a negative value and red a positive value.

OmpR and CreB represent the highest level of the regulatory hierarchy. Lrp and Fis are important regulators in *E. coli*, additionally they belong to another important category for microorganisms: One Component Systems, OCS. I have performed a functional annotation analysis for the neighbours of these regulators and I have seen that *lrp* was highly connected to the genes encoding translational and amino acids biosynthesis pathways,

while *fis* neighbours were mostly involved in energetic and metabolic pathways (**Table 5.5**).

Table 5.5 Functional annotation analysis of *fis* and *lrp* neighbours found in the SSM.

Functional annotation of ***lrp*** neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
KEGG_PATHWAY	Two-component system	32	3.90E-18	5.80E-15
GOTERM_BP_FAT	anaerobic respiration	28	3.90E-13	4.30E-11
KEGG_PATHWAY	Ribosome	19	2.40E-13	1.70E-11
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	15	7.50E-14	6.60E-12
KEGG_PATHWAY	Nitrogen metabolism	14	1.90E-11	8.80E-10
KEGG_PATHWAY	Purine metabolism	20	1.70E-11	8.00E-10
KEGG_PATHWAY	Starch and sucrose metabolism	12	8.60E-09	2.00E-07
KEGG_PATHWAY	Glyoxylate and dicarboxylate metabolism	13	8.40E-10	2.40E-08
KEGG_PATHWAY	ABC transporters	28	2.30E-10	7.70E-09
KEGG_PATHWAY	Glycine, serine and threonine metabolism	10	1.00E-06	1.80E-05
KEGG_PATHWAY	Arginine and proline metabolism	12	2.10E-07	3.80E-06
KEGG_PATHWAY	Flagellar assembly	12	3.50E-08	7.50E-07
KEGG_PATHWAY	Pyruvate metabolism	11	2.50E-06	3.90E-05
KEGG_PATHWAY	Cysteine and methionine metabolism	9	3.60E-06	5.50E-05
KEGG_PATHWAY	Bacterial chemotaxis	7	5.80E-05	7.40E-04
KEGG_PATHWAY	Phenylalanine, tyrosine and tryptophan biosynthesis	7	5.80E-05	7.40E-04
KEGG_PATHWAY	Pyrimidine metabolism	10	9.70E-05	1.20E-03
KEGG_PATHWAY	Amino sugar and nucleotide sugar metabolism	10	2.40E-05	3.20E-04
Functional annotation of ***fis*** neighbours				
Category	**Term**	**Count**	**p-value**	**Benjamini**
GOTERM_BP_FAT	anaerobic respiration	19	9.50E-10	8.40E-08
GOTERM_BP_FAT	aerobic respiration	17	3.80E-09	2.90E-07
KEGG_PATHWAY	Two-component system	18	3.70E-09	5.30E-07
KEGG_PATHWAY	Alanine, aspartate and glutamate metabolism	14	7.30E-15	8.40E-12
KEGG_PATHWAY	Nitrogen metabolism	8	6.20E-06	1.10E-04
KEGG_PATHWAY	Arginine and proline metabolism	10	2.30E-07	7.20E-06
KEGG_PATHWAY	ABC transporters	20	2.60E-08	1.80E-06
KEGG_PATHWAY	Glycine, serine and threonine metabolism	8	6.20E-06	1.10E-04
KEGG_PATHWAY	Starch and sucrose metabolism	10	3.40E-08	2.10E-06
KEGG_PATHWAY	Glyoxylate and dicarboxylate metabolism	10	4.60E-08	2.40E-06
KEGG_PATHWAY	Citrate cycle (TCA cycle)	6	4.50E-04	4.90E-03
KEGG_PATHWAY	Ribosome	9	5.30E-05	7.20E-04
KEGG_PATHWAY	Purine metabolism	11	1.10E-05	1.70E-04
KEGG_PATHWAY	Cysteine and methionine metabolism	9	1.30E-07	4.80E-06
KEGG_PATHWAY	Glutathione metabolism	5	4.50E-04	4.90E-03

KEGG_PATHWAY	Pyruvate metabolism	7	4.90E-04	5.20E-03
KEGG_PATHWAY	Pentose phosphate pathway	7	4.00E-05	5.60E-04
KEGG_PATHWAY	Glycolysis / Gluconeogenesis	7	1.80E-04	2.10E-03

The node *lrp* was also connected to 13 TCS encoding genes (*arcB*, *phoQ*, *narL*, *cusR*, *dcuR*, *cheW*, *cheB*, *phoP*, *creB*, *cheA*, *baeR*, *uhpT* and *rstA*). This regulator belongs to an important group of transcription factors in *E. coli* (Martinez-Antonio and Collado-Vides, 2003), which are in control of several functions and stress responses. These global regulators were described as the most connected components for several regulatory mechanisms, based on experimental validation from RegulonDB (Martinez-Antonio and Collado-Vides, 2003).

In order to evaluate the reliability of the connections found in the SSM model, a different database was considered, the EcID database (Andres Leon et al., 2009), which considered EcoCyc and KEGG validated interactions, in combination with results obtained from various prediction methods.

5.3.4 SSM validation through the EcID database

The TCS centric SSM suggested that *ompR*, *dcuR* and *yehT* may be the regulators of the transcriptional response to acid adaptation in the MG1655 strain.

In order to validate this hypothesis I have used "The *Escherichia coli* Integration Database" (EcID) (Andres Leon et al., 2009), a large database of gene-gene interactions derived from several sources, such as EcoCyc (Keseler et al., 2009), KEGG (Kanehisa et al., 2008) and IntAct (Aranda et al., 2010).

The intersection between the EcID database and the SSM matrix consisted of a network of approximately 600 genes. The edges of the network were supported by experimental evidence (high-throughput experiments, EcoCyc TF co-regulators and KEGG pathways), but also by phylogenetic profiles. In order to evaluate the validated interactions between the TCS regulators and their target genes, confirmed in the EcID database, I have performed a functional annotation analysis of the TCS targets. As shown in **figure 5.5**, OmpR targets, validated by EcoCyc interactions and high throughput experiments, were involved in aerobic respiration pathways, amino acid and sugar metabolism and dicarboxylate metabolism. DcuR was found to share some of the aerobic functions with OmpR, but the rest of the targets were involved in nitrate respiration and anaerobic respiration pathways, supported by phylogenetic profiles.

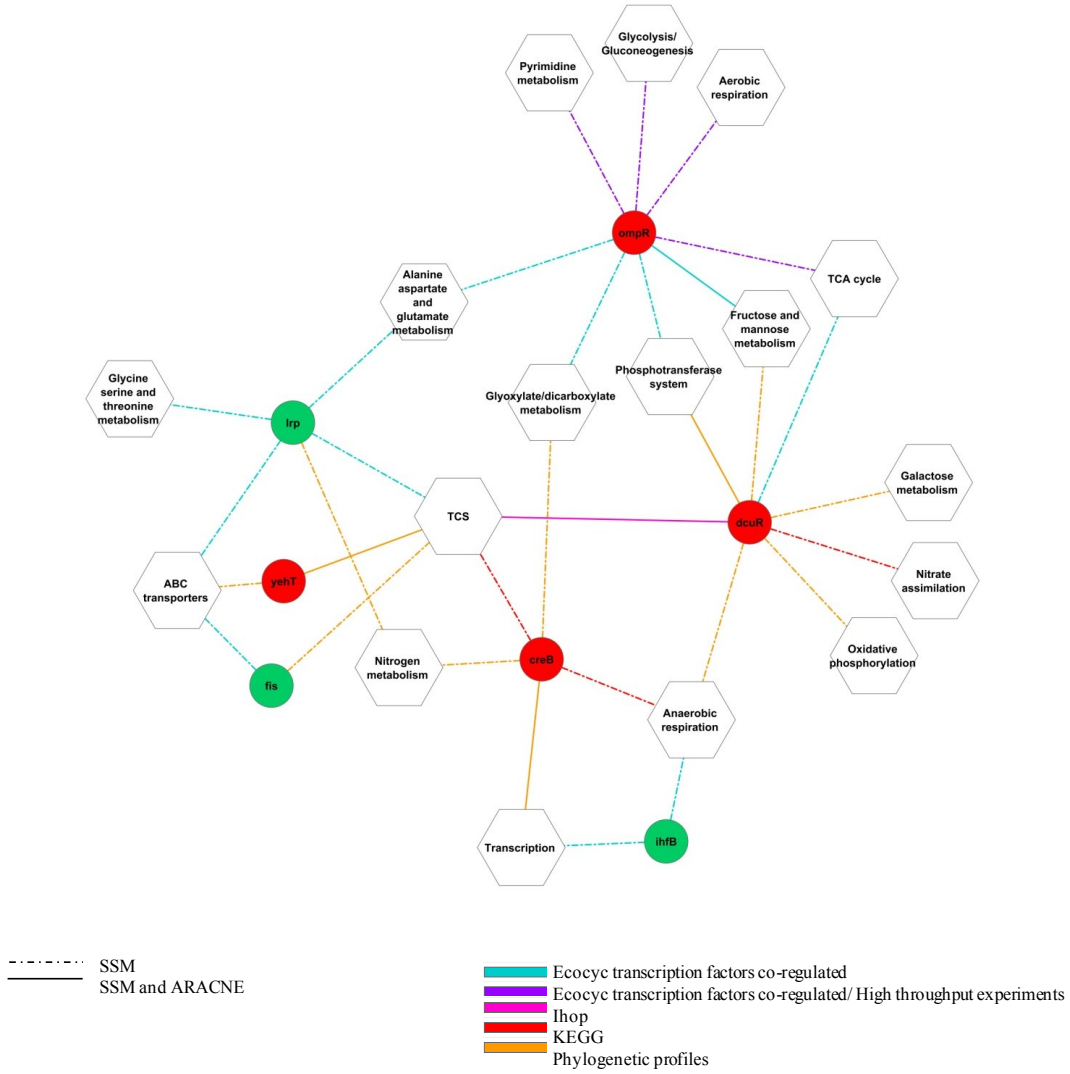


Figure 5.5 EcID based network (Andres Leon et al. 2009).

The network was obtained by selecting the potential relevant regulator of acid response in the SSM and considering the functions enriched by the genes associated to these regulators, which connections are based on experimental evidence, modelling and knowledge on interactions (Andres Leon et al., 2009). The edges overlapping with the SSM and ARACNE networks were also considered.

In the network it was not possible to find many validated connections for YehT. This is not surprising since this regulator is still uncharacterized. The global regulator CreB was also found connected to anaerobic respiration pathways and nitrogen metabolism, which interactions were validated by Ihop database and phylogenetic profiles. In the network in figure I have also considered the master regulators Lrp, Fis and IhfB, in green, since they

are known as master regulators in *E. coli*(Martinez-Antonio and Collado-Vides, 2003). All the connections seen for Lrp and IhfB were confirmed by experimental validation, while the connection with Fis and TCS was based on phylogenetic profiles evidence.

5.4 Discussion

5.4.1 A model linking TCS regulators to genome wide environmental stress response

TCS are known to be important component for microorganisms, because of their ability to sense and activate the cascade of transcriptional events allowing the cells to respond to external stimuli. TCS are sophisticated signalling systems which design has been adapted and integrated into a large variety of cellular circuits (Stock et al., 2000). Some TCS interactions are based on branched pathways, in which for example one sensor of the TCS can phosphorylate two different regulators, such the CheA-CheB/CheY TCS (Kirby, 2009). Moreover, events of cross-talking and cross-regulation via auxiliary proteins can characterise the responses to several stimuli (Goulian, 2010). However, little is known about the ability to integrate the stimuli into the complex *E. coli* regulatory network.

The results obtained from the ARACNE network were unexpected since a previous a study performed on the same dataset using a different gene inference approach (CLR) (Faith et al., 2007) failed to reveal this emerging property. Similarly to ARACNE, the CLR algorithm uses mutual information to estimate gene-to-gene connections. However, the model developed by Faith et al. was based on TFs present in RegulonDB rather than including every gene in the dataset. Moreover, in the ARACNE network I developed indirect connections were eliminated using the inequality principle. It is possible that these differences may explain why other groups before have not reported this finding.

Another important aspect to consider to fully understand the importance of our finding is the model of transcriptional regulation that emerges from the analyses of RegulonDB (Gama-Castro et al., 2011), a database based on experimentally validated transcription factor-target interactions (**Figure 5.6**).

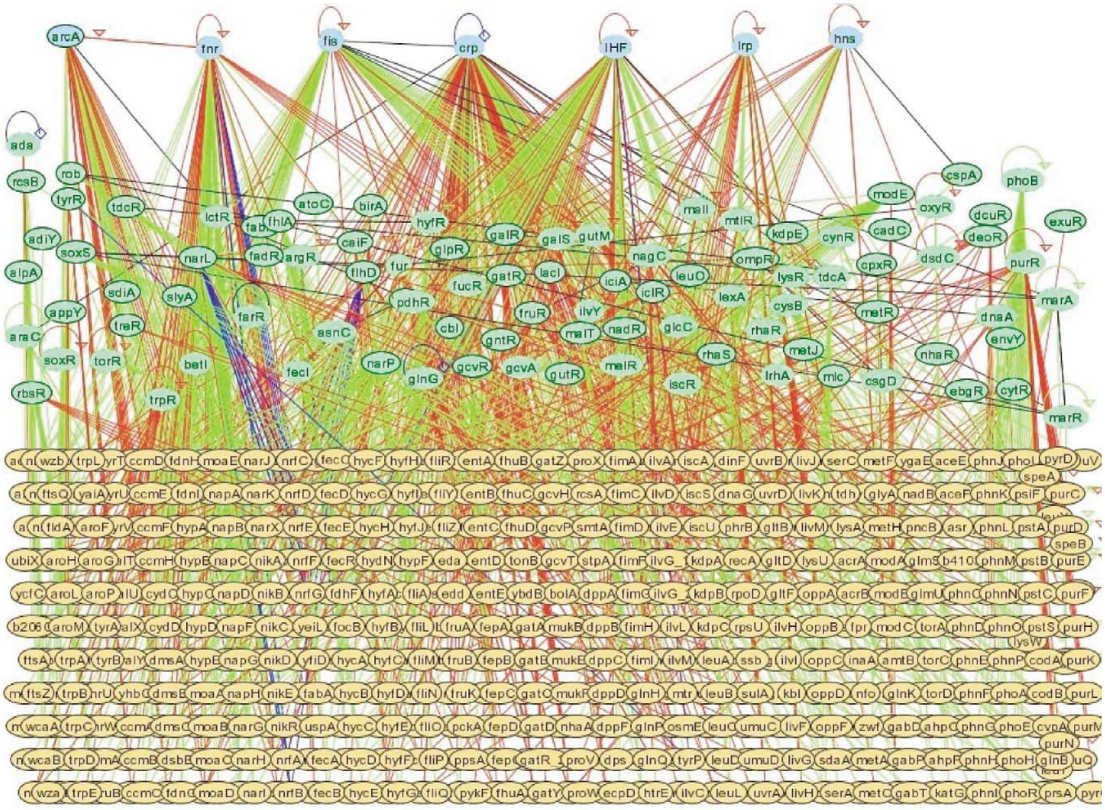


Figure 5.6 RegulonDB network showing the hierarchy behind the transcriptional regulation mechanism in *E. coli* (Martinez-Vides, 2003).

On the basis of this network, the hierarchy of gene regulation in *E. coli* is based on the action of seven *E. coli* master regulators (ArcA, Fnr, Fis, Crp, IHF, Lrp and Hns), which are located on the top of the structure in Figure 5.6. The TCS regulators were all found on the second level of the hierarchical structure, with all the relative targets downstream. Our model is therefore more representative of the hypothetical role TCS should have considering that they are able to directly sense change in the environment and signal this

change to the intracellular molecular machinery. My hypothesis relies on the central role of TCS, which could be the main factors underling the response to a stress condition. As a consequence, global regulators can be activated and modulates their targets, in dependence of the external and internal stimuli (**Figure 5.7**).

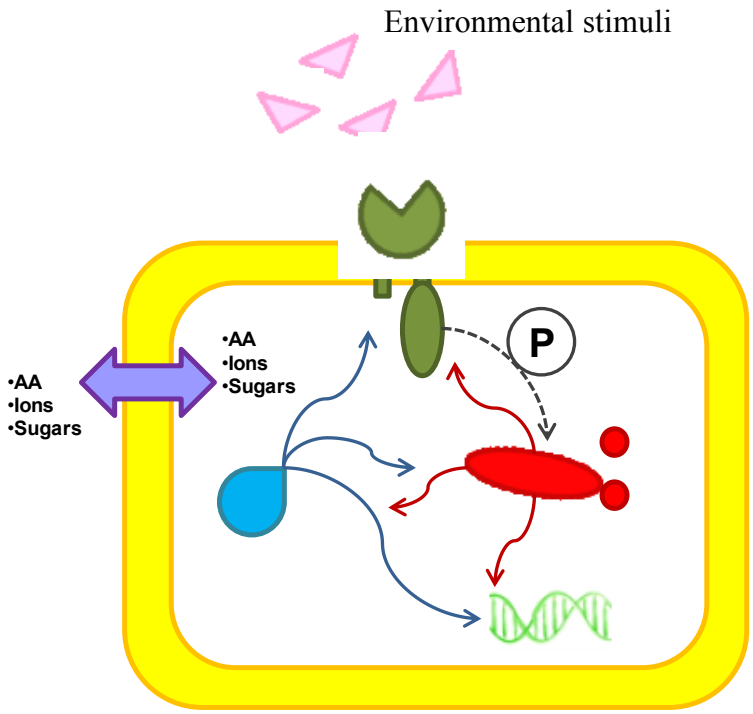


Figure 5.7 Sensing stimuli by *E. coli*.
OCS (in blue) are major regulators which allow the modulation of several functions in *E. coli*, following changes in the cytoplasmic environment. TCS (HK in green and RR in red), directly evolved from OCS (Wuichet et al., 2010), can sense the external stimuli and therefore activate the canonical events cascade. The regulation of the main cellular functions could not be attributed only to one or the other molecule, but each of them can give its contribution for maintaining the cellular physiology.

The SSM model identified experimentally validated interactions between Lrp and TCS. The Leucine Responsive regulatory Protein is one of the global regulators of *E. coli*, which could be potentially activated by TCS for activating the regulatory cascade of events.

Stress responses require a fine mechanism of regulation, which is based on the cooperation of several components. The ability to respond to environmental changes is also

dependent on how the microorganisms can adapt to those changes. Since TCS are primarily involved in sensing the external stimuli, they can also be considered important components of the adaptive responses to stress conditions. The ARACNE model could identify several TCS regulators potentially involved in detecting stimuli; hence, the model could be used as a platform for mapping transcriptional regulations in different environmental conditions.

5.4.2 OmpR, DcuR and YehT are the key regulators of the response of *E. coli* MG1655 to acid stress

During acid exposure, the inner membrane of *E. coli* helps to avoid further acidification of the cytoplasm (Foster, 2004). The membrane has in principle another important role: it senses change and triggers the resistance regulatory mechanism via TCSs.

Previous studies have demonstrated that cytoplasmic sensors could respond to changes in internal pH, by cooperating with systems on the membrane, involved in sensing external pH changes (Slonczewski et al., 1987). Between the pH sensing components, some TCSs are able to detect external pH or cytoplasmic pH(Krulwich et al., 2011). In *H. pylori*, for example, the TCS ArsRS probably responds to cytoplasmic pH changes (Casey et al., 2010). Other TCSs detect changes that can possibly be anticipatory to a change in pH. For example, in *E. coli* the reduction of trimethylamine N-oxide (TMAO) forms trimethylamine, which can alkalinize the cytoplasm. The TCS TorSR senses TMAO and induces the tryptophanase TnaA, which produces acid to counteract the cytoplasmic alkalinisation (Padan et al., 2005). Considering acid response, I have previously described

the TCS involved in this mechanism: EvgA/EvgS and PhoP/PhoQ (Eguchi et al., 2007; Eguchi et al., 2011).

Sensing external stimuli can involve also events of cross-talking within TCS. The mechanism of action of a TCS is based on the molecular specificity with the interacting domains in the histidine kinase and cognate response regulator (Laub and Goulian, 2007; Skerker et al., 2008). When those domains are constituted by similar sequences in the TCS, the cross-talking and phosphorelay events can happen in couples of different TCS, such the case of EnvZ/OmpR and CpxRA (Siryaporn and Goulian, 2008).

The gene regulatory hierarchical structure from the SSM model identified several connections between TCS (sensors and regulators), which could underlie potential cross-talking regulatory events during acid stress conditions. OmpR and CreB were located on the top of the hierarchical structure, which led to the idea that OmpR (possibly with CreB?) could be the TCS activating the events cascade for acid adaptation in *E. coli* MG1655 (**Figure 5.4**). In the gene network, OmpR was connected with YehT, with a negative interaction. OmpR gene expression was highly induced after few seconds of exposure in the time course data described in chapter 4, while *yehT* expression profile was down-regulated after 5 minutes of acid exposure.

On the basis of the hierarchy of the model structure, OmpR could have an inhibitory effect on YehT gene expression. In the model, YehT was negatively connected to RstA; *rstA* was up-regulated during acid exposure (**Figure 5.8**). Therefore the putative inhibition exerted by YehT on RstA was not seen, because the regulator was down-regulated itself. It is known that RstA is an important regulator of curli fimbria formation, anaerobic respiration but also acid tolerance (Ogasawara et al., 2007). Following the model structure, the interaction between RstA and DcuR was negative: if *rstA* was up-regulated,

dcuR expression should have been decreased. Indeed, dcuR expression is highly induced after few seconds of exposure and lately inhibited, when *rstA* expression increased.

Knowledge about these potential connections and their involvement in stress responses is poor to date. However, a new interaction between the regulator OmpR and the TCS YehTU in *S. enterica* under hypotonic conditions was recently discovered (Zhang et al., 2012).

OmpR gene expression was induced in the MG1655 strain, while it was down-regulated in the BW25113 strain. In this strain the connection between OmpR and switch towards the anaerobic respiration was evident, since in a Δ*ompR* background BW25113 was unable to respond to the stress. In consideration of the MG1655 strain, the up-regulation of this RR could also be connected with the increased expression of the aerobic respiration encoding genes.

The regulation of acid resistance does not only involve the canonical response based on the ARs because of the substantial changes at the cell system level. Acid and general stress conditions involve several changes in the cells, which could either bring to discover new acid response systems or to understand that ARs could be only a part of a tuned mechanism underlying the bacterium homeostasis. The balance within the different components and their modulation is probably achieved through several processes involving central metabolism components and the main regulators.

5.5 Conclusions.

The role of TCS during stress conditions is still unclear and a great challenge for future research. However, several models were created to address questions arising from recent findings on this field. In the last part of my PhD I have applied System Biology approaches

and I discovered that OmpR, YehT and DcuR could be potentially involved in the mechanism of acid adaptation in *E. coli* MG1655.

Chapter 6: A comparison between molecular responses to different sources of stress

6.1 Introduction

The response of biological systems to environmental perturbations is characterised by the adjusting of physiology on every level of the cellular and molecular network. The stress response is either characterised by specific responses to a stress condition and general responses, with the down-regulation of genes encoding translational pathways (Jozefczuk et al.) RpoS is the central regulator *of E. coli* general stress response and can modulate the expression of several genes, including the genes involved in metabolism, stress adaptation, transport and transcription regulation (Weber et al., 2005). In the previous chapters I have described the response to mild acidic conditions in BW25113 and MG1655 strains. I found that pH 5.5 exposure had different effects on the central metabolism in the two strains considered. Here I am going to briefly describe an analysis performed on a transcriptomic dataset from Jozefczuk et al. My question was that if acid conditions can have strong effects on the central metabolism of *E. coli* MG1655 and BW26113, other stress conditions might induce other responses. In order to address this question, I have compared the effects of mild acid exposure (pH 5.5) on the expression of central metabolism genes in several stress conditions (cold, heat and oxidative stress) from a previous work by Jozefczuk et al..

6.2 Materials and Methods

6.2.1 Datasets and data analysis

Acid exposure: I have considered the datasets as previously described in Chapters 2, 3 and 4.

Cold, Heat and Oxidative stress: The dataset were originally published by Jozefczuk et al. I have performed VSN normalization as described by authors and performed Significant Analysis for Microarrays of each stress condition with the physiological conditions.

6.3 Results and Discussion

6.3.1 Central metabolism changes in response to different stress conditions.

The main result emerging from the BW25113 analysis is the inhibition of the aerobic associated pathways (**Figure 6.1**) and the up-regulation of the anaerobic and respiratory enzymes during acid exposure (Chapter 2).

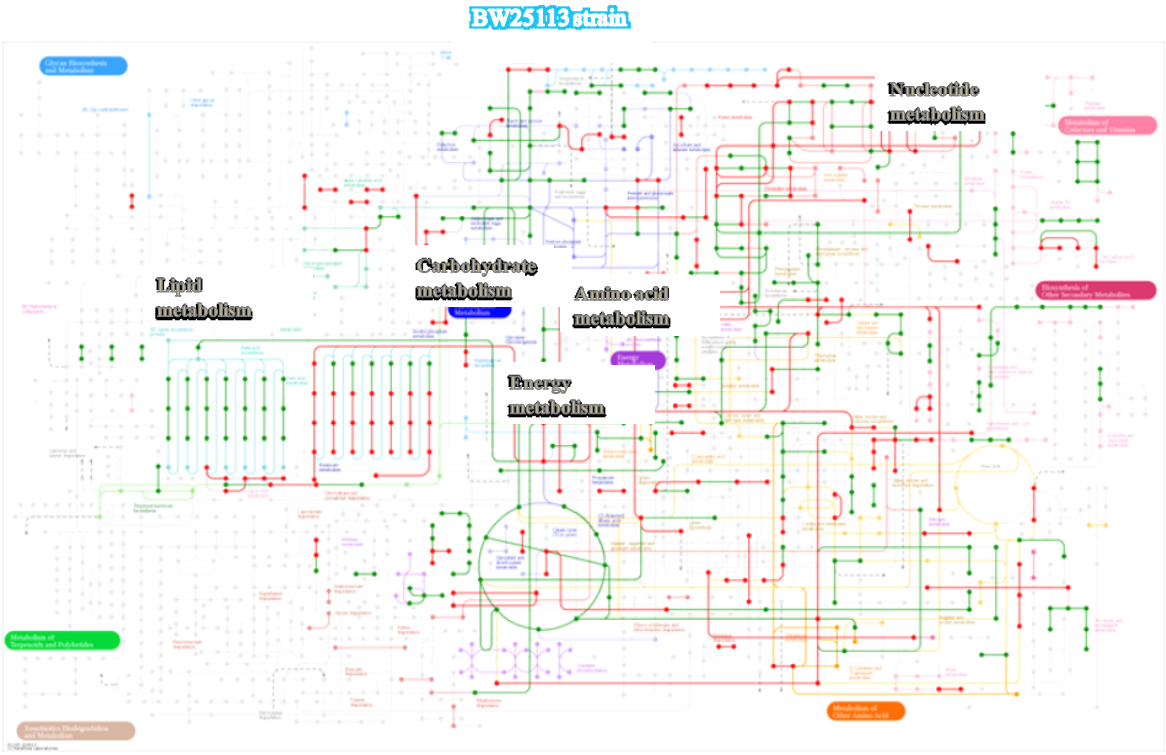


Figure 6.1 Central metabolism gene regulation during acid adaptation in *E. coli* BW25113 (Kegg Pathways).
The identified pathways are highlighted: in red up-regulated and green down-regulated. The effects of acid on BW25113 bring the cells to switch to anaerobic respiration, with consequent inhibition of TCA cycle and oxidative phosphorylation.

The nucleotide metabolism was following a mixed trend while the fatty acid biosynthesis was mostly repressed. However the general behaviour of the strain, at the transcriptomic level, was clearly in accordance with the principle of the energy conservation,

demonstrated in earlier studies (Weber et al., 2005) and described in Jozefczuk studies. The energy conservation program is an important feature for all stress response and involves the repression of genes involved in cells division and aerobic metabolism (Jozefczuk et al.). Different effects of acid treatment were seen in the MG1655 strain, which response was showing the reverse trend seen in the BW25113 (**Figure 6.2**).

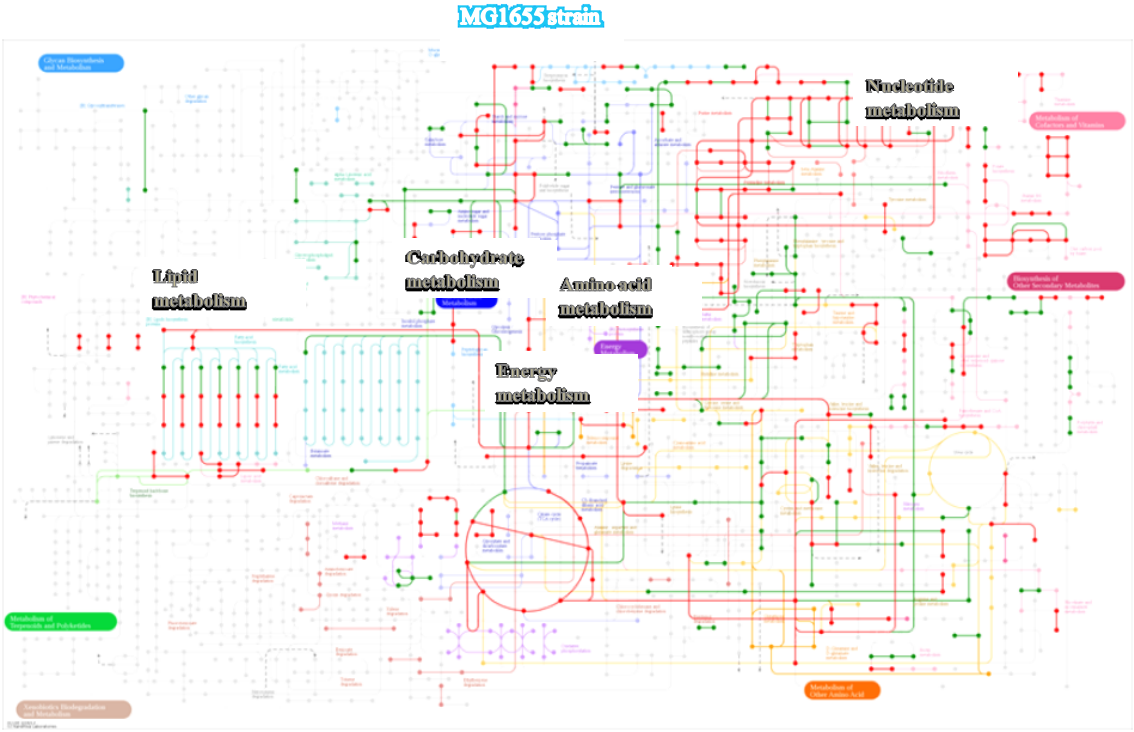


Figure 6.2 Central metabolism gene regulation during acid adaptation in *E. coli* MG1655 (Kegg Pathways).

The identified pathways are highlighted: in red up-regulated and green down-regulated. The effects of acid in MG1655 are opposite to those previously observed in BW25113. The pH shift increased the expression of the aerobic respiration genes.

In accordance to other analysis performed in several stress conditions, the energy metabolism can be modulated on the basis of the external environment. I have considered three stress conditions from a recent study: cold, heat and oxidative stress (Jozefczuk et al., 2010) (**Figures 6.3, 6.4 and 6.5**).

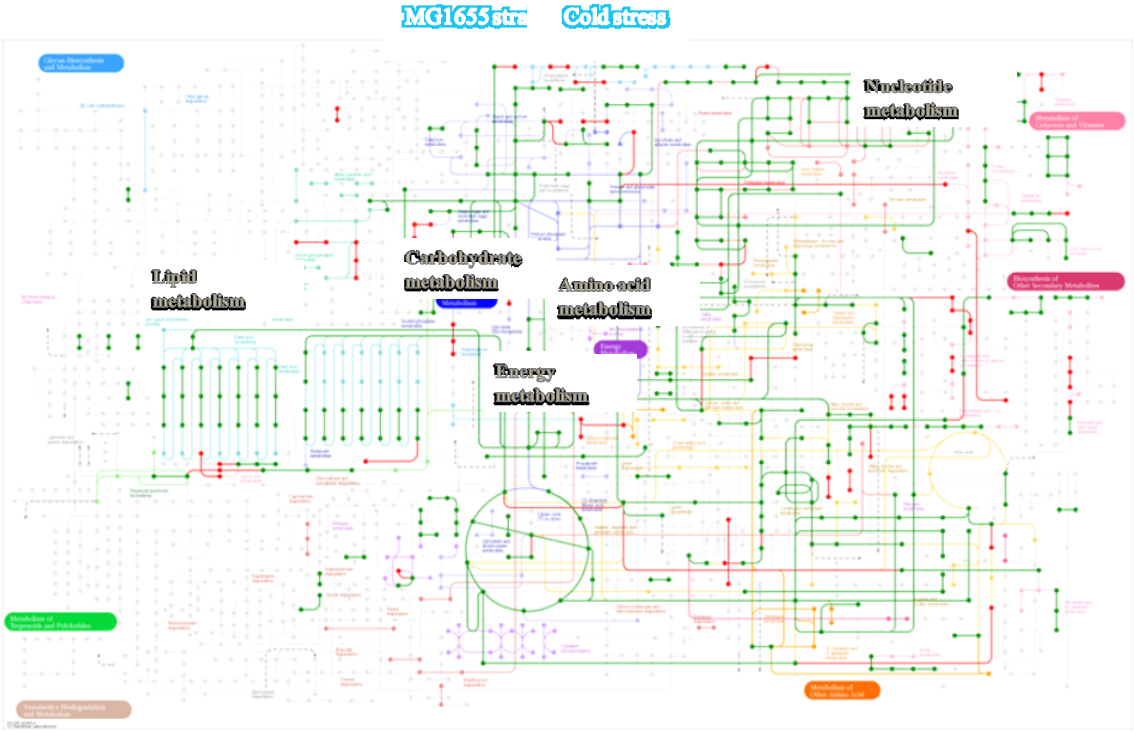


Figure 6.3 Central metabolism gene regulation during cold stress conditions in *E. coli* MG1655 (Kegg Pathways).

The identified pathways are highlighted: in red up-regulated and green down-regulated. The response to the stress brought the cells to switch off the aerobic pathways.

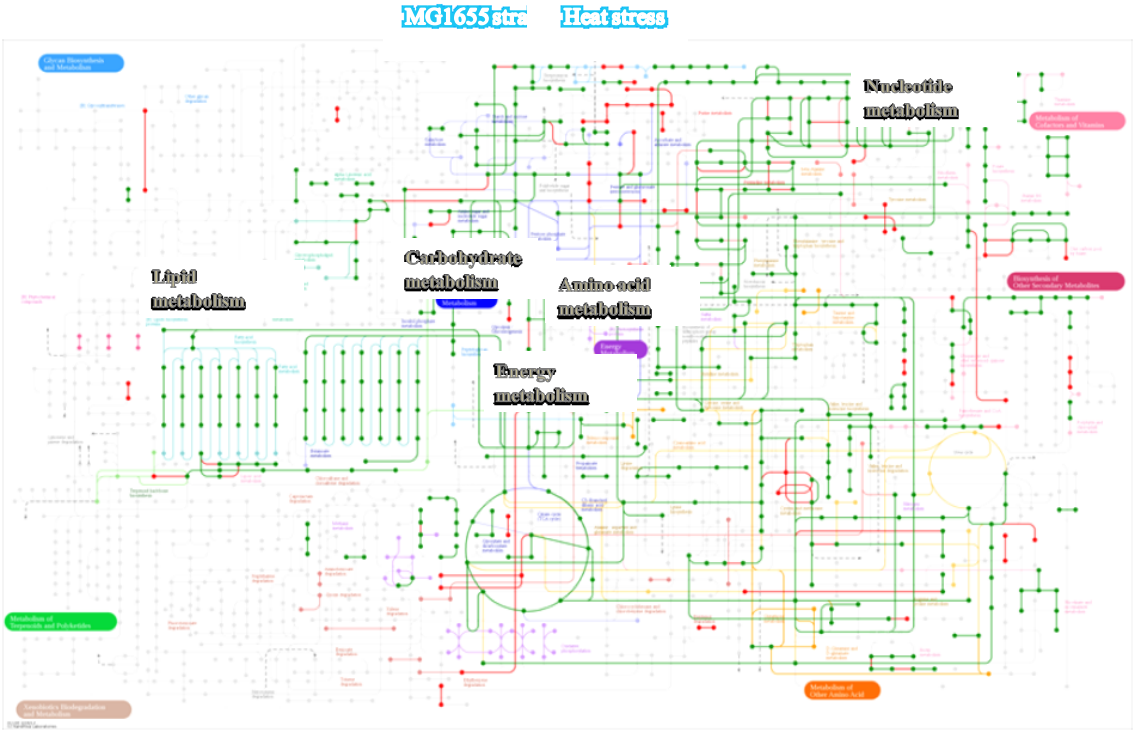


Figure 6.4 Central metabolism gene regulation during heat stress conditions in *E. coli* MG1655 (Kegg Pathways).

The identified pathways are highlighted: in red up-regulated and green down-regulated. As for the cold stress, the aerobic functions are strongly repressed, in agreement with the general principle of energy conservation.

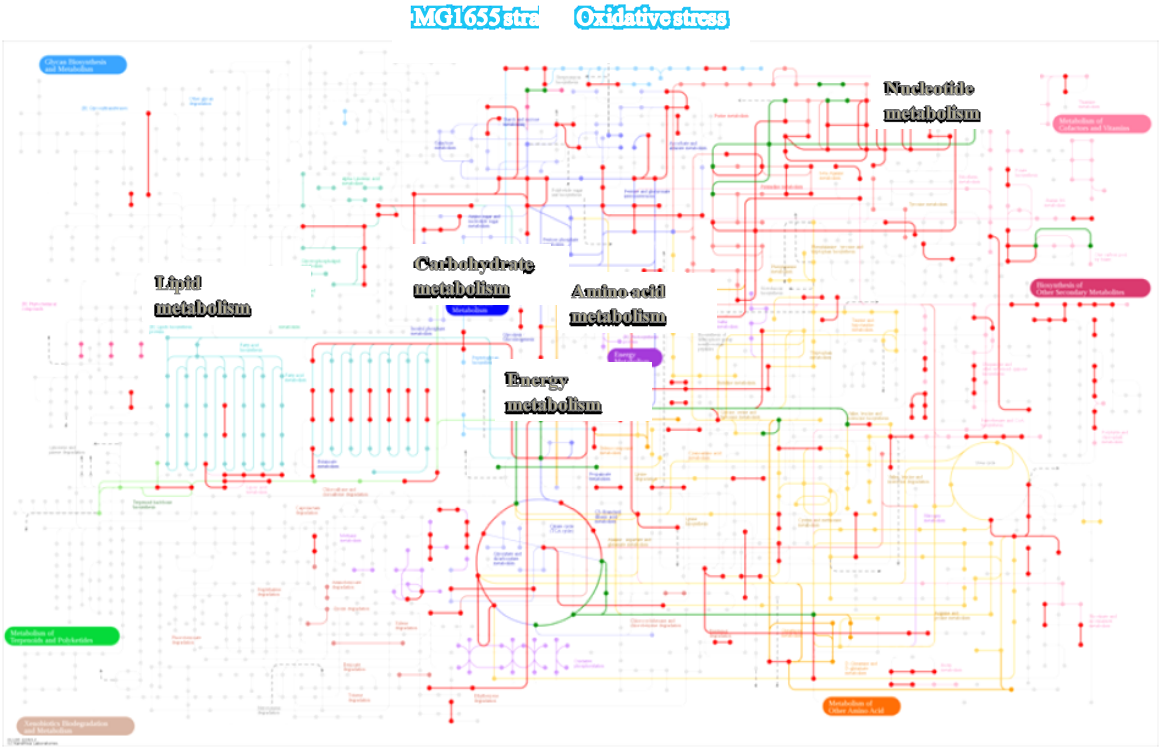


Figure 6.5 Central metabolism gene regulation during oxidative stress conditions in *E. coli* MG1655 (Kegg Pathways).
The identified pathways are highlighted: in red up-regulated and green down-regulated. Opposite behaviour of the strain during oxidative stress. Most of the central metabolism functions were highly induced in behalf of the aerobic switch.

The temperature stress conditions (cold and heat) were characterised by the same regulatory trend, similar to the behaviour previously observed for the BW25113 strain during acid adaption. The genes encoding for the main metabolic aerobic functions were down-regulated. Different considerations could be envisaged for the oxidative stress which involved a higher level expression of the main metabolic pathways. These results showed that the strategies used by *E. coli* in response to environmental stress can involve mechanisms which are dependent on two main aspects: the strain and the stress. Considering a single stress condition, several other aspects need to be addressed, such as the growth phase, the medium of growth etc. Therefore, describing a general trend of

response is not possible, since even a little change in the external conditions can easily turn over the expected pattern of regulation for a given stress response.

Chapter 7: General Discussion

7.1 A new perspective on acid resistance in the *Escherichia coli* BW25113 and MG1655 strains.

7.1.2 An integrative biology framework to understand acid response

E. coli response to acid exposure has been intensively studied. Although research has mainly focussed on the four ARs (Foster, 2004), some studies reported that a consequence of exposure to acid is the transcriptional modulation of genes involved in metabolic and energy pathways (Maurer et al., 2005, Hayes et al., 2006, Kannan et al., 2008) and that this feature may be conserved in different strains (King et al., 2010).

Our work has shown that the ompR dependent modulation of central metabolism pathways is an important determinant of acid resistance in the *E. coli* BW25113 strain. We also discovered that the response of the related strain MG1655 is very different but in principle still regulated by OmpR.

A number of questions however remain open. In the future it will be important to develop a model of acid response that explains the divergent and complex molecular response to acid exposure in different E coli strains. It is also of interest to ask whether the new mechanism we have discovered may play a role in other stress responses.

7. 1.2 Characterization of the canonical acid stress response ARs in E. coli BW25113 and MG1655 strains

In order to define the differences between the strains considered in my work, I have firstly considered the mechanisms of the canonical response, based on the contribution of

the ARs. The mechanism of acid resistance has been intensively studied during stationary phase conditions (Small et al., 1994). At this growth phase, the regulatory events characterising acid resistance events allow the cells to survive at lower nutrients concentrations and as a consequence further stress conditions (Siegele and Kolter, 1992). Additionally, RpoS is directly involved in the whole process, since it is the main stress regulator of the stationary phase and also the pre-stationary phase(Hengge, 2009). Furthermore it is also known to regulate the GAD system. The induction of *rpoS* expression is rather a consequence of pH drop, than an early stationary phase entry (Ma et al., 2002). The GAD system genes are known to be up-regulated in the presence of RpoS. In my experimental conditions they were effectively highly expressed after acid exposure in the MG1655 strain.

Nevertheless, some intriguing points were found about the gene encoding the AraC-like regulator GadX in the BW25113 strain; in effect this regulator was repressed in response to acid exposure based on cluster analysis. A similar result was already observed in previous studies, i.e. the regulator of acid response, GadW, which activates the GAD system together with GadX, it could also have an inhibitory effect on *gadX* (Tucker et al., 2003; Tramonti et al., 2006). Moreover, this repression can involve the genes coding for the glutamate decarboxylase (*gadA* and *gadB*) and the glutamate/GABA antiporter (*gadC*) (Ma et al., 2003; Tucker et al., 2003). Interestingly, the cluster analysis performed on BW25513 time course data showed that, between the genes differentially expressed, *gadA*, *gadB*, *gadC*, *gadX* were located in the down-regulated clusters. The expression of the gene coding for GadW was instead induced under the same conditions, together with *gadE*. In the *E. coli* K-12 MG1655 strain, the GAD genes together with *rpoS* were following the trend classically acknowledged in the literature of acid response, that is to say that they

were highly expressed after few minutes of exposure. I demonstrated that the two considered strains showed different ARs regulation under the same culture conditions.

One aspect common to both considered strains was the inhibition of the genes belonging to the *atp* operon. The F_OF_1 ATPase is supposed to be involved in the mechanism of acid resistance, under the control of RpoS and CRP, in the AR1. The action of the pump was observed in strongly acidic conditions (pH 2-2.5) and not detected in mild acidic conditions (Richard and Foster, 2004). The inhibition of the gene encoding for the ATPase could be important for a proton scavenging strategy under mild acidic conditions, in order to avoid further acidification of the cytoplasm.

. Some studies have demonstrated that various genes contribute to acid resistance (Foster, 2004). Carbon dioxide, a substrate for nucleic acid and amino acid biosynthesis, induces acid resistance (Sun et al., 2005). Hence Sun et al. assumed that nucleotide biosynthesis could also be important for acid resistance. They examined the participation of purine nucleotide biosynthesis in survival under acidic conditions and found that the genes *purA* and *purB*, encoding for the adenylosuccinate synthetase and lyase respectively, have been observed as increasing factors of the ability of *E. coli* to survive acid stress conditions (Sun et al., 2011). This is in contrast to my data in which I was not able to identify any differential expression pattern related to those 2 genes in the strains considered.

Three of the canonical ARs are amino acid dependent: glutamate, arginine and lysine. The amino acids are an essential requirement for acid response, as their decarboxylation reduces the intracellular H^+ concentration via the production of CO_2 and their corresponding amine (Bearson et al., 1998, Foster, 2004). Metabolomics data obtained from the MG1655 strain during the time course in acid conditions showed a considerable

increase of some amino acids such as threonine, glutamate, leucine, valine and isoleucine. The increase of amino acids was found in other stress responses (Jozefczuck et al., 2010) and is a characteristic of the general and the stringent stress response (Weber et al., 2005). Valine production was observed during biofilm formation in continuous cultures (Valle et al., 2008). In a recent study performed during heat stress conditions, the increases of lysine, glutamate, isoleucine, leucine and valine, were observed in *E. coli* HB101, associated to the decrease of betaine and putrescine (Ye et al., 2012); I also observed this phenomenon in my data. It is then possible that *E. coli* acid resistance would require more amino acids for the mechanism to take place, hence the involvement of more ARs. Since the enhancement of intracellular amino acids quantities is not only a characteristic of acid response, this enhancement might also be a requirement for the general stress response in *E. coli*.

7.2 Overlap between osmotic and acid stress responses

The colonization of the gut could require *E. coli* to adapt to several stress conditions, such as pH drop but also anaerobic conditions and osmotic changes. Exposure of *E. coli* to high osmolarity results in a loss of water and shrinking of the cells (Weber et al., 2006). The first response to osmotic changes includes the uptake of potassium, which result in secondary mechanisms: accumulation of glutamate, synthesis of trehalose and release of putrescine (Weber et al., 2006). Some osmotic-shift induced proteins were identified and between them was OmpC, an outer membrane protein controlled by the TCS EnvZ/OmpR (Hall and Silhavy, 1981)

One aspect in common between the BW25113 and MG1655 strains was the down-regulation of the genes encoding the main sugar metabolism, transport pathways and the

PTS (Phosphotransferase system) pathway. Metabolites data collected from the MG1655 strain, showed a huge decrease of the UDP-glucose levels during acid exposure. The UDP-glucose is a precursor of the lipopolysaccharide, which are major constituents of the outer leaflet of the outer membranes in *E. coli* (Wang et al., 2006)(Wang and Quinn, 2010). In addition, it was suggested that this molecule can exert another important function in response to osmotic stress conditions (Kempf and Bremer, 1998). *E. coli* has developed many mechanisms in order to respond to osmotic stresses: for example, when glycine betaine is not available in the external environment, the production of trehalose is preferred to overcome this adverse condition (Kempf and Bremer, 1998). Trehalose is a compound which helps the cells to maintain the integrity of their membranes (Crowe et al., 1984). The molecule is produced by de-novo synthesis from glucose-6-phosphate and UDP-glucose when glycine betaine and proline are not available (Kempf and Bremer, 1998). The reaction is catalysed by the OtsBA enzyme, the trehalose-6-phosphate synthase, a mutation of which can highly affect the cells during stationary phase of growth (Hengge-Haronis et al., 1991). The decrease of UDP-glucose in the metabolites analysis of MG1655 revealed that this molecule might be used for increasing the concentrations of trehalose in response to the stress, since glycine betaine levels were really low and the corresponding transporter genes (*proVWX*) were also strongly repressed during the acid time course. Moreover, after 5 minutes of exposure, MG1655 showed an increase in the expression of *otsB*, while in BW25113 both *otsAB* were induced after 15 minutes of exposure. The results confirmed the previously discovered cross protection within several stress responses exerted by the mechanisms of acid resistance. However, it is not possible to explain why *E. coli* needs to silence the classic response to osmotic stress through the glycine betaine compound, by preferring an alternative pathway, which could probably involve trehalose and several

mechanosensors. My previous work also showed that the strain BW25113 can induce the expression of mechanosensors and aquaporins, which underlie the response to hyper-osmotic stress (Stincone et al., 2011).

7.3 Two-component systems: sensors of the general stress response

TCS allow *E. coli* and enteric bacteria to respond quickly and efficiently to a large scale of external stimuli and conditions. Moreover, the HK, which can be located either on the external or on the cytoplasmic side of the membrane, can sense a wide spectrum of signals, in association with other components, such as the OCS.

The knowledge about regulatory networks involving TCS and main components required for stress responses is rapidly expanding (DeRouchey and Rau, 2011). For example, the general stress RpoS factor expression is induced by the PhoP/PhoQ intervention, together with the BarA/UvrY TCS (Battesti et al., 2011). Moreover, RpoS degradation is controlled by the protease RssB, which can be activated by ArcA/ArcB (Mika and Hengge, 2005). This TCS can also inhibit the expression of *rpoS* under some stress conditions (Mika and Hengge, 2005). The main implication of TCS regulating expression and proteolysis of RpoS relies on their fundamental role not only in detecting the stimuli, but also to contribute with the main global regulators to promptly activate the mechanisms helping the cells to survive.

The network obtained in ARACNE from a compendium of wide spread conditions data identified a large number of connections between the TCS regulators and the rest of the genes of *E. coli* (Chapter 5). What is the relation underlying the ARACNE network and the

big matrix in RegulonDB? The main aspect to consider when looking at gene inference data is the fact that genes are connected on the basis of statistical analysis and not on experimental evidence. Therefore, even though the models previously described (ARACNE and SSM) could show some contradictions, it is important to consider that the hierarchical structure of RegulonDB reports interactions experimentally validated, therefore having the global regulators on the top of the structure (**Figure 5.6**) is obviously related to their ability in controlling several functions in *E. coli*. What is the triggering event activating the global regulators? And on the basis of which relationship do they modulate different functions? Most of the important metabolic transcription factors are OCS, therefore they can sense cytoplasmic changes and induce the cascade of transcriptional events (Ulrich et al., 2005). In the presence of an external stimulus, the HK, which autophosphorylates, activates the correspondent transcription factor. Interactions of the transcriptional and phosphorylative events can also enhance and spread the response. In the mutual information network, many interactions between regulators and sensing proteins were identified. The ARACNE network could be considered as a prospective tool which enables us to understand the targets potentially modulated in response to each stimulus.

In the network, the node showing the highest number of connections, UvrY, is an important regulator of the switch between glycolytic and glycogenic metabolism (Pernestig et al., 2003). Most of the genes connected to UvrY in the mutual information network were involved in central metabolism functions. The TCS BarA/UvrY has many orthologs in several pathogenic bacteria, for example BarA/SirA in *Salmonella enterica* (Goodier and Ahmer, 2001), which was found involved in regulating *Salmonella* virulence(Teplitski et al., 2003). It has been recently reported also the contribution of UvrY in the

Uropathogenic *E. coli* during the infection process (Palaniyandi et al., 2012). Not only mutation in the TCS decreased the virulence in a mouse model, but they also reduced the ability to infect the uroepithelial cells. The same TCS was controlling virulence of another *E. coli* strain, the avian pathogen O78:K80:H9, which was unable to infect in vivo when the TCS was removed (Herren et al., 2006). UvrY and OmpR have an important role in controlling virulence in both *Salmonella* and *E. coli*.

These findings confirm the great potential of the models previously described in the present work, which both could discover the available connections known in the literature and identify regulators which great potential involvement in stress responses, which is currently under investigation in several research projects.

In a previous study, the gene KO of *E. coli* TCS were analysed for global characterization purposes (Oshima et al., 2002). I have considered those data in order to assess which mutants were showing a similar transcriptomic profile (**Figure 7.1**) (Data were downloaded from http://ecoli.naist.jp/Lab/joomla/index.php/jp/downloads-jp, a Principal Component Analysis was performed as described in Chapter 2, section Materials and Methods).

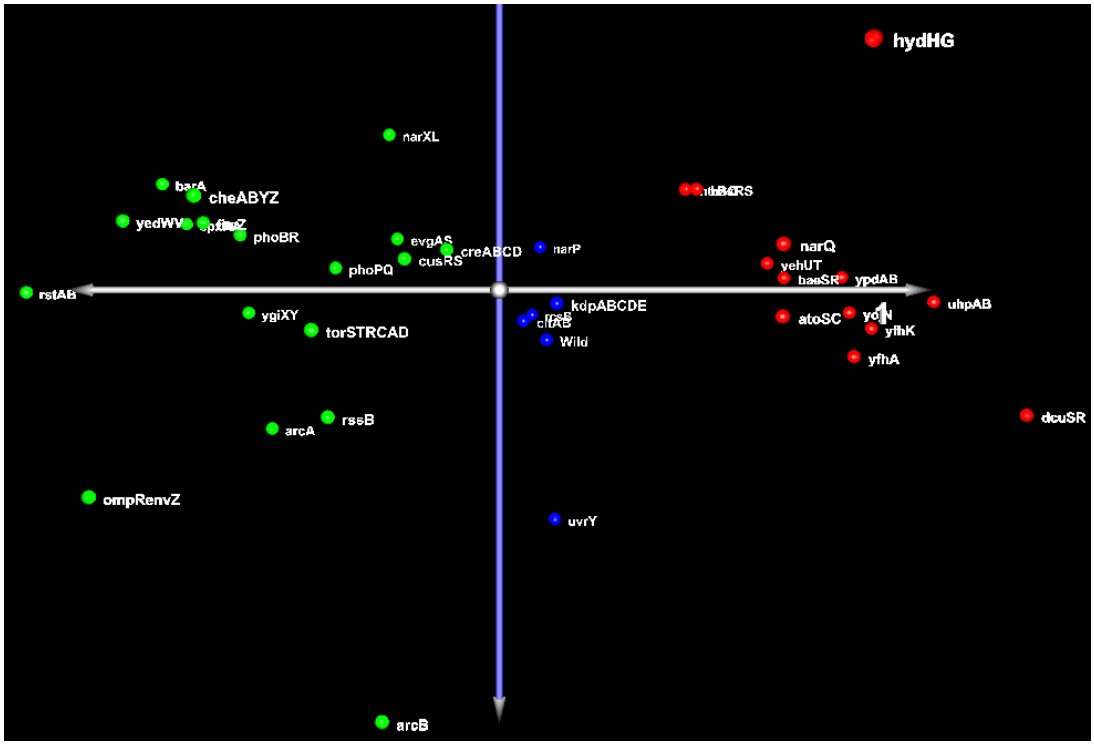


Figure 7.1 PCA plot of the TCS gene KO.

3 clusters were obtained after the analysis; one of them was containing few TCS and the Wild type (blue dots). Interestingly, the mutation in *uvrY* did not very much affect the transcriptomic profile of the bacterium. Oppositely, the mutants of OmpR/EnvZ (green cluster) and DcuRS (red cluster) were shifted across the first component, denoting a strong dissimilarity compared to the wild type (Oshima et al., 2002).

Three clusters were identified in this analysis, with the wild type belonging to the middle one. Interestingly, the mutation in UvrY did not result in a substantial difference when compared to the wild type. OmpR, YehT and DcuR were instead in two different clusters, denoting the big effects of these mutations on the bacteria. The functional analysis of the genes differentially expressed between the three clusters identified that the main differences belong to the regulation of the cellular respiration, ribosome and translational pathways and the flagellar biosynthesis. Further studies will be required for understanding and characterising the differences between the mutant strains.

The role of the TCS in both eukaryotes and prokaryotes is involving different aspects of research; one interesting approach is their use for construction of synthetic genetic networks.

Recently there was an increase in the research for production of bacteria which can sense presence of various compounds and conditions. The use of synthetic networks is also important for the creation of metabolic pathways, in order to produce molecules and compounds with fermentative processes. The use of biosensors based on the TCS scheme has increased the quality of the products and overcame the limitations due to imbalanced metabolism of the cells (Zhang and Keasling, 2011)

7.4 The importance of the studies on microorganism's behaviour during stress conditions

In the last years the role of food industry became predominant for the increasing demand of quality products and market globalization (Matsa, 2011). The best approaches to obtain high quality standards in food production require regular collaboration between scientific organizations, states and industry, and frequent updates for the control procedures (Taylor, 2011). One of the aspects underlying the safety of food processing is related to the impact of contamination by pathogenic microorganisms and following studies to circumvent this problem, for example trough multi-factorial risk assessments (Ruzante et al., 2010).

The understanding of the microorganism's behaviour is important not only for food safety considerations but also in order to prevent and fight pathogen outbreaks and to improve the fermentative processes (Abee and Kuipers, 2011). A book to control safety parameters in food production by understanding their physiology has been recently

published (McMeekin et al., 2010). New approaches and technologies to study the pathogens and microorganism in general are useful tools to amend these parameter (Abee and Kuipers, 2011). During fermentative processes, the control of the cultures parameters is important in order to achieve good quality products. During food processing, combinations of different stresses can take place. It has been observed that for example a mild re-heating (54–62 °C) of products with reduced pH due to pre-drying acidic treatments (meat products) could increase the resistance of bacteria (Calicioglu et al., 2003, DiPersio et al., 2003, Faith et al., 1998, Yoon et al., 2006) Conventional food processing technologies, like pasteurization, heating, drying, are gradually replaced by new technologies with the production of products of higher quality, therefore higher nutritional and sensory characteristics. Nevertheless, when technologies such high pressure processing, pulsed electric field are applied, the food products could be exposed to milder environmental conditions, which can bring to a specific stress response and further tolerance via cross protection (Velliou et al., 2011). Hence, the comprehension of microbial adaptation when applying innovative technologies is also of great importance for safety parameter selection (Yousef and Courtney, 2003).

The analysis of the food production chart is a useful tool for considering the different conditions which bacteria can encounter during the process. A recent study analysed the various stress conditions encountered by pathogenic *E. coli* during the cheese production process from raw milk (Peng et al., 2011). A general flow chart considered for the production process involves several steps of heating and different conditions of humidity, therefore osmolarity and pH parameters can easily vary (**Figure 7.2**).

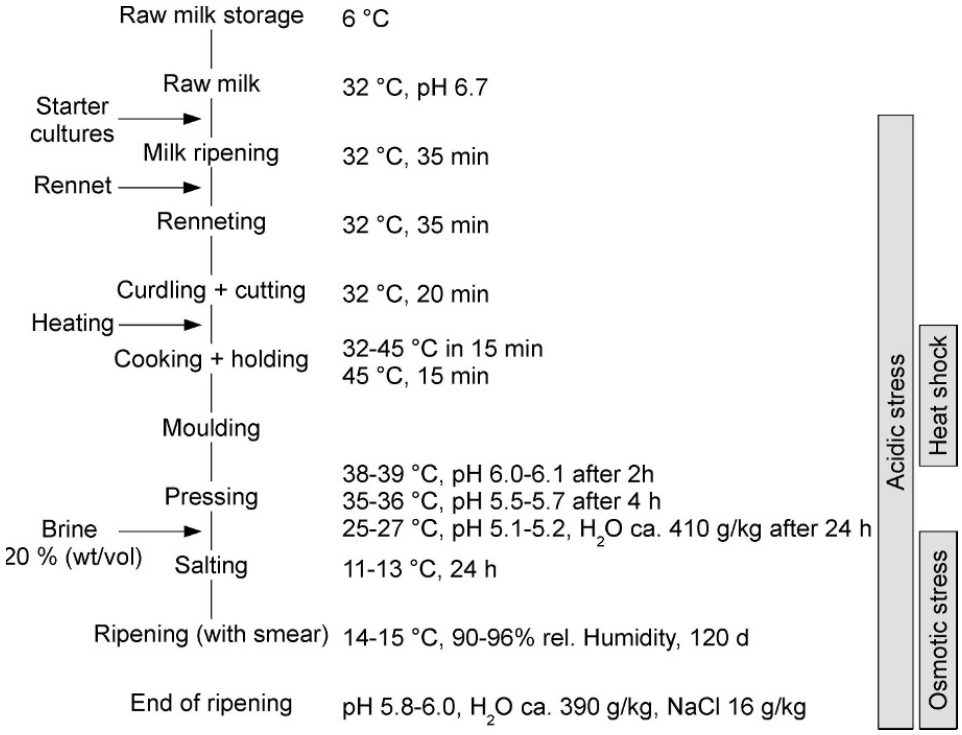


Figure 7.2 Flow chart for production of a typical semi-hard cheese

For each production step are shown the temperatures used and the stress conditions encountered by the bacteria. Acid stress covers most of the process (Peng et al., 2011).

Cheese production involves a lot of stresses which can happen simultaneously; therefore cross-protection can affect the responses of bacteria (Peng et al., 2011). A key factor for the general stress response, as previously described is RpoS, involved in protection from acid stress. The cross-protections showed that in sub lethal conditions for a stress can bring protection and/ or cross protection for another stress (Chung et al., 2006). Hence, the control of the timing of the different stress applications could be a good parameter for avoiding resistance of pathogens during food processing (Montet et al., 2009). Moreover the expression of the virulence genes is highly influences by the stress conditions, for example, GadE, the GAD central regulator, also down-regulates locus of enterocytes effacement expression under moderately acidic conditions (Kailasan Vanaja, 2009).

The understanding of the mechanism involved in the resistance of bacteria to stress conditions becomes therefore fundamental for several aspects such as health, economic and safety in our society. My PhD project was focused in a global understanding of the processes underlying acid resistance in two strains, which can have two main consequences for further studies. Firstly, my experiments were performed in early stationary phase which is usually considered for food processing and infectious diseases studies (King et al., 2010). The classic knowledge, based on the ARs mechanisms, is not enough for the explanation of the regulative mechanisms happening in the cells during acid exposure. Gene inference approaches, required for generating hypotheses at system level, can identify several regulators involved in the regulatory processes. I believe that any study involving resistance to environmental stress conditions should look at the whole cell, more than on single functions or transcriptional events. Systems Biology approaches can be generally applicable to the analysis of stress condition and can have a real impact on improving food processing protocols, decrease the costs of manufacturing and reduce risks of food contamination.

LIST of REFERENCES

ABDEL-HAMID, A. M., ATTWOOD, M. M. & GUEST, J. R. 2001. Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli. *Microbiology,* 147**,** 1483-98.

ABEE, T. & KUIPERS, O. P. 2011. Understanding microbial behavior within and outside the host to improve food functionality and safety. *Curr Opin Biotechnol,* 22**,** 133-5.

AKITAKE, B., ANISHKIN, A. & SUKHAREV, S. 2005. The "dashpot" mechanism of stretch-dependent gating in MscS. *J Gen Physiol,* 125**,** 143-54.

ALTEKRUSE, S., COHEN, M. & SWERDLOW, D. 1997. Emerging foodborne diseases. *Emerging infectious diseases,* 3**,** 285.

ALTERI, C. J., LINDNER, J. R., REISS, D. J., SMITH, S. N. & MOBLEY, H. L. 2011. The broadly conserved regulator PhoP links pathogen virulence and membrane potential in Escherichia coli. *Mol Microbiol,* 82**,** 145-63.

ALVES, R. & SAVAGEAU, M. A. 2003. Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function. *Molecular microbiology,* 48**,** 25-51.

ANDRES LEON, E., EZKURDIA, I., GARCIA, B., VALENCIA, A. & JUAN, D. 2009. EcID. A database for the inference of functional interactions in E. coli. *Nucleic Acids Res,* 37**,** D629-35.

ANDREWS, S. C., BERKS, B. C., MCCLAY, J., AMBLER, A., QUAIL, M. A., GOLBY, P. & GUEST, J. R. 1997. A 12-cistron Escherichia coli operon (hyf) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology,* 143 (Pt 11)**,** 3633-47.

ARANDA, B., ACHUTHAN, P., ALAM-FARUQUE, Y., ARMEAN, I., BRIDGE, A., DEROW, C., FEUERMANN, M., GHANBARIAN, A. T., KERRIEN, S., KHADAKE, J., KERSSEMAKERS, J., LEROY, C., MENDEN, M., MICHAUT, M., MONTECCHI-PALAZZI, L., NEUHAUSER, S. N., ORCHARD, S., PERREAU, V., ROECHERT, B., VAN EIJK, K. & HERMJAKOB, H. 2010. The IntAct molecular interaction database in 2010. *Nucleic Acids Res,* 38**,** D525-31.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K. A., TOMITA, M., WANNER, B. L. & MORI, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol,* 2**,** 2006 0008.

BACHHAWAT, P. & STOCK, A. M. 2007. Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride. *Journal of bacteriology,* 189**,** 5987-95.

BACHMANN, B. J. 1990. Linkage map of Escherichia coli K-12, edition 8. *Microbiol Rev,* 54**,** 130-97.

BAIKALOV, I., SCHRODER, I., KACZOR-GRZESKOWIAK, M., GRZESKOWIAK, K., GUNSALUS, R. P. & DICKERSON, R. E. 1996. Structure of the Escherichia coli response regulator NarL. *Biochemistry,* 35**,** 11053-61.

BAKER-AUSTIN, C. & DOPSON, M. 2007. Life in acid: pH homeostasis in acidophiles. *TRENDS in Microbiology,* 15**,** 165-171.

BANG, I. S., AUDIA, J. P., PARK, Y. K. & FOSTER, J. W. 2002. Autoinduction of the ompR response regulator by acid shock and control of the Salmonella enterica acid tolerance response. *Mol Microbiol,* 44**,** 1235-50.

BANG, I. S., KIM, B. H., FOSTER, J. W. & PARK, Y. K. 2000. OmpR regulates the stationary-phase acid tolerance response of Salmonella enterica serovar typhimurium. *J Bacteriol,* 182**,** 2245-52.
BANSAL, M. & DI BERNARDO, D. 2007. Inference of gene networks from temporal gene expression profiles. *IET systems biology,* 1**,** 306-12.
BAR-JOSEPH, Z., GIFFORD, D. K. & JAAKKOLA, T. S. 2001. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics,* 17 Suppl 1**,** S22-9.
BARTHOLOMEW, J. W. & MITTWER, T. 1952. The Gram stain. *Bacteriological reviews,* 16**,** 1-29.
BATTESTI, A., MAJDALANI, N. & GOTTESMAN, S. 2011. The RpoS-mediated general stress response in Escherichia coli. *Annu Rev Microbiol,* 65**,** 189-213.
BEAL, M. J., FALCIANI, F., GHAHRAMANI, Z., RANGEL, C. & WILD, D. L. 2005. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics,* 21**,** 349-56.
BEARSON, B. L., WILSON, L. & FOSTER, J. W. 1998. A low pH-inducible, PhoPQ-dependent acid tolerance response protects Salmonella typhimurium against inorganic acid stress. *Journal of bacteriology,* 180**,** 2409-17.
BEIER, D. & GROSS, R. 2006. Regulation of bacterial virulence by two-component systems. *Current opinion in microbiology,* 9**,** 143-52.
BENJAMIN, M. M. & DATTA, A. R. 1995. Acid tolerance of enterohemorrhagic Escherichia coli. *Appl Environ Microbiol,* 61**,** 1669-72.
BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 289-300.
BERGSTROM, L. C., QIN, L., HARLOCKER, S. L., EGGER, L. A. & INOUYE, M. 1998. Hierarchical and co-operative binding of OmpR to a fusion construct containing the ompC and ompF upstream regulatory sequences of Escherichia coli. *Genes to cells : devoted to molecular & cellular mechanisms,* 3**,** 777-88.
BERN, M. & GOLDBERG, D. 2005. Automatic selection of representative proteins for bacterial phylogeny. *BMC evolutionary biology,* 5**,** 34.
BIRCK, C., CHEN, Y., HULETT, F. M. & SAMAMA, J. P. 2003. The crystal structure of the phosphorylation domain in PhoP reveals a functional tandem association mediated by an asymmetric interface. *Journal of bacteriology,* 185**,** 254-61.
BLATTNER, F. R., PLUNKETT, G., 3RD, BLOCH, C. A., PERNA, N. T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J. D., RODE, C. K., MAYHEW, G. F., GREGOR, J., DAVIS, N. W., KIRKPATRICK, H. A., GOEDEN, M. A., ROSE, D. J., MAU, B. & SHAO, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science,* 277**,** 1453-62.
BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.
BOOT, I. R., CASH, P. & O'BYRNE, C. 2002. Sensing and adapting to acid stress. *Antonie Van Leeuwenhoek,* 81**,** 33-42.
BOOTH, I. R., EPSTEIN, W., GIFFARD, P. M. & ROWLAND, G. C. 1985. Roles of the trkB and trkC gene products of Escherichia coli in K+ transport. *Biochimie,* 67**,** 83-9.
BORGNIA, M. J., KOZONO, D., CALAMITA, G., MALONEY, P. C. & AGRE, P. 1999. Functional reconstitution and characterization of AqpZ, the E. coli water channel protein. *J Mol Biol,* 291**,** 1169-79.
BOSSEMEYER, D., BORCHARD, A., DOSCH, D. C., HELMER, G. C., EPSTEIN, W., BOOTH, I. R. & BAKKER, E. P. 1989a. K+-transport protein TrkA of Escherichia coli is a peripheral membrane protein that requires other trk gene products for attachment to the cytoplasmic membrane. *J Biol Chem,* 264**,** 16403-10.

BOSSEMEYER, D., SCHLOSSER, A. & BAKKER, E. P. 1989b. Specific cesium transport via the Escherichia coli Kup (TrkD) K+ uptake system. *J Bacteriol,* 171**,** 2219-21.
BURTON, N. A., JOHNSON, M. D., ANTCZAK, P., ROBINSON, A. & LUND, P. A. 2010. Novel aspects of the acid response network of E. coli K-12 are revealed by a study of transcriptional dynamics. *J Mol Biol,* 401**,** 726-42.
BURY-MONE, S., THIBERGE, J. M., CONTRERAS, M., MAITOURNAM, A., LABIGNE, A. & DE REUSE, H. 2004. Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen Helicobacter pylori. *Mol Microbiol,* 53**,** 623-38.
CALICIOGLU, M., SOFOS, J. N., KENDALL, P. A. & SMITH, G. C. 2003. Effects of acid adaptation and modified marinades on survival of postdrying Salmonella contamination on beef jerky during storage. *J Food Prot,* 66**,** 396-402.
CAPITANI, G., DE BIASE, D., AURIZI, C., GUT, H., BOSSA, F. & GRUTTER, M. G. 2003. Crystal structure and functional analysis of Escherichia coli glutamate decarboxylase. *The EMBO journal,* 22**,** 4027-37.
CASCANTE, M. & MARIN, S. 2008. Metabolomics and fluxomics approaches. *Essays in biochemistry,* 45**,** 67-81.
CASTANIE-CORNET, M. P., CAM, K., BASTIAT, B., CROS, A., BORDES, P. & GUTIERREZ, C. Acid stress response in Escherichia coli: mechanism of regulation of gadA transcription by RcsB and GadE. *Nucleic Acids Res,* 38**,** 3546-54.
CASTANIE-CORNET, M. P., PENFOUND, T. A., SMITH, D., ELLIOTT, J. F. & FOSTER, J. W. 1999. Control of acid resistance in Escherichia coli. *J Bacteriol,* 181**,** 3525-35.
CASTRILLO, J. I., HAYES, A., MOHAMMED, S., GASKELL, S. J. & OLIVER, S. G. 2003. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry,* 62**,** 929-37.
CHANG, L., WEI, L. I., AUDIA, J. P., MORTON, R. A. & SCHELLHORN, H. E. 1999. Expression of the Escherichia coli NRZ nitrate reductase is highly growth phase dependent and is controlled by RpoS, the alternative vegetative sigma factor. *Mol Microbiol,* 34**,** 756-66.
CHUNG, H., BANG, W. & DRAKE, M. 2006. Stress response of Escherichia coli. *Comprehensive reviews in food science and food safety,* 5**,** 52-64.
CICCARELLI, F. D., DOERKS, T., VON MERING, C., CREEVEY, C. J., SNEL, B. & BORK, P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science,* 311**,** 1283-7.
CLARK, D. P. 1989. The fermentation pathways of Escherichia coli. *FEMS Microbiol Rev,* 5**,** 223-34.
CONNER, D. E. & KOTROLA, J. S. 1995. Growth and survival of Escherichia coli O157:H7 under acidic conditions. *Appl Environ Microbiol,* 61**,** 382-5.
CRICK, F. H. 1970. DNA: test of structure? *Science,* 167**,** 1694.
CULHAM, D. E., LASBY, B., MARANGONI, A. G., MILNER, J. L., STEER, B. A., VAN NUES, R. W. & WOOD, J. M. 1993. Isolation and sequencing of Escherichia coli gene proP reveals unusual structural features of the osmoregulatory proline/betaine transporter, ProP. *J Mol Biol,* 229**,** 268-76.
DATSENKO, K. A. & WANNER, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A,* 97**,** 6640-5.
DAUB, C. O., STEUER, R., SELBIG, J. & KLOSKA, S. 2004. Estimating mutual information using B-spline functions--an improved similarity measure for analysing gene expression data. *BMC Bioinformatics,* 5**,** 118.
DE BIASE, D., TRAMONTI, A., BOSSA, F. & VISCA, P. 1999. The response to stationary-phase stress conditions in Escherichia coli: role and regulation of the glutamic acid decarboxylase system. *Mol Microbiol,* 32**,** 1198-211.
DE JONG, H., GOUZE, J. L., HERNANDEZ, C., PAGE, M., SARI, T. & GEISELMANN, J. 2004. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol,* 66**,** 301-40.

DE SMET, R. & MARCHAL, K. 2010. Advantages and limitations of current network inference methods. *Nature reviews. Microbiology,* 8**,** 717-29.
DEININGER, K. N., HORIKAWA, A., KITKO, R. D., TATSUMI, R., ROSNER, J. L., WACHI, M. & SLONCZEWSKI, J. L. 2011. A requirement of TolC and MDR efflux pumps for acid adaptation and GadAB induction in Escherichia coli. *PloS one,* 6**,** e18960.
DEISINGH, A. K. & THOMPSON, M. 2004. Strategies for the detection of Escherichia coli O157:H7 in foods. *Journal of applied microbiology,* 96**,** 419-29.
DENNIS, G., JR., SHERMAN, B. T., HOSACK, D. A., YANG, J., GAO, W., LANE, H. C. & LEMPICKI, R. A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol,* 4**,** P3.
DEROUCHEY, J. E. & RAU, D. C. 2011. Role of amino acid insertions on intermolecular forces between arginine peptide condensed DNA helices: implications for protamine-DNA packaging in sperm. *J Biol Chem,* 286**,** 41985-92.
DIPERSIO, P. A., KENDALL, P. A., CALICIOGLU, M. & SOFOS, J. N. 2003. Inactivation of Salmonella during drying and storage of apple slices treated with acidic or sodium metabisulfite solutions. *J Food Prot,* 66**,** 2245-51.
DUNN, W. B. 2008. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical biology,* 5**,** 011001.
EGUCHI, Y., ISHII, E., HATA, K. & UTSUMI, R. 2011. Regulation of acid resistance by connectors of two-component signal transduction systems in Escherichia coli. *Journal of bacteriology,* 193**,** 1222-8.
EGUCHI, Y., ITOU, J., YAMANE, M., DEMIZU, R., YAMATO, F., OKADA, A., MORI, H., KATO, A. & UTSUMI, R. 2007. B1500, a small membrane protein, connects the two-component systems EvgS/EvgA and PhoQ/PhoP in Escherichia coli. *Proc Natl Acad Sci U S A,* 104**,** 18712-7.
EGUCHI, Y., OKADA, T., MINAGAWA, S., OSHIMA, T., MORI, H., YAMAMOTO, K., ISHIHAMA, A. & UTSUMI, R. 2004. Signal transduction cascade between EvgA/EvgS and PhoP/PhoQ two-component systems of Escherichia coli. *J Bacteriol,* 186**,** 3006-14.
EPPS, H. M. & GALE, E. F. 1942. The influence of the presence of glucose during growth on the enzymic activities of Escherichia coli: comparison of the effect with that produced by fermentation acids. *The Biochemical journal,* 36**,** 619-23.
FAGAN, A., CULHANE, A. C. & HIGGINS, D. G. 2007. A multivariate analysis approach to the integration of proteomic and gene expression data. *Proteomics,* 7**,** 2162-71.
FAITH, J. J., HAYETE, B., THADEN, J. T., MOGNO, I., WIERZBOWSKI, J., COTTAREL, G., KASIF, S., COLLINS, J. J. & GARDNER, T. S. 2007. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biology,* 5**,** e8.
FAITH, N. G., PARNIERE, N., LARSON, T., LORANG, T. D., KASPAR, C. W. & LUCHANSKY, J. B. 1998. Viability of Escherichia coli O157:H7 in salami following conditioning of batter, fermentation and drying of sticks, and storage of slices. *J Food Prot,* 61**,** 377-82.
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.
FORSTER, J., GOMBERT, A. K. & NIELSEN, J. 2002. A functional genomics approach using metabolomics and in silico pathway analysis. *Biotechnology and bioengineering,* 79**,** 703-12.
FOSTER, J. W. 1993. The acid tolerance response of Salmonella typhimurium involves transient synthesis of key acid shock proteins. *Journal of bacteriology,* 175**,** 1981-7.
FOSTER, J. W. 1999. When protons attack: microbial strategies of acid adaptation. *Current opinion in microbiology,* 2**,** 170-4.

FOSTER, J. W. 2004. Escherichia coli acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol,* 2**,** 898-907.
FOSTER, J. W. & HALL, H. K. 1990. Adaptive acidification tolerance response of Salmonella typhimurium. *J Bacteriol,* 172**,** 771-8.
GAJIWALA, K. S. & BURLEY, S. K. 2000. HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *Journal of molecular biology,* 295**,** 605-12.
GAMA-CASTRO, S., SALGADO, H., PERALTA-GIL, M., SANTOS-ZAVALETA, A., MUNIZ-RASCADO, L., SOLANO-LIRA, H., JIMENEZ-JACINTO, V., WEISS, V., GARCIA-SOTELO, J. S., LOPEZ-FUENTES, A., PORRON-SOTELO, L., ALQUICIRA-HERNANDEZ, S., MEDINA-RIVERA, A., MARTINEZ-FLORES, I., ALQUICIRA-HERNANDEZ, K., MARTINEZ-ADAME, R., BONAVIDES-MARTINEZ, C., MIRANDA-RIOS, J., HUERTA, A. M., MENDOZA-VARGAS, A., COLLADO-TORRES, L., TABOADA, B., VEGA-ALVARADO, L., OLVERA, M., OLVERA, L., GRANDE, R., MORETT, E. & COLLADO-VIDES, J. 2011. RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic acids research,* 39**,** D98-105.
GAO, R. & STOCK, A. M. 2009. Biological insights from structures of two-component proteins. *Annual review of microbiology,* 63**,** 133-54.
GARCIA, D. E., BAIDOO, E. E., BENKE, P. I., PINGITORE, F., TANG, Y. J., VILLA, S. & KEASLING, J. D. 2008. Separation and mass spectrometry in microbial metabolomics. *Current opinion in microbiology,* 11**,** 233-9.
GEORGE, S. M., RICHARDSON, L. C., POL, I. E. & PECK, M. W. 1998. Effect of oxygen concentration and redox potential on recovery of sublethally heat-damaged cells of Escherichia coli O157:H7, Salmonella enteritidis and Listeria monocytogenes. *J Appl Microbiol,* 84**,** 903-9.
GEORGELLIS, D., KWON, O., LIN, E. C., WONG, S. M. & AKERLEY, B. J. 2001. Redox signal transduction by the ArcB sensor kinase of Haemophilus influenzae lacking the PAS domain. *Journal of bacteriology,* 183**,** 7206-12.
GIANNELLA, R. A., BROITMAN, S. A. & ZAMCHECK, N. 1972. Gastric acid barrier to ingested microorganisms in man: studies in vivo and in vitro. *Gut,* 13**,** 251-6.
GIANNELLA, R. A., BROITMAN, S. A. & ZAMCHECK, N. 1973. Influence of gastric acidity on bacterial and parasitic enteric infections. A perspective. *Annals of internal medicine,* 78**,** 271-6.
GIBNEY, M. J., WALSH, M., BRENNAN, L., ROCHE, H. M., GERMAN, B. & VAN OMMEN, B. 2005. Metabolomics in human nutrition: opportunities and challenges. *The American journal of clinical nutrition,* 82**,** 497-503.
GILBERT, J. A., FIELD, D., HUANG, Y., EDWARDS, R., LI, W., GILNA, P. & JOINT, I. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PloS one,* 3**,** e3042.
GOODIER, R. I. & AHMER, B. M. 2001. SirA orthologs affect both motility and virulence. *J Bacteriol,* 183**,** 2249-58.
GOULIAN, M. 2010. Two-component signaling circuit structure and properties. *Current opinion in microbiology,* 13**,** 184-9.
GRAHAM, K. & OLSON, M. F. 2007. The ras signalling pathway as a target in cancer therapy. *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer,* 172**,** 125-53.
GROBAN, E. S., CLARKE, E. J., SALIS, H. M., MILLER, S. M. & VOIGT, C. A. 2009. Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of molecular biology,* 390**,** 380-93.
GUIGLIARELLI, B., MAGALON, A., ASSO, M., BERTRAND, P., FRIXON, C., GIORDANO, G. & BLASCO, F. 1996. Complete coordination of the four Fe-S centers of the beta subunit from Escherichia coli nitrate reductase. Physiological, biochemical, and EPR

characterization of site-directed mutants lacking the highest or lowest potential [4Fe-4S] clusters. *Biochemistry,* 35**,** 4828-36.
GUNN, J. S. & MILLER, S. I. 1996. PhoP-PhoQ activates transcription of pmrAB, encoding a two-component regulatory system involved in Salmonella typhimurium antimicrobial peptide resistance. *Journal of bacteriology,* 178**,** 6857-64.
GUPTA, R., STINCONE, A., ANTCZAK, P., DURANT, S., BICKNELL, R., BIKFALVI, A. & FALCIANI, F. 2011. A computational framework for gene regulatory network inference that combines multiple methods and datasets. *BMC systems biology,* 5**,** 52.
GUTKNECHT, J. & WALTER, A. 1981. Transport of protons and hydrochloric acid through lipid bilayer membranes. *Biochimica et biophysica acta,* 641**,** 183-8.
HALL, H. K., KAREM, K. L. & FOSTER, J. W. 1995. Molecular responses of microbes to environmental pH stress. *Advances in microbial physiology,* 37**,** 229-72.
HALL, M. N. & SILHAVY, T. J. 1981. Genetic analysis of the ompB locus in Escherichia coli K-12. *J Mol Biol,* 151**,** 1-15.
HALL, N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *The Journal of experimental biology,* 210**,** 1518-25.
HAYASHI, K., MOROOKA, N., YAMAMOTO, Y., FUJITA, K., ISONO, K., CHOI, S., OHTSUBO, E., BABA, T., WANNER, B. L., MORI, H. & HORIUCHI, T. 2006. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Mol Syst Biol,* 2**,** 2006 0007.
HAYES, E. T., WILKS, J. C., SANFILIPPO, P., YOHANNES, E., TATE, D. P., JONES, B. D., RADMACHER, M. D., BONDURANT, S. S. & SLONCZEWSKI, J. L. 2006. Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in Escherichia coli K-12. *BMC Microbiol,* 6**,** 89.
HECKER, M., LAMBECK, S., TOEPFER, S., VAN SOMEREN, E. & GUTHKE, R. 2009. Gene regulatory network inference: data integration in dynamic models-a review. *Bio Systems,* 96**,** 86-103.
HENGGE, R. 2009. Proteolysis of sigmaS (RpoS) and the general stress response in Escherichia coli. *Res Microbiol,* 160**,** 667-76.
HERBERT, D., ELSWORTH, R. & TELLING, R. C. 1956. The continuous culture of bacteria; a theoretical and experimental study. *J Gen Microbiol,* 14**,** 601-22.
HEUVELING, J., POSSLING, A. & HENGGE, R. 2008. A role for Lon protease in the control of the acid resistance genes of Escherichia coli. *Mol Microbiol,* 69**,** 534-47.
HEWITT, C. J., NEBE-VON CARON, G., NIENOW, A. W. & MCFARLANE, C. M. 1999. The use of multi-parameter flow cytometry to compare the physiological response of Escherichia coli W3110 to glucose limitation during batch, fed-batch and continuous culture cultivations. *J Biotechnol,* 75**,** 251-64.
HICKS, D. B. & KRULWICH, T. A. 1986. The membrane ATPase of alkalophilic Bacillus firmus RAB is an F1-type ATPase. *The Journal of biological chemistry,* 261**,** 12896-902.
HILL, C., O'DRISCOLL, B. & BOOTH, I. 1995. Acid adaptation and food poisoning microorganisms. *Int J Food Microbiol,* 28**,** 245-54.
HIROSE, O., YOSHIDA, R., IMOTO, S., YAMAGUCHI, R., HIGUCHI, T., CHARNOCK-JONES, D. S., PRINT, C. & MIYANO, S. 2008. Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics,* 24**,** 932-42.
HOMMAIS, F., KRIN, E., COPPEE, J. Y., LACROIX, C., YERAMIAN, E., DANCHIN, A. & BERTIN, P. 2004. GadE (YhiE): a novel activator involved in the response to acid environment in Escherichia coli. *Microbiology,* 150**,** 61-72.
HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.
HUNTER, S., JONES, P., MITCHELL, A., APWEILER, R., ATTWOOD, T. K., BATEMAN, A., BERNARD, T., BINNS, D., BORK, P., BURGE, S., DE CASTRO, E., COGGILL, P., CORBETT, M., DAS, U., DAUGHERTY, L., DUQUENNE, L., FINN, R. D., FRASER,

M., GOUGH, J., HAFT, D., HULO, N., KAHN, D., KELLY, E., LETUNIC, I., LONSDALE, D., LOPEZ, R., MADERA, M., MASLEN, J., MCANULLA, C., MCDOWALL, J., MCMENAMIN, C., MI, H., MUTOWO-MUELLENET, P., MULDER, N., NATALE, D., ORENGO, C., PESSEAT, S., PUNTA, M., QUINN, A. F., RIVOIRE, C., SANGRADOR-VEGAS, A., SELENGUT, J. D., SIGRIST, C. J., SCHEREMETJEW, M., TATE, J., THIMMAJANARTHANAN, M., THOMAS, P. D., WU, C. H., YEATS, C. & YONG, S. Y. 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic acids research,* 40**,** D306-12.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. & SPEED, T. P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-64.

ISHII, N., NAKAHIGASHI, K., BABA, T., ROBERT, M., SOGA, T., KANAI, A., HIRASAWA, T., NABA, M., HIRAI, K., HOQUE, A., HO, P. Y., KAKAZU, Y., SUGAWARA, K., IGARASHI, S., HARADA, S., MASUDA, T., SUGIYAMA, N., TOGASHI, T., HASEGAWA, M., TAKAI, Y., YUGI, K., ARAKAWA, K., IWATA, N., TOYA, Y., NAKAYAMA, Y., NISHIOKA, T., SHIMIZU, K., MORI, H. & TOMITA, M. 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science,* 316**,** 593-7.

IUCHI, S. & LIN, E. C. 1988. arcA (dye), a global regulatory gene in Escherichia coli mediating repression of enzymes in aerobic pathways. *Proc Natl Acad Sci U S A,* 85**,** 1888-92.

IYER, R., WILLIAMS, C. & MILLER, C. 2003. Arginine-agmatine antiporter in extreme acid resistance in Escherichia coli. *J Bacteriol,* 185**,** 6556-61.

JAWOROWSKI, A., MAYO, G., SHAW, D. C., CAMPBELL, H. D. & YOUNG, I. G. 1981. Characterization of the respiratory NADH dehydrogenase of Escherichia coli and reconstitution of NADH oxidase in ndh mutant membrane vesicles. *Biochemistry,* 20**,** 3621-8.

JOHNSON, J. M., WEAGANT, S. D., JINNEMAN, K. C. & BRYANT, J. L. 1995. Use of pulsed-field gel electrophoresis for epidemiological study of Escherichia coli O157:H7 during a food-borne outbreak. *Applied and environmental microbiology,* 61**,** 2806-8.

JOHNSON, M. D., BURTON, N. A., GUTIERREZ, B., PAINTER, K. & LUND, P. A. 2011. RcsB is required for inducible acid resistance in Escherichia coli and acts at gadE-dependent and -independent promoters. *Journal of bacteriology,* 193**,** 3653-6.

JONES, R. W., LAMONT, A. & GARLAND, P. B. 1980. The mechanism of proton translocation driven by the respiratory nitrate reductase complex of Escherichia coli. *Biochemical Journal,* 190**,** 79.

JOYCE, A. R. & PALSSON, B. O. 2006. The model organism as a system: integrating 'omics' data sets. *Nature reviews. Molecular cell biology,* 7**,** 198-210.

JOZEFCZUK, S., KLIE, S., CATCHPOLE, G., SZYMANSKI, J., CUADROS-INOSTROZA, A., STEINHAUSER, D., SELBIG, J. & WILLMITZER, L. Metabolomic and transcriptomic stress response of Escherichia coli. *Mol Syst Biol,* 6**,** 364.

KAILASAN VANAJA, S. 2009. *Interplay between acid resistance and virulence in Escherichia coli O157| H7.* MICHIGAN STATE UNIVERSITY.

KANEHISA, M., ARAKI, M., GOTO, S., HATTORI, M., HIRAKAWA, M., ITOH, M., KATAYAMA, T., KAWASHIMA, S., OKUDA, S., TOKIMATSU, T. & YAMANISHI, Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic acids research,* 36**,** D480-4.

KANNAN, G., WILKS, J. C., FITZGERALD, D. M., JONES, B. D., BONDURANT, S. S. & SLONCZEWSKI, J. L. 2008. Rapid acid treatment of Escherichia coli: transcriptomic response and recovery. *BMC Microbiol,* 8**,** 37.

KAPER, J. B., NATARO, J. P. & MOBLEY, H. L. 2004. Pathogenic Escherichia coli. *Nature reviews. Microbiology,* 2**,** 123-40.

KARLEBACH, G. & SHAMIR, R. 2008. Modelling and analysis of gene regulatory networks. *Nature reviews. Molecular cell biology,* 9**,** 770-80.

KASHKET, E. R. 1985. Effects of K+ and Na+ on the proton motive force of respiring Escherichia coli at alkaline pH. *Journal of bacteriology,* 163**,** 423-9.
KELL, D. B. 2004. Metabolomics and systems biology: making sense of the soup. *Current opinion in microbiology,* 7**,** 296-307.
KELLIS, M. & RINN, J. L. 2010. Sequences to systems. *Genome biology,* 11**,** 303.
KESELER, I. M., BONAVIDES-MARTINEZ, C., COLLADO-VIDES, J., GAMA-CASTRO, S., GUNSALUS, R. P., JOHNSON, D. A., KRUMMENACKER, M., NOLAN, L. M., PALEY, S., PAULSEN, I. T., PERALTA-GIL, M., SANTOS-ZAVALETA, A., SHEARER, A. G. & KARP, P. D. 2009. EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic acids research,* 37**,** D464-70.
KHODURSKY, A. B., PETER, B. J., COZZARELLI, N. R., BOTSTEIN, D., BROWN, P. O. & YANOFSKY, C. 2000. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 12170-5.
KIM, J. R. & CHO, K. H. 2006. The multi-step phosphorelay mechanism of unorthodox two-component systems in E. coli realizes ultrasensitivity to stimuli while maintaining robustness to noises. *Comput Biol Chem,* 30**,** 438-44.
KING, T., LUCCHINI, S., HINTON, J. C. & GOBIUS, K. 2010. Transcriptomic analysis of Escherichia coli O157:H7 and K-12 cultures exposed to inorganic and organic acids in stationary phase reveals acidulant- and strain-specific acid tolerance responses. *Applied and environmental microbiology,* 76**,** 6514-28.
KIRBY, J. R. 2009. Chemotaxis-like regulatory systems: unique roles in diverse bacteria. *Annual review of microbiology,* 63**,** 45-59.
KITANO, H. 2002. Systems biology: a brief overview. *Science,* 295**,** 1662-4.
KITKO, R. D., WILKS, J. C., GARDUQUE, G. M. & SLONCZEWSKI, J. L. Osmolytes contribute to pH homeostasis of Escherichia coli. *PLoS One,* 5**,** e10078.
KOBAYASHI, H. & ANRAKU, Y. 1972. Membrane-bound adenosine triphosphatase of Escherichia coli. I. Partial purification and properties. *Journal of biochemistry,* 71**,** 387-99.
KOHL, P., CRAMPIN, E. J., QUINN, T. A. & NOBLE, D. 2010. Systems biology: an approach. *Clinical pharmacology and therapeutics,* 88**,** 25-33.
KRULWICH, T. A., SACHS, G. & PADAN, E. 2011. Molecular aspects of bacterial pH sensing and homeostasis. *Nature reviews. Microbiology,* 9**,** 330-43.
LACERDA, C. M. & REARDON, K. F. 2009. Environmental proteomics: applications of proteome profiling in environmental microbiology and biotechnology. *Briefings in functional genomics & proteomics,* 8**,** 75-87.
LAMARK, T., KAASEN, I., ESHOO, M. W., FALKENBERG, P., MCDOUGALL, J. & STROM, A. R. 1991. DNA sequence and analysis of the bet genes encoding the osmoregulatory choline-glycine betaine pathway of Escherichia coli. *Mol Microbiol,* 5**,** 1049-64.
LAUB, M. T. & GOULIAN, M. 2007. Specificity in two-component signal transduction pathways. *Annual review of genetics,* 41**,** 121-45.
LECLERC, R. D. 2008. Survival of the sparsest: robust gene networks are parsimonious. *Molecular systems biology,* 4.
LEYER, G. J. & JOHNSON, E. A. 1993. Acid adaptation induces cross-protection against environmental stresses in Salmonella typhimurium. *Applied and environmental microbiology,* 59**,** 1842-7.
LEYER, G. J., WANG, L. L. & JOHNSON, E. A. 1995. Acid adaptation of Escherichia coli O157: H7 increases survival in acidic foods. *Applied and Environmental Microbiology,* 61**,** 3752-3755.
LIBEREK, K., LEWANDOWSKA, A. & ZIETKIEWICZ, S. 2008. Chaperones in control of protein disaggregation. *EMBO J,* 27**,** 328-35.

LIN, J., LEE, I. S., FREY, J., SLONCZEWSKI, J. L. & FOSTER, J. W. 1995. Comparative analysis of extreme acid survival in Salmonella typhimurium, Shigella flexneri, and Escherichia coli. *Journal of bacteriology,* 177**,** 4097-104.

LIN, J., SMITH, M. P., CHAPIN, K. C., BAIK, H. S., BENNETT, G. N. & FOSTER, J. W. 1996. Mechanisms of acid resistance in enterohemorrhagic Escherichia coli. *Appl Environ Microbiol,* 62**,** 3094-100.

LUDWIG, C., EASTON, J., LODI, A., TIZIANI, S., MANZOOR, S., SOUTHAM, A., BYRNE, J., BISHOP, L., HE, S., ARVANITIS, T., GÜNTHER, U. & VIANT, M. 2012. Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). *Metabolomics,* 8**,** 8-18.

LUKAT, G. S., MCCLEARY, W. R., STOCK, A. M. & STOCK, J. B. 1992. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 718-22.

MA, Z., GONG, S., RICHARD, H., TUCKER, D. L., CONWAY, T. & FOSTER, J. W. 2003. GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in Escherichia coli K-12. *Molecular microbiology,* 49**,** 1309-20.

MA, Z., MASUDA, N. & FOSTER, J. W. 2004. Characterization of EvgAS-YdeO-GadE branched regulatory circuit governing glutamate-dependent acid resistance in Escherichia coli. *J Bacteriol,* 186**,** 7378-89.

MA, Z., RICHARD, H., TUCKER, D. L., CONWAY, T. & FOSTER, J. W. 2002. Collaborative regulation of Escherichia coli glutamate-dependent acid resistance by two AraC-like regulators, GadX and GadW (YhiW). *J Bacteriol,* 184**,** 7001-12.

MAIER, T., GUELL, M. & SERRANO, L. 2009. Correlation of mRNA and protein in complex biological samples. *FEBS Lett,* 583**,** 3966-73.

MANDELSTAM, J. 1958. Turnover of protein in growing and non-growing populations of Escherichia coli. *Biochem J,* 69**,** 110-9.

MARBACH, D., SCHAFFTER, T., MATTIUSSI, C. & FLOREANO, D. 2009. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *Journal of computational biology : a journal of computational molecular cell biology,* 16**,** 229-39.

MARGOLIN, A. A., NEMENMAN, I., BASSO, K., WIGGINS, C., STOLOVITZKY, G., DALLA FAVERA, R. & CALIFANO, A. 2006a. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics,* 7 Suppl 1**,** S7.

MARGOLIN, A. A., WANG, K., LIM, W. K., KUSTAGI, M., NEMENMAN, I. & CALIFANO, A. 2006b. Reverse engineering cellular networks. *Nat Protoc,* 1**,** 662-71.

MARTEYN, B., WEST, N. P., BROWNING, D. F., COLE, J. A., SHAW, J. G., PALM, F., MOUNIER, J., PREVOST, M. C., SANSONETTI, P. & TANG, C. M. 2010. Modulation of Shigella virulence in response to available oxygen in vivo. *Nature,* 465**,** 355-8.

MARTIN-GALIANO, A. J., FERRANDIZ, M. J. & DE LA CAMPA, A. G. 2001. The promoter of the operon encoding the F0F1 ATPase of Streptococcus pneumoniae is inducible by pH. *Mol Microbiol,* 41**,** 1327-38.

MARTINEZ-ANTONIO, A. & COLLADO-VIDES, J. 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol,* 6**,** 482-9.

MARTINSEN, T. C., BERGH, K. & WALDUM, H. L. 2005. Gastric juice: a barrier against infectious diseases. *Basic & clinical pharmacology & toxicology,* 96**,** 94-102.

MASUDA, N. & CHURCH, G. M. 2002. Escherichia coli gene expression responsive to levels of the response regulator EvgA. *J Bacteriol,* 184**,** 6225-34.

MATES, A. K., SAYED, A. K. & FOSTER, J. W. 2007. Products of the Escherichia coli acid fitness island attenuate metabolite stress at extremely low pH and mediate a cell density-dependent acid resistance. *J Bacteriol,* 189**,** 2759-68.

MATSA, D. A. 2011. Competition and product quality in the supermarket industry. *Q J Econ,* 126**,** 1539-91.
MATSUSHITA, K., OHNISHI, T. & KABACK, H. R. 1987. NADH-ubiquinone oxidoreductases of the Escherichia coli aerobic respiratory chain. *Biochemistry,* 26**,** 7732-7.
MATTHIES, D., HABERSTOCK, S., JOOS, F., DOTSCH, V., VONCK, J., BERNHARD, F. & MEIER, T. 2011. Cell-free expression and assembly of ATP synthase. *J Mol Biol,* 413**,** 593-603.
MAURER, L. M., YOHANNES, E., BONDURANT, S. S., RADMACHER, M. & SLONCZEWSKI, J. L. 2005. pH regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia coli K-12. *J Bacteriol,* 187**,** 304-19.
MCMEEKIN, T. A., HILL, C., WAGNER, M., DAHL, A. & ROSS, T. 2010. Ecophysiology of food-borne pathogens: Essential knowledge to improve food safety. *Int J Food Microbiol,* 139 Suppl 1**,** S64-78.
MELLMANN, A., HARMSEN, D., CUMMINGS, C. A., ZENTZ, E. B., LEOPOLD, S. R., RICO, A., PRIOR, K., SZCZEPANOWSKI, R., JI, Y., ZHANG, W., MCLAUGHLIN, S. F., HENKHAUS, J. K., LEOPOLD, B., BIELASZEWSKA, M., PRAGER, R., BRZOSKA, P. M., MOORE, R. L., GUENTHER, S., ROTHBERG, J. M. & KARCH, H. 2011. Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One,* 6**,** e22751.
MELTON-CELSA, A. R., ROGERS, J. E., SCHMITT, C. K., DARNELL, S. C. & O'BRIEN, A. D. 1998. Virulence of Shiga toxin-producing Escherichia coli (STEC) in orally-infected mice correlates with the type of toxin produced by the infecting strain. *Japanese journal of medical science & biology,* 51 Suppl**,** S108-14.
MIKA, F. & HENGGE, R. 2005. A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of sigmaS (RpoS) in E. coli. *Genes Dev,* 19**,** 2770-81.
MILNER, P. F. 1963. The differentiation of Enterobacteriaceae infecting the urinary tract: A study in male paraplegics. *Journal of clinical pathology,* 16**,** 39-45.
MITCHELL, P. & MOYLE, J. 1954. The Gram reaction and cell composition: nucleic acids and other phosphate fractions. *Journal of general microbiology,* 10**,** 533-40.
MIZUNO, T. & TANAKA, I. 1997. Structure of the DNA-binding domain of the OmpR family of response regulators. *Molecular microbiology,* 24**,** 665-7.
MONTET, M., CHRISTIEANS, S., THEVENOT, D., COPPET, V., GANET, S., MULLER, M., DUNIÈRE, L., MISZCZYCHA, S. & VERNOZY-ROZAND, C. 2009. Fate of acid-resistant and non-acid resistant Shiga toxin-producing< i> Escherichia coli</i> strains in experimentally contaminated French fermented raw meat sausages. *International journal of food microbiology,* 129**,** 264-270.
MOXON, E. R., RAINEY, P. B., NOWAK, M. A. & LENSKI, R. E. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol,* 4**,** 24-33.
MUNCH, R., HILLER, K., GROTE, A., SCHEER, M., KLEIN, J., SCHOBERT, M. & JAHN, D. 2005. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics,* 21**,** 4187-9.
NATARO, J. P. & KAPER, J. B. 1998. Diarrheagenic Escherichia coli. *Clinical microbiology reviews,* 11**,** 142-201.
NEELY, M. N., DELL, C. L. & OLSON, E. R. 1994. Roles of LysP and CadC in mediating the lysine requirement for acid induction of the Escherichia coli cad operon. *Journal of bacteriology,* 176**,** 3278-85.
NI BHRIAIN, N., DORMAN, C. J. & HIGGINS, C. F. 1989. An overlap between osmotic and anaerobic stress responses: a potential role for DNA supercoiling in the coordinate regulation of gene expression. *Mol Microbiol,* 3**,** 933-42.
NIE, L., WU, G., BROCKMAN, F. J. & ZHANG, W. 2006. Integrated analysis of transcriptomic and proteomic data of Desulfovibrio vulgaris: zero-inflated Poisson regression models to predict abundance of undetected proteins. *Bioinformatics,* 22**,** 1641-7.

NIE, L., WU, G. & ZHANG, W. 2008. Statistical application and challenges in global gel-free proteomic analysis by mass spectrometry. *Critical reviews in biotechnology,* 28**,** 297-307.
NIKAIDO, H. & VAARA, M. 1985. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev,* 49**,** 1-32.
NOBLE, D. 2002. The rise of computational biology. *Nature reviews. Molecular cell biology,* 3**,** 459-63.
NOBLE, D. 2008. Claude Bernard, the first systems biologist, and the future of physiology. *Experimental physiology,* 93**,** 16-26.
NUNEZ, C., ESTEVE-NUNEZ, A., GIOMETTI, C., TOLLAKSEN, S., KHARE, T., LIN, W., LOVLEY, D. R. & METHE, B. A. 2006. DNA microarray and proteomic analyses of the RpoS regulon in Geobacter sulfurreducens. *Journal of bacteriology,* 188**,** 2792-800.
OGATA, H., GOTO, S., FUJIBUCHI, W. & KANEHISA, M. 1998. Computation with the KEGG pathway database. *Biosystems,* 47**,** 119-28.
OLIVIER, A., LEE, H. Y. & COTE, J. C. 2005. Study of the heterogeneity of 16S rRNA genes in gamma-proteobacteria: implications for phylogenetic analysis. *The Journal of general and applied microbiology,* 51**,** 395-405.
ONG, I. M., GLASNER, J. D. & PAGE, D. 2002. Modelling regulatory pathways in E. coli from time series expression profiles. *Bioinformatics,* 18 Suppl 1**,** S241-8.
OPDYKE, J. A., KANG, J. G. & STORZ, G. 2004. GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J Bacteriol,* 186**,** 6698-705.
ORTEGA, F., SAMEITH, K., TURAN, N., COMPTON, R., TREVINO, V., VANNUCCI, M. & FALCIANI, F. 2008. Models and computational strategies linking physiological response to molecular networks from large-scale data. *Philos Transact A Math Phys Eng Sci,* 366**,** 3067-89.
OSHIMA, T., AIBA, H., MASUDA, Y., KANAYA, S., SUGIURA, M., WANNER, B. L., MORI, H. & MIZUNO, T. 2002. Transcriptome analysis of all two-component regulatory system mutants of Escherichia coli K-12. *Molecular microbiology,* 46**,** 281-91.
OZSOLAK, F., PLATT, A. R., JONES, D. R., REIFENBERGER, J. G., SASS, L. E., MCINERNEY, P., THOMPSON, J. F., BOWERS, J., JAROSZ, M. & MILOS, P. M. 2009. Direct RNA sequencing. *Nature,* 461**,** 814-8.
PALANIYANDI, S., MITRA, A., HERREN, C. D., LOCKATELL, C. V., JOHNSON, D. E., ZHU, X. & MUKHOPADHYAY, S. 2012. BarA-UvrY two-component system regulates virulence of uropathogenic E. coli CFT073. *PLoS One,* 7**,** e31348.
PENG, S., TASARA, T., HUMMERJOHANN, J. & STEPHAN, R. 2011. An Overview of Molecular Stress Response Mechanisms in Escherichia coli Contributing to Survival of Shiga ToxinProducing Escherichia coli during Raw Milk Cheese Production. *Journal of Food Protection®,* 74**,** 849-864.
PERNESTIG, A. K., GEORGELLIS, D., ROMEO, T., SUZUKI, K., TOMENIUS, H., NORMARK, S. & MELEFORS, O. 2003. The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources. *J Bacteriol,* 185**,** 843-53.
PERRIN, B. E., RALAIVOLA, L., MAZURIE, A., BOTTANI, S., MALLET, J. & D'ALCHE-BUC, F. 2003. Gene networks inference using dynamic Bayesian networks. *Bioinformatics,* 19 Suppl 2**,** ii138-48.
PIERARD, D., VAN DAMME, L., MORIAU, L., STEVENS, D. & LAUWERS, S. 1997. Virulence factors of verocytotoxin-producing Escherichia coli isolated from raw meats. *Applied and environmental microbiology,* 63**,** 4585-7.
RAAMSDONK, L. M., TEUSINK, B., BROADHURST, D., ZHANG, N., HAYES, A., WALSH, M. C., BERDEN, J. A., BRINDLE, K. M., KELL, D. B., ROWLAND, J. J., WESTERHOFF, H. V., VAN DAM, K. & OLIVER, S. G. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature biotechnology,* 19**,** 45-50.

RANGEL, C., ANGUS, J., GHAHRAMANI, Z., LIOUMI, M., SOTHERAN, E., GAIBA, A., WILD, D. L. & FALCIANI, F. 2004. Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics,* 20**,** 1361-72.

RAYCHAUDHURI, S., STUART, J. M. & ALTMAN, R. B. 2000. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput***,** 455-66.

REID, S. D., HERBELIN, C. J., BUMBAUGH, A. C., SELANDER, R. K. & WHITTAM, T. S. 2000. Parallel evolution of virulence in pathogenic Escherichia coli. *Nature,* 406**,** 64-7.

RICHARD, H. & FOSTER, J. W. 2004. Escherichia coli glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J Bacteriol,* 186**,** 6032-41.

RICHARD, H. T. & FOSTER, J. W. 2003. Acid resistance in Escherichia coli. *Adv Appl Microbiol,* 52**,** 167-86.

RILEY, L. W., REMIS, R. S., HELGERSON, S. D., MCGEE, H. B., WELLS, J. G., DAVIS, B. R., HEBERT, R. J., OLCOTT, E. S., JOHNSON, L. M., HARGRETT, N. T., BLAKE, P. A. & COHEN, M. L. 1983. Hemorrhagic colitis associated with a rare Escherichia coli serotype. *N Engl J Med,* 308**,** 681-5.

ROBERTSON, D. G. 2005. Metabonomics in toxicology: a review. *Toxicological sciences : an official journal of the Society of Toxicology,* 85**,** 809-22.

ROTTENBERG, H. 1973. The thermodynamic description of enzyme-catalyzed reactions. The linear relation between the reaction rate and the affinity. *Biophys J,* 13**,** 503-11.

ROTTENBERG, H. & KOEPPE, R. E., 2ND 1989. Mechanism of uncoupling of oxidative phosphorylation by gramicidin. *Biochemistry,* 28**,** 4355-60.

RUZANTE, J. M., DAVIDSON, V. J., CASWELL, J., FAZIL, A., CRANFIELD, J. A., HENSON, S. J., ANDERS, S. M., SCHMIDT, C. & FARBER, J. M. 2010. A multifactorial risk prioritization framework for foodborne pathogens. *Risk Anal,* 30**,** 724-42.

SAEED, A. I., BHAGABATI, N. K., BRAISTED, J. C., LIANG, W., SHAROV, V., HOWE, E. A., LI, J., THIAGARAJAN, M., WHITE, J. A. & QUACKENBUSH, J. 2006. TM4 microarray software suite. *Methods Enzymol,* 411**,** 134-93.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374-8.

SAMELIS, J., IKEDA, J. S. & SOFOS, J. N. 2003. Evaluation of the pH-dependent, stationary-phase acid tolerance in Listeria monocytogenes and Salmonella Typhimurium DT104 induced by culturing in media with 1% glucose: a comparative study with Escherichia coli O157:H7. *Journal of applied microbiology,* 95**,** 563-75.

SANDERSON, K. E. 1976. Genetic relatedness in the family Enterobacteriaceae. *Annual review of microbiology,* 30**,** 327-49.

SARGENT, F., BALLANTINE, S. P., RUGMAN, P. A., PALMER, T. & BOXER, D. H. 1998. Reassignment of the gene encoding the Escherichia coli hydrogenase 2 small subunit--identification of a soluble precursor of the small subunit in a hypB mutant. *Eur J Biochem,* 255**,** 746-54.

SAYED, A. K., ODOM, C. & FOSTER, J. W. 2007. The Escherichia coli AraC-family regulators GadX and GadW activate gadE, the central activator of glutamate-dependent acid resistance. *Microbiology,* 153**,** 2584-92.

SCHWAN, W. R. 2009. Survival of uropathogenic Escherichia coli in the murine urinary tract is dependent on OmpR. *Microbiology,* 155**,** 1832-9.

SENIOR, A. E. 1990. The proton-translocating ATPase of Escherichia coli. *Annu Rev Biophys Biophys Chem,* 19**,** 7-41.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res,* 13**,** 2498-504.
SHAPIRO, H. M. & NEBE-VON-CARON, G. 2004. Multiparameter flow cytometry of bacteria. *Methods Mol Biol,* 263**,** 33-44.
SHI, J. & BIEK, D. P. 1995. A versatile low-copy-number cloning vector derived from plasmid F. *Gene,* 164**,** 55-8.
SHI, X. & BENNETT, G. N. 1994. Effects of rpoA and cysB mutations on acid induction of biodegradative arginine decarboxylase in Escherichia coli. *Journal of bacteriology,* 176**,** 7017-23.
SIEGELE, D. A. & KOLTER, R. 1992. Life after log. *J Bacteriol,* 174**,** 345-8.
SIRYAPORN, A. & GOULIAN, M. 2008. Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in E. coli. *Molecular microbiology,* 70**,** 494-506.
SLONCZEWSKI, J. L., GONZALEZ, T. N., BARTHOLOMEW, F. M. & HOLT, N. J. 1987. Mu d-directed lacZ fusions regulated by low pH in Escherichia coli. *J Bacteriol,* 169**,** 3001-6.
SMALL, P., BLANKENHORN, D., WELTY, D., ZINSER, E. & SLONCZEWSKI, J. L. 1994. Acid and base resistance in Escherichia coli and Shigella flexneri: role of rpoS and growth pH. *J Bacteriol,* 176**,** 1729-37.
SMITH, D. K., KASSAM, T., SINGH, B. & ELLIOTT, J. F. 1992. Escherichia coli has two homologous glutamate decarboxylase genes that map to distinct loci. *J Bacteriol,* 174**,** 5820-6.
SMITH, J. L. 2003. The role of gastric acid in preventing foodborne disease and how bacteria overcome acid conditions. *Journal of food protection,* 66**,** 1292-303.
SOFOS, J. N. 2008. Challenges to meat safety in the 21st century. *Meat Sci,* 78**,** 3-13.
SOGA, T., BARAN, R., SUEMATSU, M., UENO, Y., IKEDA, S., SAKURAKAWA, T., KAKAZU, Y., ISHIKAWA, T., ROBERT, M., NISHIOKA, T. & TOMITA, M. 2006. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *The Journal of biological chemistry,* 281**,** 16768-76.
SONCINI, F. C. & GROISMAN, E. A. 1996. Two-component regulatory systems can interact to process multiple environmental signals. *Journal of bacteriology,* 178**,** 6796-801.
STANCIK, L. M., STANCIK, D. M., SCHMIDT, B., BARNHART, D. M., YONCHEVA, Y. N. & SLONCZEWSKI, J. L. 2002. pH-dependent expression of periplasmic proteins and amino acid catabolism in Escherichia coli. *J Bacteriol,* 184**,** 4246-58.
STEDTFELD, R. D., BAUSHKE, S. W., TOURLOUSSE, D. M., MILLER, S. M., STEDTFELD, T. M., GULARI, E., TIEDJE, J. M. & HASHSHAM, S. A. 2008. Development and experimental validation of a predictive threshold cycle equation for quantification of virulence and marker genes by high-throughput nanoliter-volume PCR on the OpenArray platform. *Applied and environmental microbiology,* 74**,** 3831-8.
STEINFATH, M., REPSILBER, D., SCHOLZ, M., WALTHER, D. & SELBIG, J. 2007. Integrated data analysis for genome-wide research. *EXS,* 97**,** 309-29.
STINCONE, A., DAUDI, N., RAHMAN, A. S., ANTCZAK, P., HENDERSON, I., COLE, J., JOHNSON, M. D., LUND, P. & FALCIANI, F. 2011. A systems biology approach sheds new light on Escherichia coli acid resistance. *Nucleic acids research,* 39**,** 7512-28.
STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000. Two-component signal transduction. *Annual review of biochemistry,* 69**,** 183-215.
SUKHAREV, S. I., MARTINAC, B., ARSHAVSKY, V. Y. & KUNG, C. 1993. Two types of mechanosensitive channels in the Escherichia coli cell envelope: solubilization and functional reconstitution. *Biophys J,* 65**,** 177-83.
SUN, L., FUKAMACHI, T., SAITO, H. & KOBAYASHI, H. 2005. Carbon dioxide increases acid resistance in Escherichia coli. *Lett Appl Microbiol,* 40**,** 397-400.
SUN, Y., FUKAMACHI, T., SAITO, H. & KOBAYASHI, H. 2011. ATP requirement for acidic resistance in Escherichia coli. *Journal of bacteriology,* 193**,** 3072-7.

SWERDLOW, D. L., WOODRUFF, B. A., BRADY, R. C., GRIFFIN, P. M., TIPPEN, S., DONNELL, H. D., JR., GELDREICH, E., PAYNE, B. J., MEYER, A., JR., WELLS, J. G. & ET AL. 1992. A waterborne outbreak in Missouri of Escherichia coli O157:H7 associated with bloody diarrhea and death. *Annals of internal medicine,* 117**,** 812-9.

TANABE, M., NISHIO, K., IKO, Y., SAMBONGI, Y., IWAMOTO-KIHARA, A., WADA, Y. & FUTAI, M. 2001. Rotation of a complex of the gamma subunit and c ring of Escherichia coli ATP synthase. The rotor and stator are interchangeable. *The Journal of biological chemistry,* 276**,** 15269-74.

TAYLOR, M. R. 2011. Will the Food Safety Modernization Act help prevent outbreaks of foodborne illness? *N Engl J Med,* 365**,** e18.

TEPLITSKI, M., GOODIER, R. I. & AHMER, B. M. 2003. Pathways leading from BarA/SirA to motility and virulence gene expression in Salmonella. *J Bacteriol,* 185**,** 7257-65.

TETSCH, L., KOLLER, C., DONHOFER, A. & JUNG, K. 2011. Detection and function of an intramolecular disulfide bond in the pH-responsive CadC of Escherichia coli. *BMC microbiology,* 11**,** 74.

TOMOYASU, T., MOGK, A., LANGEN, H., GOLOUBINOFF, P. & BUKAU, B. 2001. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. *Mol Microbiol,* 40**,** 397-413.

TOYA, Y., ISHII, N., HIRASAWA, T., NABA, M., HIRAI, K., SUGAWARA, K., IGARASHI, S., SHIMIZU, K., TOMITA, M. & SOGA, T. 2007. Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *Journal of chromatography. A,* 1159**,** 134-41.

TRAMONTI, A., DE CANIO, M. & DE BIASE, D. 2008. GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: transcriptional control at the gadY-gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol Microbiol,* 70**,** 965-82.

TRAMONTI, A., DE CANIO, M., DELANY, I., SCARLATO, V. & DE BIASE, D. 2006. Mechanisms of transcription activation exerted by GadX and GadW at the gadA and gadBC gene promoters of the glutamate-based acid resistance system in Escherichia coli. *J Bacteriol,* 188**,** 8118-27.

TRAMONTI, A., VISCA, P., DE CANIO, M., FALCONI, M. & DE BIASE, D. 2002. Functional characterization and regulation of gadX, a gene encoding an AraC/XylS-like transcriptional activator of the Escherichia coli glutamic acid decarboxylase system. *J Bacteriol,* 184**,** 2603-13.

TUCKER, D. L., TUCKER, N., MA, Z., FOSTER, J. W., MIRANDA, R. L., COHEN, P. S. & CONWAY, T. 2003. Genes of the GadX-GadW regulon in Escherichia coli. *Journal of bacteriology,* 185**,** 3190-201.

TUSHER, V. G., TIBSHIRANI, R. & CHU, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* 98**,** 5116-21.

TWEEDDALE, H., NOTLEY-MCROBB, L. & FERENCI, T. 1998. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. *Journal of bacteriology,* 180**,** 5109-16.

TZSCHOPPE, M., MARTIN, A. & BEUTIN, L. 2012. A rapid procedure for the detection and isolation of enterohaemorrhagic Escherichia coli (EHEC) serogroup O26, O103, O111, O118, O121, O145 and O157 strains and the aggregative EHEC O104:H4 strain from ready-to-eat vegetables. *International journal of food microbiology,* 152**,** 19-30.

ULRICH, L. E., KOONIN, E. V. & ZHULIN, I. B. 2005. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol,* 13**,** 52-6.

URBANCZYK-WOCHNIAK, E., LUEDEMANN, A., KOPKA, J., SELBIG, J., ROESSNER-TUNALI, U., WILLMITZER, L. & FERNIE, A. R. 2003. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO reports,* 4**,** 989-93.

UYAR, E., KUROKAWA, K., YOSHIMURA, M., ISHIKAWA, S., OGASAWARA, N. & OSHIMA, T. 2009. Differential binding profiles of StpA in wild-type and h-ns mutant

cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis. *Journal of bacteriology,* 191**,** 2388-91.
VELLIOU, E. G., VAN DERLINDEN, E., CAPPUYNS, A. M., NIKOLAIDOU, E., GEERAERD, A. H., DEVLIEGHERE, F. & VAN IMPE, J. F. 2011. Towards the quantification of the effect of acid treatment on the heat tolerance of Escherichia coli K12 at lethal temperatures. *Food Microbiol,* 28**,** 702-11.
VIDAL, O., LONGIN, R., PRIGENT-COMBARET, C., DOREL, C., HOOREMAN, M. & LEJEUNE, P. 1998. Isolation of an Escherichia coli K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression. *Journal of bacteriology,* 180**,** 2442-9.
VOGT, R. L. & DIPPOLD, L. 2005. Escherichia coli O157:H7 outbreak associated with consumption of ground beef, June-July 2002. *Public health reports,* 120**,** 174-8.
WANG, Q. Z., WU, C. Y., CHEN, T., CHEN, X. & ZHAO, X. M. 2006. Integrating metabolomics into a systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms. *Appl Microbiol Biotechnol,* 70**,** 151-61.
WANNER, B. L. 1992. Is cross regulation by phosphorylation of two-component response regulator proteins important in bacteria? *Journal of bacteriology,* 174**,** 2053-8.
WATERS, T. M., WARNECKE, R. B., PARSONS, J., ALMAGOR, O. & BUDETTI, P. P. 2006. The role of the national practitioner data bank in the credentialing process. *American journal of medical quality : the official journal of the American College of Medical Quality,* 21**,** 30-9.
WATSON, B. S., LEI, Z., DIXON, R. A. & SUMNER, L. W. 2004. Proteomics of Medicago sativa cell walls. *Phytochemistry,* 65**,** 1709-20.
WEBER, A., KOGL, S. A. & JUNG, K. 2006. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in Escherichia coli. *J Bacteriol,* 188**,** 7165-75.
WEBER, H., POLEN, T., HEUVELING, J., WENDISCH, V. F. & HENGGE, R. 2005. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *Journal of bacteriology,* 187**,** 1591-603.
WERBER, D., KRAUSE, G., FRANK, C., FRUTH, A., FLIEGER, A., MIELKE, M., SCHAADE, L. & STARK, K. 2012. Outbreaks of virulent diarrhoeagenic Escherichia coli - are we in control? *BMC medicine,* 10**,** 11.
WEST, A. H. & STOCK, A. M. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in biochemical sciences,* 26**,** 369-76.
WILLIAMS, K. P., GILLESPIE, J. J., SOBRAL, B. W., NORDBERG, E. K., SNYDER, E. E., SHALLOM, J. M. & DICKERMAN, A. W. 2010. Phylogeny of gammaproteobacteria. *Journal of bacteriology,* 192**,** 2305-14.
WOESE, C. R. 1987. Bacterial evolution. *Microbiological reviews,* 51**,** 221-71.
WOESE, C. R., STACKEBRANDT, E., MACKE, T. J. & FOX, G. E. 1985. A phylogenetic definition of the major eubacterial taxa. *Systematic and applied microbiology,* 6**,** 143-51.
WU, X., LI, P., WANG, N., GONG, P., PERKINS, E. J., DENG, Y. & ZHANG, C. 2011. State Space Model with hidden variables for reconstruction of gene regulatory networks. *BMC Syst Biol,* 5 Suppl 3**,** S3.
WUICHET, K., CANTWELL, B. J. & ZHULIN, I. B. 2010. Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol,* 13**,** 219-25.
YAMAMOTO, Y., MIWA, Y., MIYOSHI, K., FURUYAMA, J. & OHMORI, H. 1997. The Escherichia coli ldcC gene encodes another lysine decarboxylase, probably a constitutive enzyme. *Genes & genetic systems,* 72**,** 167-72.
YANG, Y. & INOUYE, M. 1993. Requirement of both kinase and phosphatase activities of an Escherichia coli receptor (Taz1) for ligand-dependent signal transduction. *Journal of molecular biology,* 231**,** 335-42.
YE, Y., ZHANG, L., HAO, F., ZHANG, J., WANG, Y. & TANG, H. 2012. Global Metabolomic Responses of Escherichia coli to Heat Stress. *Journal of proteome research*.

YOON, K. Y., WOODAMS, E. E. & HANG, Y. D. 2006. Production of probiotic cabbage juice by lactic acid bacteria. *Bioresour Technol,* 97**,** 1427-30.

YOON, S. H., HAN, M. J., LEE, S. Y., JEONG, K. J. & YOO, J. S. 2003. Combined transcriptome and proteome analysis of Escherichia coli during high cell density culture. *Biotechnology and bioengineering,* 81**,** 753-67.

YOUSEF, A. E. & COURTNEY, P. D. 2003. Basics of stress adaptation and implications in new-generation foods. *Microbial stress adaptation and food safety,* 1**,** 1-30.

YU, J., SMITH, V. A., WANG, P. P., HARTEMINK, A. J. & JARVIS, E. D. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics,* 20**,** 3594-603.

ZASLAVER, A., MAYO, A. E., ROSENBERG, R., BASHKIN, P., SBERRO, H., TSALYUK, M., SURETTE, M. G. & ALON, U. 2004. Just-in-time transcription program in metabolic pathways. *Nature genetics,* 36**,** 486-91.

ZAVARZIN, G. A., STACKEBRANDT, E. & MURRAY, R. G. 1991. A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. *Canadian journal of microbiology,* 37**,** 1-6.

ZHANG, F. & KEASLING, J. 2011. Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol,* 19**,** 323-9.

ZHANG, W., LI, F. & NIE, L. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology,* 156**,** 287-301.

ZHOU, L., LEI, X. H., BOCHNER, B. R. & WANNER, B. L. 2003. Phenotype microarray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems. *Journal of bacteriology,* 185**,** 4956-72.

ZWIR, I., SHIN, D., KATO, A., NISHINO, K., LATIFI, T., SOLOMON, F., HARE, J. M., HUANG, H. & GROISMAN, E. A. 2005. Dissecting the PhoP regulatory network of Escherichia coli and Salmonella enterica. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 2862-7.