Transcriptome analysis of the lactic acid and NaCl-stress response of *Lactobacillus plantarum*

Bart Pieterse

Promotor	**Prof. Dr. W. M. de Vos** Hoogleraar Microbiologie Wageningen Universiteit
Co-promotor	**Dr. Ir. M. J. van der Werf** Productmanager microbiële productieprocessen TNO Kwaliteit van Leven, Zeist
Promotiecommissie	**Prof. Dr. T. Abee** Wageningen Universiteit
	Dr. M. Kleerebezem NIZO Food Research, Ede
	Dr. R. van Kranenburg Purac, Gorinchem
	Prof. Dr. O. P. Kuipers Rijksuniversiteit Groningen

Dit onderzoek is uitgevoerd binnen de onderzoekschool VLAG

Transcriptome analysis of the lactic acid and NaCl-stress response of *Lactobacillus plantarum*

Bart Pieterse

Proefschrift

Ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. Dr. M. J. Kropff
in het openbaar te verdedigen op
maandag 24 april 2006
des namiddags te 16.00 in de Aula

B. Pieterse- Transcriptome analysis of the lactic acid and NaCl-stress response of *Lactobacillus plantarum*-2006

Thesis Wageningen University, Wageningen, The Netherlands – with summary in Dutch – 160 p.

Keywords: transcriptome, microarray, experimental design, *Lactobacillus plantarum*, lactic acid, NaCl, stress

ISBN 90-8504-404-9

Voor mijn ouders

Abstract

Many cellular processes are regulated at the transcriptional level. For this reason the transcriptome of a cell, the total set of RNAs under a specific condition, contains information about the biological state of the cell and the genes that play a role under specific circumstances. Transcriptomics is the research method that studies the effect of specific conditions on alterations in the expression levels of complete sets of genes. This thesis describes the application and optimization of transcriptomics and transcriptomics-related techniques in studies on lactic acid and sodium chloride stress-related gene expression in *Lactobacillus plantarum* WCFS1. The complete genomic sequence of this strain is available, and the organism is used as a model strain for lactic acid bacteria in industrial applications and host-microbe interactions.

In order to maximize specificity and reliability of transcriptome data, a quenching protocol for the isolation of microbial cells from submerged cultures was implemented and validated. The results from this study revealed a significant positive effect of the quenching protocol on the conservation of the transcriptome, and consequently on the reliability and representativity of the gene expression data.

Transcriptome analyses were performed with clone-based microarrays. In order to optimise the potential of clone-based arrays for transcriptomics studies of strains from which the complete genomic sequence is not available, two complementary equations were developed that allow for mathematical design of such an array. The first equation predicts the fraction of genes that is represented on the array in a detectable way and cover at least a set part of the genomic fragment on the array. The second equation predicts the fraction of genes for which information can be interpreted in a quantitative way.

The effect of lactic acid on gene expression in *L. plantarum* was studied by means of an experimental design that allows for the distinction between the long term effects of lactate, lactic acid, pH, water activity and growth rate. A group of genes was identified that specifically responded to the undissociated form of the lactic acid molecule. This group contained, among others, several genes that have been associated with other stress responses, as well as a highly coherent group of cell surface protein-encoding genes. From these genes, two putative regulators, a peroxide-resistance gene, and one of the cell surface protein encoding clusters were selected for validation of their role under lactic acid-stress. Experiments with mutant strains in which these genes were inactivated or overexpressed gave no specific indication on their relevance under the tested conditions.

In a study on the long term effects of NaCl stress on *L. plantarum* no increased expression was observed for genes that are typically related to acute osmotic stress. The

results indicate a different nature of NaCl-imposed stress on the cells on the long term for which adaptations over multiple functional classes are required. A remarkable observation was the detection of a high number of differentially expressed genes under continous NaCl stress with the consensus sequence for the catabolite responsive element *cre* in their upstream region.

The results presented contribute to the reliability and applicability of transcriptomics research in microorganisms. Moreover, the results from the studies on the response of *L. plantarum* towards lactic acid- and NaCl-stress provided new and global insights in the nature of these stresses and may serve a role in strain- and fermentation-optimizations.

Contents

Abstract		
Chapter 1	Introduction	11
Chapter 2	Quenching of microbial samples for increased reliability of microarray data	31
Chapter 3	Mathematical design of prokaryotic clone-based microarrays	50
Chapter 4	Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling	72
Chapter 5	Analysis of lactic acid tolerance in *Lactobacillus plantarum* derivatives mutated in lactic acid-stimulon associated functions	98
Chapter 6	Transcriptome analysis of the effect of continuous NaCl stress on gene expression of *Lactobacillus plantarum*	114
Chapter 7	Summary and concluding remarks	139
Samenvatting		151
Nawoord		153
List of publications		157
Curriculum vitae		159

Chapter 1

Introduction

Lactobacillus plantarum

Lactobacillus plantarum is one of the predominant species of the *Lactobacillus* genus, both with respect to its distribution in a large number of habitats and its applications (73). This lactic acid bacterium plays an important role in plant fermentations, such as the fermentations of sauerkraut, cucumber, cabbage, olives and silage (73). Moreover, *L. plantarum* has been used as a protective organism to suppress growth of spoilage organisms in meat products (62) and has been isolated from various types of cheese (73). Also in the wine industry *L. plantarum* has shown great potential to positively contribute to the quality of the end product due to its malolactic fermentation properties (53, 73). Associated with its presence in fermented foods, is the intake of *L. plantarum* by human consumption. It has been shown that this *Lactobacillus* species are able to persist the conditions encountered in the gastro-intestinal tract of both mice (48) and humans (72). Moreover, probiotic properties have been ascribed to various *L. plantarum* strains (18, 45).

Recently, the complete genome of *L. plantarum* WCFS1 has been sequenced and annotated (37). The 3.3 Mbp genome contains over 3000 putative protein encoding genes. This strain currently functions as a model organism for lactic acid bacteria in industrial applications and host-microbe interactions.

Stress related to (food) processing

Lactic acid bacteria are usually a relatively low cost ingredient in their various applications. Nevertheless, their growth performance and robustness are essential properties for the final product. Hence, they should be able to successfully counteract various adverse conditions during processing, such as acid, osmotic, temperature and oxidative stress as reviewed by Van de Guchte *et al.*(69). Whereas selection strategies can be useful for the identification of new strains with desired properties, physiological and expression studies can serve a role in obtaining new information on the effect of certain stress factors on the cell. This insight can be applied for optimizing of processing as well as the selection of new targets for strain improvement. In this thesis, the effects of lactic acid stress and NaCl-imposed osmotic stress on gene expression in *L. plantarum* were studied.

Lactic acid stress

Lactic acid is the predominant metabolite formed during fermentations with lactobacilli. Although lactobacilli are generally more resistant towards lactic acid than many other microorganisms, they are still strongly inhibited by the concentrations of this organic acid that are reached during fermentations in both bioreactors and plant fermentations (22, 56). The inhibitory effect of lactic acid and other organic acids is mainly attributed to the undissociated form of the molecule. This uncharged form easily diffuses over the cell membrane towards the more alkaline cytosol (5, 29). Inside the cell, where the pH is higher, the lactic acid dissociates in correspondence with the Henderson-Hasselbach equation, resulting in a dissipation of the ΔpH. This process continuous until an equilibrium has been reached at which the intra- and extracellular concentrations of the undissociated, membrane-permeable form of lactic acid are equal. It has been shown that *L. plantarum* is also highly susceptible to the undissociated form of the lactic acid (43). We confirmed this finding for *L. plantarum* WCFS1 by determining the effect of pH and the initial total lactate concentration

Total lactate (mM) \ pH	6	5.5	5	4.5	4	3.5
0					0	0
50					21	35
75						52
100					42	70
125					52	
150					63	
175					73	
200				37	84	139
250				46		
300			20	56	126	
350				65		
400			27	74	167	
500		11	34	93		
600	4.3	13	40			
700	5.0	16	47			
800	5.7	18	54			
900	6.4	20	60			
1000	7.1	22	67			

Figure 1. Effect of lactate concentration and pH on the ability of *L. plantarum* WCFS1 to grow. Growth experiments were performed in duplicate at 37°C on 0.25x MRS broth with additional sodium lactate at different pH values. For all combinations tested, the initial undissociated lactic acid concentration in the growth medium is depicted in the table. Growth was defined as an increase in OD_{600} of more than 3 times the mean plus the average deviation of the non-inoculated control samples after a period of 3 days. Growth was observed in all combinations of pH and total lactate for which the cells in the table are gray. The light gray filling (900mM total lactate, pH 5) indicates that no growth was observed in the duplicate experiment.

on the ability to grow (Fig. 1). The undissociated lactic acid concentration was found to have a larger effect on the ability of this strain to grow than the total lactate concentration or the pH. At lactic acid concentrations above 50-80 mM in the medium, growth ceased. Moreover, it was shown that the addition of sodium lactate to the medium initiates an acid adaptation response in *L. plantarum* WCFS1: cells that were pregrown in the presence of lactate at pH

5.5 showed a higher survival rate at low pH than cells that were pregrown at pH 7.0 or at pH 5.5 in the absence of sodium lactate (Fig. 2). This indicates that the adaptation response can not be explained just by the lower pH effect of the organic acid on the extracellular pH (Fig. 2).

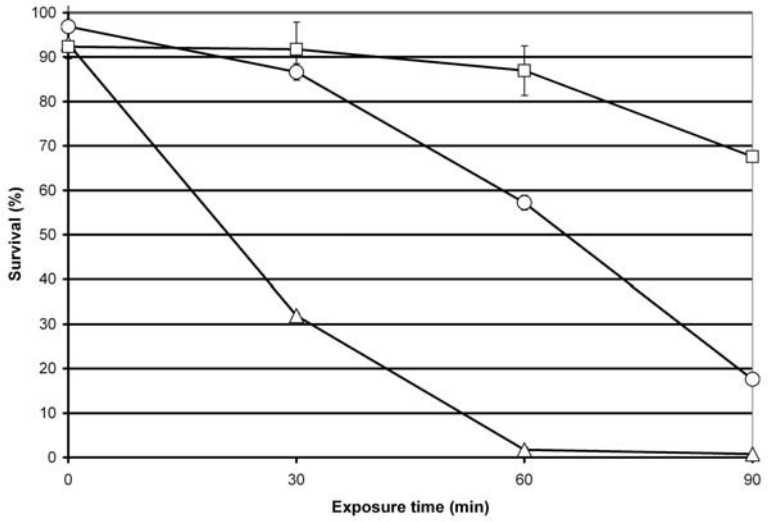

Figure 2. Survival of *L. plantarum* WCFS1 in 0.25 x MRS pH 2.0 after a 1 hour preincubation of logarithmically growing cells in 0.25 x MRS (37ºC) at pH 7.0 (Δ), pH 5.5 (O), or pH 5.5 with 300mM sodium lactate (□), respectively. Survival was determined by viable cell counts and related to the viable cell counts prior to the exposure to 0.25 x MRS pH 2.0.

Although the inhibitory effect of lactic acid is related to the undissociated form of the molecule, the mechanism underlying this toxicity is still not clear. The uncoupling theory considers organic acids as uncouplers that are responsible for a ΔpH-driven influx of protons, resulting in a complete dissipation of the proton motive force (5, 30). Another theory attributes the inhibitory effect of organic acids by the decreasing effect on the intracellular pH (60). This theory is in disagreement with the observation that most fermenting bacteria seem to lower their intracellular pH in order to avoid a large ΔpH, and as a result increase their tolerance against organic acids (10, 14, 19). Adaptation to low extracellular pH values by a decrease of the intracellular pH has also been observed for *L. plantarum,* which can grow at intracellular pH values as low as 4.6-4.8 (43). A third theory attributes the inhibitory effect of organic acids to the intracellular accumulation of anions, which could lead to both end-product inhibition and a loss of water-activity (55, 56).

Table 1. Systems that play a role in alkalisation of the cytosol and / or in the generation of a proton motive force at low pH in lactobacilli.

System	Mechanism	Organism	(Homologous) system in L. plantarum WCFS1	References
F_0F_1-ATPase	Proton expulsion at the expense of ATP.	*L. helveticus*	lp_2363-lp_2370	79
ADI-pathway	Conversion of arginine into ornithine, ammonia and carbon dioxide. The decarboxylylation was associated with the net formation of ATP. The ammonia formation results in alkalisation of the cytosol.	Various lactobacilli		3
Gad	Decarboxylation of glutamate to gamma-aminobutyrate resulting in alkalisation of the cytosol. The decarboxylylation was associated with the net formation of ATP.	*Lactobacillus sp.* *L. brevis*		28, 66
Malolactic fermentation	Conversion of the dicarboxylic malic acid to the monocarboxylic lactic acid. In *L. plantarum* the produced lactate is excreted in electrogenic uniport with malate, yielding one ATP.	*L. plantarum* *L. sake*	lp_1115_lp_1119	38, 47
Asp/Ala exchange	Amino acid decarboxylse / antiporter system: an amino acid is imported and decarboxylated. This results in the consumption of a proton.	*Lactobacillus sp*.		1
Amino acid antiporter	Implicated role in an amino acid decarboxylse/antiporter system.	*L. acidophilus*		6
Ornithine decarboxylase/ amino acid permease	Suggested decarboxylse / antiporter system resulting in the net consumption of a proton.	*L. acidophilus*		6
Histidine/Histamine exchange	Decarboxylation of histidine to histamine. The histamine is transported from the cell in electrogenic in electrogenic uniport with histidine resulting in a proton motive force.	*L. buchneri*		44

Several systems have been identified in lactobacilli that serve a role in alkalization of the cytosol and / or generation of a proton motive force (Table 1): the F_0F_1-ATPase system is capable of generating ATP by the uptake of a proton, but can also expulse protons at the expense of 1 ATP (79); In the arginine deiminase system (ADI), arginine is converted into ornithine, ammonia and carbon dioxide with a net energy gain of 1 ATP. The ornithine is transported from the cell in antiport with arginine. The ammonia production results in alkalization of the cytosol (17); Several systems have been reported that import a carboxylic acidic compound (e.g. amino acids) into the cell (39). After decarboxylation, the product is transported from the cell by an electrogenic transporter, which results in the generation of a proton motive force. In the malolactic fermentation, malate is oxidatively decarboxylated into pyruvate. The generated pyruvate is subsequently reduced to lactate. In *L. plantarum*, the lactate is transported from the cell in electrogenic uniport with malate (47). To our knowledge, this is the only system in *L. plantarum* for which the relevance under acidic conditions has been tested and confirmed.

The majority of studies on the effect of undissociated lactic acid on prokaryotes concentrated on physiological parameters, like growth rate and intracellular pH (e.g. 30, 43, 57). The impact of organic acids on gene expression in prokaryotes has been the subject of several studies (4, 36, 50). However, none of these studies allowed for the distinction between the effects of the dissociated and the undissociated form.

Hyperosmotic stress

Hyperosmosis is a frequently encountered stress factor for prokaryotes during processing steps in the food industry. In many cases, suppression of bacterial growth by high osmolarity is a desired effect and forms a frequently applied strategy to avoid pathogenic growth (2). However, in cases where the living microorganism is needed during the process or in the end-product, growth inhibition or death by hyperosmosis may be an unwanted side effect.

The initial effect of hyperosmotic stress on biological cells is caused by an outward flux of water due to the higher extracellular osmolyte concentrations. This results in a decreased turgor pressure, which is essential for cell elongation (16). Obviously, cells that are capable of growing under hyperosmotic conditions, have overcome the initial problem of a decreased turgor. Nevertheless, they may still suffer from high osmolarity: since the restoration of turgor requires intracellular accumulation of osmotically active compounds the water activity in the cells will decrease. Consequently, the cells may suffer from ongoing stress that is caused by the altered composition of the cytoplasm (11).

The majority of studies on bacterial osmotic-stress response focus on the adaptation towards the initial effect by the accumulation of osmolytes (for reviews see: 15, 16, 46, 51, 54, 61, 76). An important group of osmolytes that serve this role are the so-called compatible solutes: osmoprotective compounds that are highly soluble and compatible with cellular functions, even at high concentrations (16). The uptake of compatible solutes by *L. plantarum*, and their effect on growth, has been the subject of several studies (23-26, 33-35). In contrast to enteric bacteria and *Bacillus subtilis*, *L. plantarum* has little or no possibilities to synthesize these compatible solutes *de novo*, and depends on their uptake from the environment (51). On the genome of *L. plantarum* WCFS1 at least three predicted transcription units are present with genes for compatible solute transporters (lp_0367-lp_0369, lp_1607-lp_1610, and lp_3324) (37).

Transcriptomics

Many cellular processes are regulated at the transcriptional level. For this reason the transcriptome of a cell, which is the total set of RNA under a specific condition, contains information on the biological state of the cell and the genes that play a role under specific circumstances.

Transcriptomics is the research approach that studies the effect of specific conditions on alterations in the expression levels of large numbers of genes. To achieve this, the transcripts present in cells obtained from different samples are quantified by microarray analysis, and the resulting datasets are compared by means of various mathematical approaches (Fig. 3). Since the expression levels are determined for "all" the genes that are represented on the microarray, this approach is highly unbiased. This is an important advantage over most classical expression studies in which the expression levels of specific genes thought to be of relevance are determined in order to confirm a hypothesis that is based on prior knowledge. Below, an overview is presented of the different steps that are of importance in a transcriptomics experiment.

Microarray analysis

Microarray analysis was first described in scientific literature in 1995 (58), and allows for the simultaneous comparison of the expression levels of large sets of genes in response to the conditions of interest. For a microarray analysis, RNA is isolated from cells grown or incubated under conditions of interest. This RNA is subsequently converted into labeled cDNA or amplified RNA. Most labeling protocols, including the one that was used in

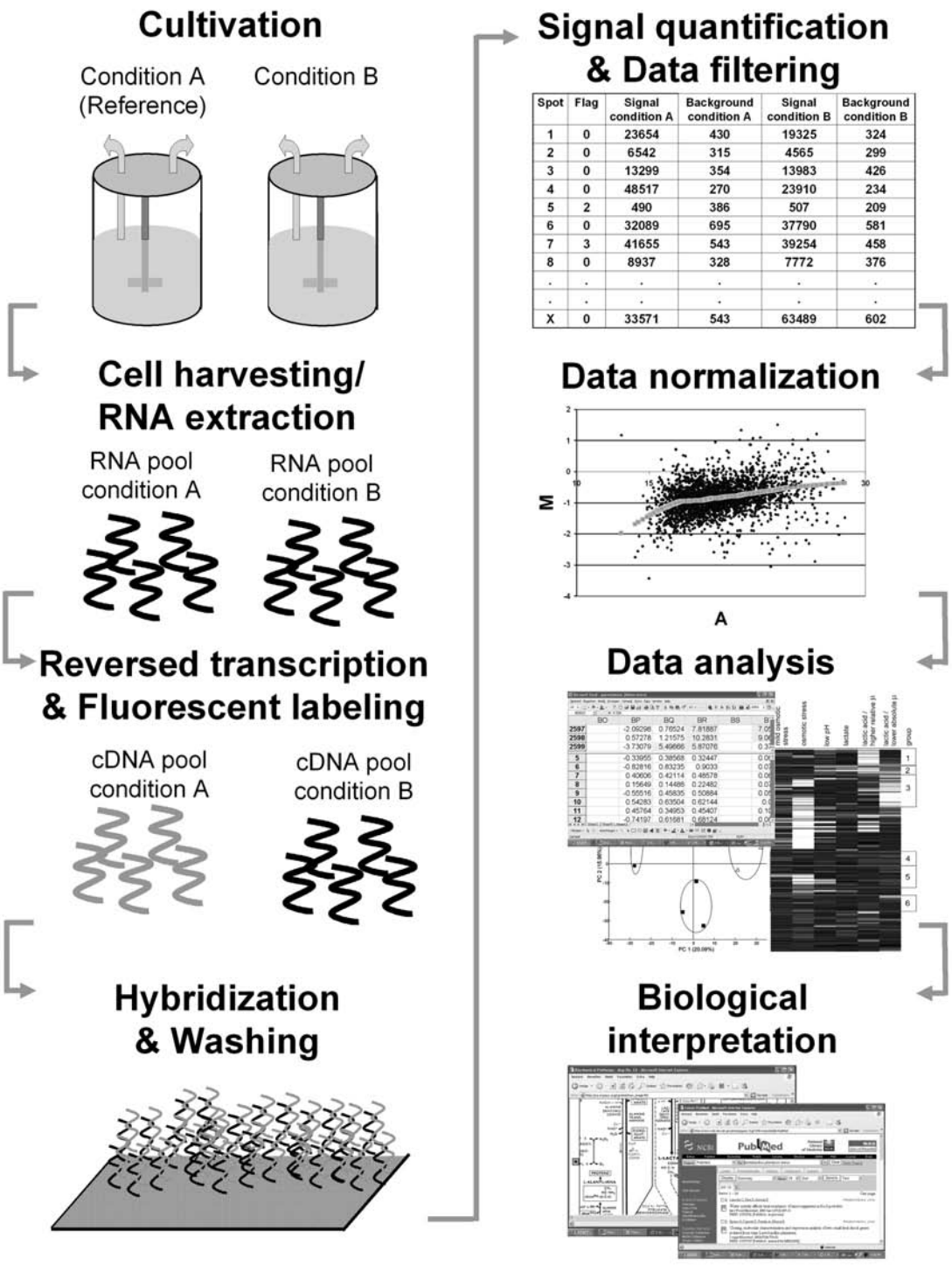

Figure 3. General outline of a transcriptomics experiment.

this thesis, are based on the linear, enzymatic amplification approach first described by Van Gelder *et al.* (71). The RNA samples are transcribed into fluorescent cDNA by reversed transcription. During or after the reversed transcription, a label is bound to the cDNA. The

labeled cDNA's are hybridized to the microarray, which is a small glass slide containing thousands of spots. Usually each spot represents one gene of the strain of interest and consists of a gene or gene fragment that has been spotted or printed on the slide. Whereas for some microarray approaches labeled cDNA from a single sample is uniquely hybridized onto the microarray, other approaches rely on the co-hybridization of two samples onto the same slide. In the latter case, the samples can be distinguished by using discrete labels (usually fluorescent Cy3 and Cy5 labels). Since the labeled cDNA's will hybridize to the equivalent spots on the microarray, the total amount of signal from each spot is a measure for the level at which the corresponding genes were expressed. For the analyses that were performed for this thesis, the co-hybridization approach was applied.

A drawback of the standard microarray, on which each spot represents a single gene of the organism, is that the design requires a completely annotated genome sequence. An alternative strategy is the construction of a clone-based microarray on which each spot represents a genome fragment from the organism of interest. For this approach no prior knowledge on the genome sequence is required, and the fragments that are to be spotted on the array slide can be amplified from the genome library with generic primers. Drawbacks of this approach are that (i) it is beforehand uncertain whether a specific gene is present on the array, (ii) the identity(ies) of the gene(s) present in the spotted fragments are not known beforehand, and (iii) one spotted fragment on the array may represent multiple genes.

Experimental design

Although transcriptomics approaches can be highly efficient with respect to the time and money that are spend in relation to the amount of data that is generated, the investment that could be wasted due to improper experimental design may even be more impressive. All steps from biological culturing to data interpretation should be comprised and integrated in the experimental design. Parameters that are directly related to statistical analysis of microarray data, such as the number of biological replicates, number of arrays, and array hybridization schemes have been discussed on multiple occasions (e.g. 12, 31, 59, 80). Biological considerations, on the other hand, should also be accounted for. The goal of many transcriptomics studies is to identify those genes that specifically respond to a condition of interest and serve a role in the physiological adaptation to that condition. Unfortunately, other genes may also be differentially expressed that may not serve a specific function under that condition and from which the differential expression is the consequence of side effects like e.g. a diminished growth rate or general stress. In the classical, hypothesis-driven, expression studies the impact of such effects on the interpretability and specificity of the data has been

limited since the expression levels were mostly determined to confirm a hypothesis that was already founded on a biological rationale. However, in case of an unbiased transcriptomics approach, the presence of such genes in the final selection will complicate data interpretation and target selection. For this reason, straightforward pairwise comparisons will often not result in the desired specificity of the data. The biological specificity of a transcriptomics experiment could be improved by including samples from cultures that mimic the sub-effects of the condition of interest. Another issue that relates to the reliability and specificity of transcriptomics data is the short term on which changes can be introduced to the transcriptome. Literature on microbial RNA half lives and the rapid transcriptional response towards changes in the environment indicate that potential changes introduced to the transcriptome during the harvesting of microbial cells include both degradation and *de novo* transcription (8, 49, 74, 75, 78). This implies that unwanted alterations could be introduced to the transcriptome during harvesting and work-up of the biological sample. Therefore, experimental strategies should be applied that limit these changes.

Signal quantification and data filtering

The fluorescent signals and background signals from the different labels in the different spots on the array are quantified by confocal laser scanning. After the quantification of the signal the data are filtered. During this phase the quality of the array is judged on criteria like the signal-to-background ratio in each spot, homogeneity of the signal throughout the spot, whether the signal is within the detection limits of the confocal laser scanner, and whether the signals are homogenously distributed over the entire array. If the quality does not fulfill the required standards, data from the unreliable spot, or in extreme cases from the complete microarray, may de removed from the dataset. Another option is to mark data points as less reliable.

Data normalization

The next step in the data processing is normalization of the data, a step that is often required due to a deviation between the observed signals and the actual amount of hybridized cDNA. These deviations may be introduced by different labeling efficiencies, different hybridization efficiencies, variation in the total amount of probed material, and a non-linear relation between the measured signal and the actual amount of cDNA (9). Several types of strategies for the normalization of microarray data can be distinguished (9, 41): (i) using of housekeeping genes; (ii) inclusion of a known amount of exogenous control genes as internal controls (52); (iii) using genomic DNA as a reference sample; (iv) mathematical selection of non-differentially expressed genes to normalize to; (v) using all genes on the array for the

normalization. This latter approach is based on the assumptions that the expression of most genes is not affected, and that the total amount of RNA is equal under both conditions. Adjustments have been made to the various normalization methods that allow for the anticipation to specific sources of error like e.g. print-tip variations (63).

Currently, the most commonly used normalization approach, is the LOWESS normalization (13). LOWESS is a locally weighed scattered plot smoother that was implemented for the determination of signal dependent normalization factors. This approach appears superior to other normalization methods (41), and is relatively easy to implement.

Data analysis

After filtering and normalization, the data should represent the biological conditions of interest, and are suitable for data analysis to pinpoint similarities and differences in the transcriptomes compared. Two major types of data analysis can be distinguished: significance analysis and multivariate data analysis. Whereas significance analysis tools typically consider the behavior of a single variable over multiple conditions, multivariate tools (also: pattern recognition tools) consider the interrelationships between multiple variables (67). Significance analysis and multivariate data analysis methods often fulfill complementary roles in biological research: while multivariate tools reveal patterns in large datasets, significance analysis indicates the reliability of an observation for a specific variable. For instance, in transcriptomics, multivariate tools can be applied for target selection by identifying genes that specifically correlate with the biological process of interest. Subsequently, the statistical reliability of the observed expression changes for those genes can be determined by significance analysis.

Examples of significance analysis tools are ANOVA (32), Bayesian method (7, 42), and the Mann-Whitney test (77). Approaches that determine a general cut-off value for the entire dataset should be avoided, since they are based on the unjust assumption that variation between the replicates is spot independent. Therefore, these approaches involve both the risk of a high number of false positives and false negatives. One has to realize that a statistical test that determines a general cut-off value for a dataset of 5000 genes with a significance level $\alpha=0.05$, could easily result in result in 5000 x 0.05 = 250 false positives (41).

Multivariate data analysis tools are generally useful to reveal complex relations in large datasets and subsequently visualize these relations in a comprehensible format. Representative examples of multivariate data analysis tools that are specifically useful for the analysis of transcriptomics data are principal component analysis, and various types of clustering methods. Principal component analysis allows for the simultaneous visualization of thousands of variables in only a few dimensions. This is achieved by means of a mathematical

reduction of the many variables into a few principal components that account for as much of the variability between the datasets as possible. Usually, for a transcriptomics experiment each spot in the resulting score plot represents the data from one microarray. Datasets that are closest related, group together in the score plot. Clustering methods determine similarity measures for the behavior of different genes over the different conditions, and use these measures for the grouping of these genes. Various examples of clustering methods are hierarchical clustering (20), self organizing maps (64), K-means clustering (65), and quality clustering (27).

Since this part of the data analysis is a purely statistical process, no prior knowledge on gene expression in response to the tested conditions is required and the amount of bias is minimized.

Biological interpretation

After data analysis, the large dataset containing expression information about all genes present on the microarray is reduced to a limited set of genes from which the expression is affected by the condition(s) of interest. These genes form the basis for the biological interpretation of the microarray results. Numerous useful computational tools for the clarification of functional genomics data are available as commercial packages or are freely available on the internet (21). These comprise tools that allow for e.g. the data visualization by projection onto a metabolic pathway map, tools for the retrieval of structural and functional information for specific genes, and text- and literature-mining tools (68).

In some situations the function of a differentially expressed gene can positively be confirmed as being important for the situation of interest due to available knowledge on that gene and its regulation. However, in many cases the function of a gene under a specific situation only becomes clear in the context of its co-expression with other genes. Genes could, for instance, have joined functions in the same biological pathway or be classified to the same biological function. A representative example of a powerful tool for the query, visualization and linking of genomics data is the MetaCyc database, which is a metabolic database of pathways and enzymes from a wide variety of organisms (40). Recently, a *L. plantarum* WCFS1 specific version, LacPlantCyc, has become available (70).

Outline of the thesis

This thesis describes the study of the effects of two stress factors that are relevant in industrial applications in *L. plantarum* WCFS1 by transcriptome analysis: lactic acid and osmotic stress.

A major part of the research was conducted by means of a transcriptomics approach. Since this relatively young research approach is still evolving, special attention was paid to the optimization of transcriptomics-related procedures. In Chapter 2 a protocol was developed for the harvesting of prokaryotic cells under quenched conditions in order to avoid the introduction of changes in the transcriptome during harvesting.

The microarray that was used in this research was a clone-based array on which amplified fragments from a *L. plantarum* WCFS1 clone library (constructed for the sequencing of the genome) were spotted. In order to optimize this approach for unsequenced strains as well, formulas were developed that allow for optimal mathematical design of prokaryotic clone-based microarrays by the prediction of the fraction of represented genes for which interpretable results can be generated (Chapter 3).

Lactic acid stress depends on the pH of the environment and the concentration of the organic acid. By comparing the transcriptomes from multiple growth conditions, a distinction could be made between the pH effect, and the effects of the undissociated and dissociated form of the molecule and growth-related effects (Chapter 4). Genes that appeared to be upregulated in response to the undissociated form of the lactic acid molecule were selected for overexpression and / or knock out studies in order to determine their effect on growth and survival at high concentrations of undissociated lactic acid (Chapter 5).

Osmotic stress is associated with multiple processing steps that microorganisms encountered during food processing. The effect of continuous exposure to two different sodium chloride concentrations on gene expression in *L. plantarum* WCFS1 was studied (Chapter 6).

Finally, the experimental approaches and results that are presented in this thesis are summarized and discussed (Chapter 7).

Literature

1. **Abe, K., H. Hayashi, and P. C. Maloney.** 1996. Exchange of aspartate and alanine: mechanism for development of a proton-motive force in bacteria. J. Biol. Chem. **271:** 229-233.
2. **Abee, T., and J. A. Wouters.** 1999. Microbial stress response in minimal processing. Int. J. Food Microbiol. **50:**65-91.
3. **Arena, M. E., F. M. Saguir, and M. C. Manca de Nadra.** 1999. Arginine, citrulline and ornithine metabolism by lactic acid bacteria from wine. J. Bacteriol. **173:** 928-936.
4. **Arnold, C. N., J. McElhanon, A. Lee, R. Leonhart, and D. A. Siegele.** 2001. Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. J. Bacteriol. **183:**2178-2186.
5. **Axe, D. D., and J. E. Bailey.** 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol. Bioeng. **47:**8-19.
6. **Azcarate-Peril, M. A., E. Altermann, R. L. Hoover-Fitzula, Cano, and T. R. Klaenhammer.** 2004. Identification and inactivation of genetic loci involved with *Lactobacillus acidophilus* acid tolerance. Appl. Environ. Microbiol. **70:**5315-22.
7. **Baldi, P., and A. D. Long.** 2001. A Bayesian framework for the analysis of microarray expression data: regularized *t* test and statistical inferences of gene changes. Bioinformatics. **17:**509-519.
8. **Bernstein, J.A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen.** 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. USA. **99:**9697-9702.
9. **Bilban, M., L. K. Buehler, S. Head, G. Desoye, and V. Quaranta.** 2002. Normalizing DNA microarray data. Curr. Issues. Mol. Biol. **4:**57-64.
10. **Booth, I. R.** 1985. Regulation of cytoplasmic pH in bacteria. Microbiol. Rev. **49:**359-378.
11. **Cayley, S. and M. T. Record Jr.** 2004. Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate protein-DNA interactions and growth rate in osmotically stressed *Escherichia coli* K-12. J. Mol. Recognit. **17:**488-496.
12. **Churchill, G. A.** 2002. Fundamentals of experimental design for cDNA microarrays. Nat. Genet. **32 suppl.:** 490-495.

13. **Cleveland, W.S.** 1974. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Ass. **74:**829–836.
14. **Cook, G. M., and J. B. Russell.** 1994. The effect of extracellular pH and lactic acid on pH homeostasis in *Lactococcus lactis* and *Streptococcus bovis*. Curr. Microbiol. **28:**165-168.
15. **Csonka, L. N.** 1989. Physiological and genetic responses of bacteria to osmotic stress. Microbiol. Rev. **53:**121-147.
16. **Csonka, L. N., and A. D. Hanson.** 1991. Prokaryotic osmoregulation: Genetics and physiology. Annu. Rev. Microbiol. **45:** 596-606.
17. **Cunin, R., N. Glansdorff, A. Pierard, and V. Stalon.** 1986. Biosynthesis and metabolism of arginine in bacteria. Microbiol. Rev. **50:** 314-352.
18. **De Vries, M. C, E. E. Vaughan, M. Kleerebezem, and W. M. de Vos.** 2006. *Lactobacillus plantarum* - Survival, functional and potential probiotic properties in the human intestinal tract. Int. Dairy J. In press.
19. **Diez-Gonzalez, F., and J. B. Russell.** 1997. The effects of carbonylcyanide-*m*-chlorophenylhydrazone (CCCP) and acetate on *Escherichia coli* O157:H7 and K-12: uncoupling versus anion accumulation. FEMS Microbiol. Lett. **151:**71-76.
20. **Eisen M.B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. **95:**14863-14868.
21. **Galperin, M. Y.** 2005. The molecular biology database collection: 2005 update. Nucleic Acids Res. **33:**D5-D24.
22. **Giraud, E., B. Lelong, and M. Raimbault.** 1991. Influence of pH and initial lactate concentration on the growth of *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **36:** 96-99.
23. **Glaasker, E.**, **E. H. M. L. Heuberger, W. N. Konings, and B. Poolman.** 1998. Mechanism of osmotic activation of the quaternary ammonium compound transporter (QacT) of *Lactobacillus plantarum*. J. Bacteriol. **180:**5540-5546.
24. **Glaasker, E., W. N. Konings, and B. Poolman.** 1996. Glycine betaine fluxes in *Lactobacillus plantarum* during osmostasis and hyper- and hypo-osmotic shock. J. Biol. Chem. **271:**10060-10065.
25. **Glaasker, E., W. N. Konings, and B. Poolman.** 1996. Osmotic regulation of intracellular solute pools in *Lactobacillus plantarum*. J. Bacteriol. **178:**575-582.
26. **Glaasker, E., F. S. B. Tjan, P. F. Ter Steeg, W. N. Konings, and B. Poolman.** 1998. Physiological response of *Lactobacillus plantarum* to salt and nonelectrolyte stress. J. Bacteriol. **180:**4718-4723.

27. **Heyer L. J., S. Kruglyak, and S. Yooseph.** 1999. Exploring expression data: identification and analysis of coexpressed genes. Genome Res. **9:**1106-1115.
28. **Higuchi, T., H. Hayashi, and K. Abe.** 1997. Exchange of glutamateand gamma-aminobutyrate in a *Lactobacillus* strain. J. Bacteriol. **179:** 3362-3364.
29. **Hutkins, R. W., and N. L. Nannen.** 1993. pH homeostasis in lactic acid bacteria. J. Dairy. Sci. **76:** 2354-2365.
30. **Kashket, E. R.** 1987. Bioenergetics of lactic acid bacteria: cytoplasmic pH and osmotolerance. FEMS Microbiol. Rev. **46:**233-244.
31. **Kerr, M. K., and G. A. Churchill.** 2001. Experimental design for gene expression microarrays. Biostat. **2:**183-201.
32. **Kerr M. K., M. Martin, and G. A. Churchill.** 2000. Analysis of variance for gene expression microarray data. J. Comput. Biol. **7:**819-837.
33. **Kets, E. P. W., and J. A. M. de Bont.** 1997. Effect of carnitines on *Lactobacillus plantarum* subjected to osmotic stress. FEMS Microbiol. Lett. **146:**205-209.
34. **Kets, E. P. W., E. A. Galinski, and J. A. M. de Bont.** 1994. Carnitine: a novel compatible solute in *Lactobacillus plantarum*. Arch. Microbiol. **162:**243-248.
35. **Kets, E. P. W., M. Nierop Groot, E. A. Galinski, and J. A. M. de Bont.** 1997. Choline and acetylcholine: novel cationic osmolytes in *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **48:** 94-98.
36. **Kirkpatrick, C., L. M. Maurer, N. E. Oyelakin, Y. N. Yoncheva, R. Maurer, and J. L. Slonczewski.** 2001. Acetate and formate stress: opposite responses in the proteome of *Escherichia coli*. J. Bacteriol. **183:**6466-6477.
37. **Kleerebezem M., J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:** 1990-1995.
38. **Kolb, S., H. Otte, B. Nagel, and B. Schink.** 1992. Energy conservation in malolactic fermentation by *Lactobacillus plantarum* and *Lactobacillus sake*. Arch. Microbiol. **157:** 457-463.
39. **Konings, W. N., J. S. Lolkema, H. Bolhuis, H. W. van Veen, B. Poolman, and A. J. M. Driessen.** 1997. The role of transport processes in survival of lactic acid bacteria. Antonie van Leeuwenhoek. **71:** 117-128.
40. **Krieger, C.J., P. Zhang, L.A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S.Y. Rhee, and P. D. Karp.** 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. **32** (Database issue)**:** D438-442.
41. **Leung, Y. F., and D. Cavalieri.** 2003. Fundamentals of cDNA microarray data analysis. Trends Genet. **19:** 649-659.

42. **Long A. D., H. J. Mangalam, B. Y. Chan, L. Tolleri, G. W. Hatfield, and P. Baldi.** 2001. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. J. Biol. Chem. **276:**19937-19944.

43. **McDonald, M. C., H. P. Flemming, and H. M. Hassan.** 1990. Acid tolerance of *Leuconostoc mesenteroides and Lactobacillus plantarum*. Appl. Environ. Microbiol. **56:** 2120-2124.

44. **Molenaar, D., J. S. Bosscher, B. ten Brink, A. J. Driessen, and W. N. Konings.** 1993. Generation of a proton motive force by histidine decarboxylation and electrogenic histidine/histamine antiport in *Lactobacillus buchneri*. J. Bacteriol. **175:** 2864-2870.

45. **Molin, G.** 2001. Probiotics in foods not containing milk or milk constituents, with special reference to *Lactobacillus plantarum* 299v. Am. J. Clin. Nutr. **73:** 380S-385S.

46. **O'Byrne, C. P., and I. R. Booth.** 2002. Osmoregulation and its importance to food-borne microorganisms. Int. J. Food Microbiol. **74:** 203-216.

47. **Olsen, E. B., J. B. Russell, and T. Henick-Kling.** 1991. Electrogenic L-malate transport by *Lactobacillus plantarum*: a basis for energy derivation from malolactic fermentation. J. Bacteriol. **173:**6199-6206.

48. **Pavan, S. P. Desreumaux, and A. Mercenier.** 2003. Use of mouse models to evaluate the persistence, safety, and immune modulation capacities of lactic acid bacteria. Clin. Diagn. Lab. Immunol. **10:** 696-701.

49. **Pieterse, B., R. H. Jellema, and M. J. van der Werf.** 2006. Quenching of microbial samples for increased reliability of microarray data. J. Microbiol. Meth. **64:** 207-216.

50. **Polen, T., D. Rittmann, V. F. Wendisch, and H. Sahm.** 2003. DNA Microarray analyses of the long-term adaptive response of *Escherichia coli* to acetate and propionate. Appl. Environ. Microbiol. **69:**1759-1774.

51. **Poolman, B., and E. Glaasker.** 1998. Regulation of compatible solute accumulation in bacteria. Mol. Microbiol. **29:**397-407.

52. **Quackenbush, J.** 2002. Microarray data normalization and transformation. Nat. Genet. **32 suppl.:** 496-501.

53. **Renault P., C. Gaillardin, and H. Heslot.** 1988. Role of malolactic fermentation in lactic acid bacteria. Biochimie **70:** 375-379.

54. **Roeβler, M., and V. Műller.** 2001. Osmoadaptation in bacteria and archaea: common principles and differences. Env. Microbiol. **3:**743-754.

55. **Russell, J. B.** 1992. Another explanation for the accumulation of fermentation acids at low pH: anion accumulation versus uncoupling. J. Appl. Bacteriol. **73:** 363-370.

56. **Russell, J. B., and F. Diez-Gonzalez.** 1998. The effects of fermentation acids on bacterial growth. Adv. Microbial Physiol. **39:** 205-234.

57. **Salmond, C. V., R. G. Kroll, and I. R. Booth.** 1984. The effect of food preservatives on pH homeostasis in *Escherichia coli*. J. Gen. Microbiol. **130:** 2845-2850.

58. **Schena M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. **270:** 467-470.

59. **Simon, R. M., and K. Dobbin.** 2003. Experimental design of microarray experiments. Biotechniques **34 suppl.:**16-21.

60. **Shelef, L. A.** 1994. Antimicrobial effects of lactates, A review. J. Food Protection. **57:**445-450.

61. **Sleator, R. D., and C. Hill.** 2001. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. FEMS Microbiol. Rev. **26:**49-71.

62. **Smith, J. L., and S. A. Palumbo.** 1983. Use of starter cultures in meats. J. Food Protection. **46:** 997-1006.

63. **Smyth, G. K., and T. Speed.** 2003. Normalization of cDNA microarray data. Methods. **31:** 265-273.

64. **Tamayo P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky ,E. S. Lander, and T. R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. U. S. A. **96:**2907-2912.

65. **Tavazoie S., J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church.** 1999. Systematic determination of genetic network architecture. Nat. Genet. **22:**281-285.

66. **Ueno, Y., K. Hayakawa, S. Takahashi, and K. Oda.** 1997. Purification and characterization of glutamate decarboxylase from *Lactobacillus brevis* IFO 12005. Biosci. Biotechnol. Biochem. **61:** 1168-1171.

67. **van der Werf, M. J.** 2005. Towards replacing closed with open target selection strategies. Trends Biotechnol. **23:** 11-16.

68. **van der Werf, M. J., R. H. Jellema, and T. Hankemeier.** 2005. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. J. Ind. Microbiol. Biotechnol. **32:**234-252.

69. **van de Guchte, M., P. Serror, C. Chervaux, T. Smokvina, S. D. Ehrlich, and E. Maguin.** 2002. Stress responses in lactic acid bacteria. Antonie van Leeuwenhoek. **82:**187-216.

70. **Van Enckevort, F. H. J., C. Francke, B. Teusink, and R. J. Siezen.** 2005. LacplantCyc: a *Lactobacillus plantarum* Pathway / Genome Database as a reference for Lactic Acid Bacteria. www.lacplantcyc.nl.

71. **Van Gelder, R. N., M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas, and J. H. Eberwine.** 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Nat. Acad. Sci. **87:**1663-1667.

72. **Vesa, T., P. Pochart, and P. Marteau.** 2000. Pharmacokinetics of *Lactobacillus plantarum* NCIMB 8826, *Lactobacillus fermentum* KLD, and *Lactococcus lactis* MG 1363 in the human gastrointestinal tract. Aliment. Pharmacol. Ther. **14:** 823-828.

73. **Vescovo, M., S. Torriani, F. Dellaglio, and V. Bottazzi.** 1993. Basic characteristics, ecology and application of *Lactobacillus plantarum*: a review. Ann. Microbiol. Enzimol. **43:** 261-284.

74. **Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown.** 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. **99:**5860-5865.

75. **Weber, A., and K. Jung.** 2002. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. J. Bacteriol. **184:**5502-5507.

76. **Wood, J. M.** 1999. Osmosensing by bacteria: Signals and membrane-based sensors. Microbiol. Mol. Biol. Rev. **63:** 230-262.

77. **Wu, T. D.** 2001. Analysing gene expression data from DNA microarrays to identify candidate genes. J. Pathol. **195:** 53-65.

78. **Yale, J., and H. J. Bohnert.** 2001. Transcript expression in *Saccharomyces cerevisiae* at high salinity. J. Biol. Chem. **276:**15996-16007.

79. **Yamamoto, N., Y. Masujima, and T. Takano.** 1996. Reduction of membrane bound ATPase activity in a *Lactobacillus helveticus* strain with slower growth at low pH. FEMS Microbiol. Lett. **138:** 179-184.

80. **Yang, M. C. K., J. J. Yang, R. A. McIndoe, and J. X. She.** 2003. Microarray experimental design: power and sample size considerations. Physiol. Genomics **16:** 24-28.

Chapter 2

Quenching of microbial samples for increased reliability of microarray data.

Bart Pieterse, Renger H. Jellema and Mariët J. van der Werf

This chapter has been published as:
Pieterse, B., R. H. Jellema, and M. J. Van der Werf. 2006. Quenching of microbial samples for increased reliability of microarray data. J. Microbiol. Meth. **64:**207-216.

Abstract

Messenger RNA levels change on a minutes scale due to both degradation and *de novo* transcription. Consequently, alterations in the transcript profiles that are not representative for the condition of interest, are easily introduced during sample harvesting and work-up. In order to avoid these unwanted changes we have validated a -45°C methanol-based quenching method for obtaining reliable and reproducible 'snapshot' samples of *Lactobacillus plantarum* cells for transcriptome analyses. Transcript profiles of cells harvested with the quenching method were compared with transcript profiles of cells that were harvested according to two different commonly applied protocols. Significant differences between the transcript profiles of cells harvested by the different methods from the same steady-state culture were observed. In total, 42 genes or operons were identified from which the transcript levels were altered when the cells were not immediately quenched upon harvesting. Among these, several have previously been associated with cold-shock response. Furthermore, the reproducibility of transcript profiles improved, as indicated by the fact that the variation in the data sets obtained from the quenched cells was smaller than in the data sets obtained from the cells that were harvested under non-quenched conditions.

Introduction

Transcript levels can change on a minutes scale. Mean half-life values for *Escherichia coli* RNA are generally between three and seven minutes (1, 20). In the eukaryote *Saccharomyces cerevisiae,* half-life of mRNA varied between approximately 3 to over 90 minutes with a mean of 23 minutes (22). Furthermore, acute changes in the environment of the microorganism can also lead to changes in transcript levels within minutes (e.g. 23, 24). This suggests that significant changes in mRNA levels could already occur in the time required for concentrating and freezing cells obtained from liquid cultures using standard protocols.

In order to extract reliable biological information from large datasets such as transcriptomics data, it is essential that the numbers truly represent the biological situation of interest. Therefore, changes in the transcription profile should be avoided by quenching the cellular metabolism immediately upon harvesting.

Several methods have been described in literature in order to achieve this. Two methods that specifically aim at avoiding RNA degradation are the hot phenol method (15) and the guanidium thiocyanate method (14). Both methods require the use of relatively large amounts of highly toxic agents. Other quenching methods were originally developed for the isolation of metabolites, but may also be suitable for the isolation of RNA such as rapid filtration through an ultrafiltration membrane followed by immediate freezing of the cells (13, 19, 22), and dilution of the cells in a -45°C methanol solution (3).

Of these methods, the cold methanol method has the potential to be suitable for transcriptomics purposes. An important benefit of this method is that cells can easily be concentrated under quenched conditions. Moreover, since this method has already been proven useful for the isolation of metabolites, it may be highly interesting for combined approaches with multiple –omics technologies in systems biology approaches.

In this report we describe the validation of the cold methanol method for the quenching of cells to be analyzed by microarray analysis. The transcript profiles of *Lactobacillus plantarum* cells that were harvested using different protocols, among which the quenching protocol, were compared by microarray analysis. Subsequently, a combination of univariate and multivariate data analyses tools was applied to this data in order to obtain a comprehensive view of the effects of these different harvesting methods on the transcript profiles.

Materials and methods

Microorganism and cultivation

The microorganism used is *Lactobacillus plantarum* WCFS1. The complete genome sequence from this microorganism is available (12).
Continuous culturing was performed on MRS medium (55 g/l, Difco), at 37°C, pH 6.0 in an Applikon bioreactor with a working volume of 1 liter. The culture was maintained anoxic by a continuous overlay of nitrogen gas. The dilution rate was set at 0.3 h^{-1}.

Cell harvesting

Samples were taken from a single steady state culture by three different harvesting methods:
In the "standard method", 40 ml of sample was taken from the bioreactor with a syringe through the septum and added to a centrifuge tube which was cooled on ice. Cells were concentrated immediately by centrifugation at 5000 g for 8 minutes at 0°C. The supernatant was discarded and the cell material was immediately frozen in liquid nitrogen.

For the second method ("rapid spin method"), the sample was taken from the fermenter with a syringe through the septum and added to 2 ml microcentrifuge tubes. The suspensions were centrifuged for 30 seconds at 15,000 g at room temperature. The supernatant was discarded and the tubes were immediately frozen in liquid nitrogen. The estimated time between harvesting and freezing was 3 minutes.

For the third method ("quenching method")(3), the sample was collected from the overflow of the fermenter and immediately added to a centrifuge tube containing -45°C quenching buffer (60% methanol, 62.5 mM HEPES, pH 6.5). The estimated passage time in the overflow tube was 10 seconds. To 160 ml of buffer, 40 ml of culture sample was added. The content of the tube was stirred vigorously during harvesting to avoid formation of ice-crystals. During harvesting the temperature of the suspension stayed below -20°C. After addition of the sample, the tube with the quenched cell suspension was cooled down to -45°C in the ethanol bath. The samples were centrifuged for 10 minutes at 20,000 g at -20°C. The supernatant was discarded, and the cell pellet was immediately frozen in liquid nitrogen. All cell pellets were stored at -80°C until further processing.

Cell lysis and RNA isolation

RNA was isolated in triplicate from cells harvested by the three different methods. A frozen cell pellet was added to a microcentrifuge tube containing 0.5 g of 0.1 mm zirconia

beads (Biospec Products, Bartlesville), 200 µl of 40 mg/ml Macaloid gel (Kronos Titan GmbH, Leverkusen), 100 µl of 10% SDS, 500 µl of phenol (pH 5.9) and 200 µl MilliQ water. The cells were disrupted by grinding. After disruption, the samples were centrifuged for 10 minutes at 15,000 g (4°C). The aqueous phase was transferred to a clean tube and an equal volume of phenol was added after which the tubes were mixed for 1 minute. Subsequently, an equal volume of chloroform was added and the tubes were mixed again. The samples were centrifuged for 10 minutes at 15,000 g (4°C) and the aqueous phase was transferred to a clean tube. The phenol/chloroform treatment was repeated until a clear interphase appeared after centrifugation. The aqueous phase was treated once more with an equal volume of chloroform. After centrifugation the aqueous phase was transferred to a clean tube. A tenth volume 3M sodium acetate and 2.5 volumes of ethanol were added. The samples were incubated overnight at -20°C. RNA was precipitated by centrifugation at 20,000 g for 45 minutes at 4°C. The pellets were washed once with 70% of ethanol and dissolved in RNase free MilliQ water. RNA purity and concentration were determined both spectrophotometrically and on agarose gel. The RNA isolates were checked for residual RNase activity by the comparison of samples that were incubated for 1 hour at 42°C with the initial material on an agarose gel.

Array design

The microarray that was used is a clone based array. The microarray contains approximately 80% of the *L. plantarum* WCFS1 genome. In total, 3714 identified genomic fragments were amplified from the genomic *L.plantarum* WCFS1 library in pBlueScript SK+ (10) by PCR using forward and reverse primers with 5'-C6 aminolinkers to facilitate crosslinking to the aldehyde coated glass slides. PCR products were purified by ethanol precipitation. Purified PCR products were dissolved in 3xSSC (SSC: 150mM sodium chloride, 17mM sodium citrate, pH 7.2), after which clones were arrayed in a controlled atmosphere on CSS-100 silylated aldehyde glass slides (Telechem, USA), with quill pins (Telechem SMP3) in a SDDC 2 Eurogridder (ESI, Canada). After drying, the slides were blocked with borohydride. The average size of the fragments is 1.2 kb with an average deviation of 0.3 kb.

Fluorescent labeling and hybridization

Differential transcript levels were determined by two-color fluorescent hybridizations of the corresponding cDNA's on the *L. plantarum* WCFS1 clone-array. The samples were labeled by random hexamers primed *in vitro* reversed transcription with either Cy5- or Cy3-labeled dUTP (Amersham Biosciences) (Table 1). Unincorporated dyes were

removed by using autoseq G50 columns (Amersham Biosciences). The arrays were prehybridized for 45 minutes at 42°C in prehybridization solution (1% BSA, 5x SCC and 0.1 % SDS). Co-hybridizations of the labelled cDNA samples were performed overnight at 42 ºC in Easyhyb buffer (Roche) according to the manufacturer's protocol. The slides were subsequently washed twice in 1xSSC, 0.2% SDS, once in 0.5xSSC, and twice in 0.2xSSC.

Table 1. Microarray hybridization scheme for the comparison of the effects of the three different cell harvesting methods on expression profiles. The RNA of all the samples originates from the same steady state culture. RNA isolations from the cells that were harvested according to the different methods were performed in triplicate.

Array number	Cy3-labeled sample (reference)	Cy5-labeled sample	Goal
1	Quenching method, RNA isolate 1	Quenching method, RNA isolate 3	Technical variation
2	Quenching method, RNA isolate 2	Quenching method, RNA isolate 1	
3	Quenching method, RNA isolate 3	Quenching method, RNA isolate 2	
4	Quenching method, RNA isolate 1	'Rapid Spin' method, RNA isolate 1	Effect of 'rapid spin' method on transcription profile
5	Quenching method, RNA isolate 2	'Rapid spin' method, RNA isolate 2	
6	Quenching method, RNA isolate 3	'Rapid spin' method, RNA isolate 3	
7	Quenching method, RNA isolate 1	'Standard' method, RNA isolate 1	Effect of 'standard' method on transcription profile
8	Quenching method, RNA isolate 2	'Standard' method, RNA isolate 2	
9	Quenching method, RNA isolate 3	'Standard' method, RNA isolate 3	

Image analysis

The fluorescent signals from the two different labels on the hybridized arrays were quantified with a ScanArray Express (Packard Bioscience) scanner and Imagene 4.2 software (BioDiscovery, Inc.). Spots of which the difference between the mean signal of the spot and the mean signal of the background was smaller than two times the background standard deviation were depicted as empty, and were excluded from further analysis. Spots from which the signal in one or both of the channels exceeded the detection limit of the scanner were also excluded.

Normalization

Within-slide, intensity dependent normalizations were performed with the scatter plot smoother LOWESS-option (26) from the software program DataFit 8.0 (Oakdale Engineering). The user-defined fraction of data that was used for smoothing at each point, was set at 20% for all slides.

Significance analysis

Prior to significance analysis, a data transformation was applied as described in the results section. Significance analysis was performed by means of a 1way-ANOVA. Subsequently a Tukey HSD test was performed to determine whether a significant differential expression level (99% confidence interval) was observed under a specific condition. If genes or predicted operons were present on multiple spots on the array, these spots were considered as replicates in the significance analysis.

By comparison of overlapping regions of the genomic fragments on the array, it was in most cases possible to predict which of the specific genes that were present on the clones was, or were the ones affected.

Principal component analysis (PCA)

Principal Component Analysis (PCA) was performed on the normalized, transformed data using the Eigenvector PLS Toolbox (Eigenvector Research Inc.) within the Matlab environment (The MathWorks). The data were autoscaled to produce zero mean and unit variation variables prior to PCA. The complexity (number of principal components) of the model was tested using the Leave One Out (LOO) Cross validation test.

Hierarchical clustering

For the hierarchical clustering and visualization of the results, the programs CLUSTER and TREEVIEW were applied (5). Only those spots were included that fall within the 90% confidence interval of the 1way-ANOVA. Average linkage clustering was performed on the natural logarithm of the LOWESS-normalized ratios.

Results

Experimental design

The effects of three different cell harvesting methods on transcript profiles were compared by microarray analysis. In order to exclude differential expression levels due to biological spreading, all samples were harvested from the same steady state culture within one hour. Harvesting was performed according to the quenching method (Q), the “rapid spin” method (R) or the “standard” method (S) (Fig. 1; see also the materials and methods section). The main difference between these three methods was the time between harvesting and the assumed arrest of both RNA degradation and *de novo* synthesis. For the “standard” and the “rapid spin” harvesting methods, the assumed arrest took place when the samples were frozen

in liquid nitrogen. This was after approximately 10 and 3 minutes for the "standard" and the "rapid spin" method, respectively. For cells harvested by the quenching method, the arrest of metabolism took place after approximately 10 seconds, when the sample was mixed with the methanol-based buffer.

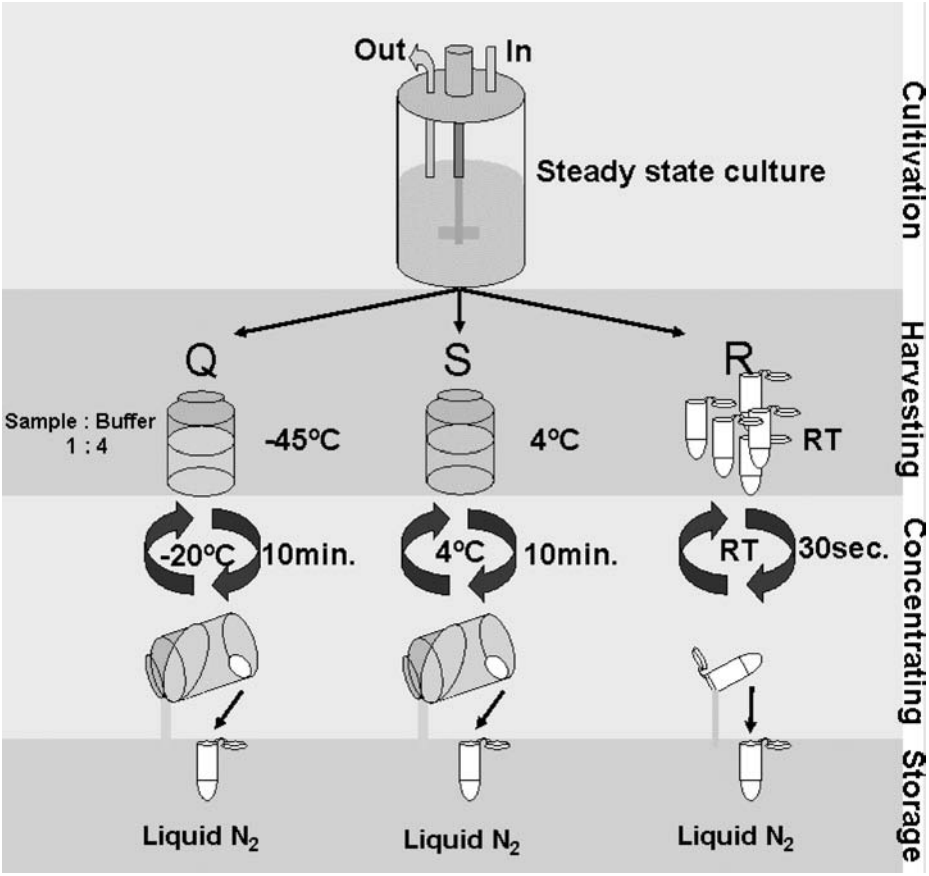

Figure 1. Schematic representation of the three cell harvesting protocols compared in this study. Concentrating was achieved by means of centrifugation and discarding of the supernatant. Cell material was immediately frozen with liquid nitrogen after the concentration step.

In order to be able to distinguish between alterations in the expression profiles introduced during the cell harvesting procedure and differences due to variations in RNA purification, labeling-, or hybridization efficiency, triplicate RNA isolations were performed after which each sample was labeled independently. No effect was observed from the different harvesting methods on the yield of the RNA isolations.

The microarray hybridizations were divided into three groups of three hybridizations (Table 1). The reference material in all groups was Cy3-labeled cDNA, transcribed from the

RNA of one of the three independent isolations from the quenched cells. In the first group the Cy5 signal represents one of the independent RNA isolates from the quenched cells. Consequently, apparent differential expression ratios in this group are introduced during RNA purification or caused by differences in labeling or hybridization efficiencies.

Data processing and significance analysis

After removal of the empty spots and the spots from which the signal exceeded the detection limit of the scanner, 2622 spots remained for further analysis.
After LOWESS normalization (26), ANOVA and a Tukey HSD test were applied to identify spots that showed a significantly different behavior (p=0.1/ 0.01) between any of the three groups. Since ANOVA requires a normal distribution of the input data, a data transformation was applied on the expression ratios:

$$Cy5 / Cy3 \geq 1 \rightarrow (Cy5 / Cy3) - 1$$
$$Cy5 / Cy3 < 1 \rightarrow - (1/ (Cy5 / Cy3)-1)$$

The obvious advantages of this transformation are that the distribution of the data of the resulting datasets is more suitable for significance analysis, and that the sign of the numbers (positive or negative) indicates whether the corresponding transcript either shows a higher or a lower expression. Meanwhile, since the input data for the transformation originate from the LOWESS normalized data from one slide, data manipulation introduced during between slide normalizations can be avoided.

Effect of the harvesting methods on the expression ratios

In order to obtain an unbiased overall view on the relatedness between the nine independent datasets, the LOWESS normalized ratio files were analyzed by principal component analysis (PCA). This method allows for the simultaneous visualization of variables (spots) in only a few dimensions by means of a mathematical reduction of the many parameters. The PCA score plot presented in Figure 3 is a 2-dimensional representation of the correlation matrix of the response variables of the nine independent transformed ratio files. In this plot the first principal component (PC1) accounts for as much variability between the datasets as possible. The second principal component (PC2) accounts for as much of the remaining variability as possible. As can be seen in this score plot, the three different harvesting methods are successfully classified by the PCA, indicating a significant effect of the harvesting methods on the transcript profiles. Interestingly, the order in which the three groups are separated on the first principal component corresponds to the assumed time between harvesting and arrest of cellular activity.

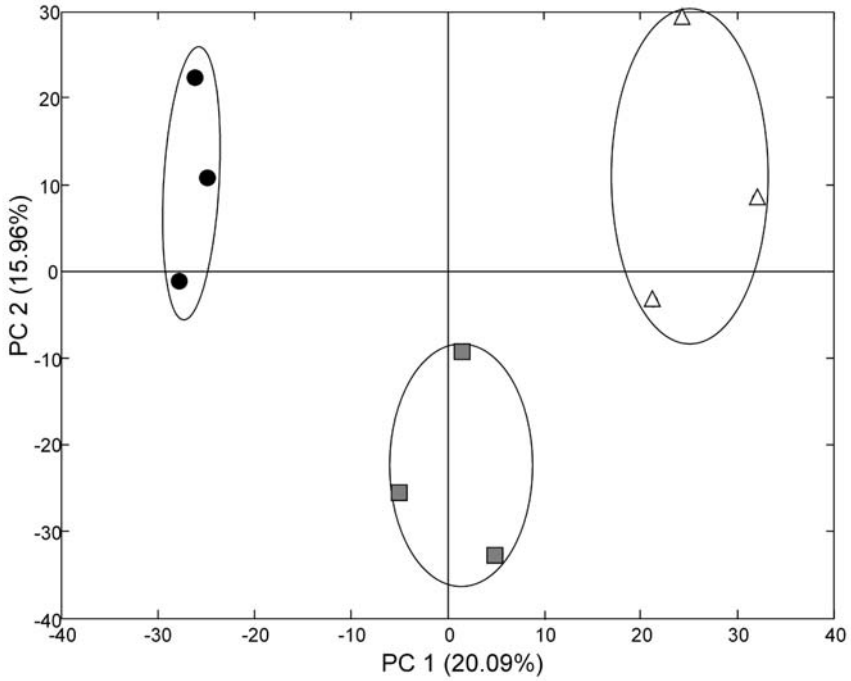

Figure 2. Score plot of the principal component analysis (PCA) on the normalized, transformed micro-array datasets of the "standard" harvesting method (Δ), the "rapid spin" harvesting method (■) and the quenching method (●).

The spreading within the three groups is presumably caused by variations introduced during harvesting, RNA isolation and micro-array analysis. This experimental variation is smaller in the group of the quenched cells since the data points in this group are less separated by the first principal component than the data points in the groups of the cells that were harvested according to the other two methods (Fig. 2), indicating a higher reproducibility for the samples that were harvested according to the quenching method.

Identification of affected transcripts

In order to obtain a visual representation of the different effects of the harvesting methods on the transcript levels, clones that showed a differential signal between any of the three harvesting methods (90% confidence interval of the ANOVA) were selected for average linkage clustering. The resulting hierarchical cluster (Fig. 3) shows the expression trends in the triplicate datasets from the three harvesting methods.

Most of the changes between the three groups are caused by differences in expression levels between the samples from the "rapid spin" and / or the "standard" harvested samples and the samples that were harvested according to the quenching protocol. For some

spots, differential expression levels are indicated in the group of the quenched versus the quenched samples. The corresponding transcripts are probably affected during RNA isolation or by technical spreading introduced by the micro-array analysis. The differences include both transcripts that showed a decreased expression level (degradation) and transcripts that showed an increased expression level (*de novo* synthesis).

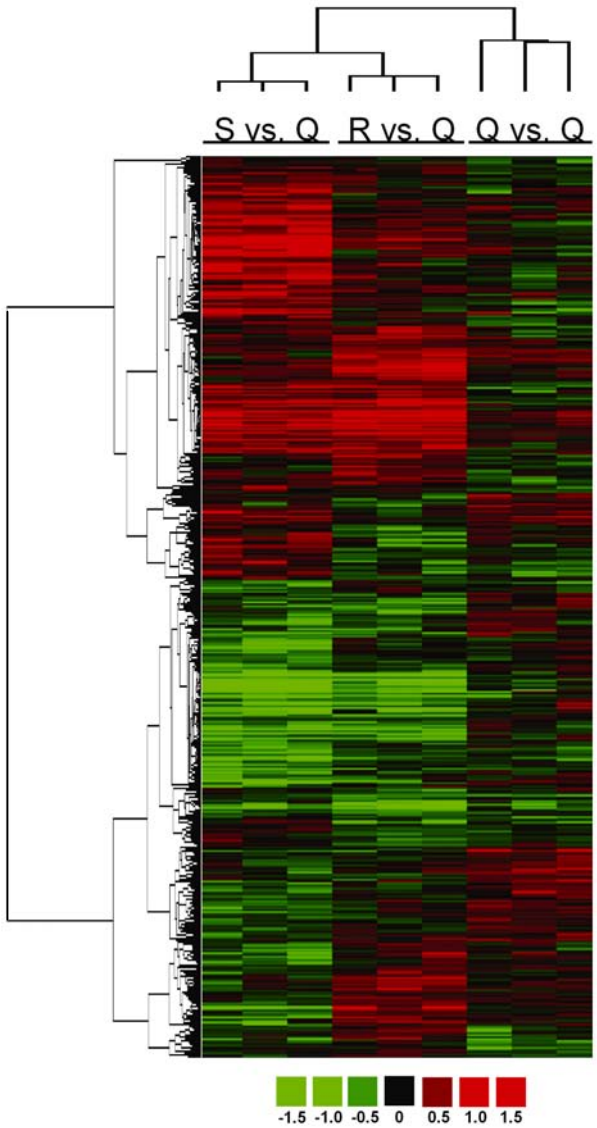

Figure 3. Hierarchical cluster of the normalized, log-transformed data of all micro-array spots from which expression ratios were within the 90% confidence interval of the 1way-ANOVA. Q = RNA from quenched cells, R = RNA from cells harvested with the "rapid spin" method, S = RNA from cells harvested with the "standard" method.

Table 2. Genes from which expression significantly changed (99% confidence interval Tukey HSD test) in response to the "rapid spin" and / or the "standard" harvesting method. Fold changes depicted in bold are significant. R = RNA from cells harvested with the "rapid spin" method, S = RNA from cells harvested with the "standard" method. Hyp. = hypothetical protein.

Orf-nos	Assigned function	Fold change	
		R	S
1700 – 1701	tspO / hyp.	0.5	**0.3**
2027 – 2029	dnaK / grpE / hrcA	0.7	**0.3**
0727 – 0728	groES / groEL	0.8	**0.4**
0725 or 0726	hyp. or putative cell surface protein	0.7	**0.5**
0729	putative transporter	0.8	**0.5**
0746 – 0751	phosphate transport system	1.1	**0.5**
0213 - 0214	Hyp. / hyp.	1.4	**0.5**
3327	cadA	0.8	**0.6**
0329	acdH	0.9	**0.6**
1928 – 1929	nrpR2 / hyp.	0.9	**0.6**
0481	pyrG	1.1	**0.6**
0360 – 0362	putative regulator / hyp. / accB3	0.8	**0.7**
0802 – 0803	glnPH1 / glnQ1	0.8	**0.7**
2325 – 2326	thiI / csd2	0.8	**0.7**
2696 - 2703	pyrimidine biosynthesis cluster	**0.4**	**0.2**
3536	bsh1	**0.5**	**0.4**
3270 – 3271	purA / guaC	**0.5**	**0.5**
3305	pflB2	**0.7**	**0.6**
3437 – 3438	trxA3 / hyp.	**0.7**	**0.7**
1670 – 1682	fatty acid biosynthesis cluster	**0.4**	0.5
2693 – 2694	rexA / rexB	**0.6**	0.7
2151 – 2152	pdhD / pdhC	1.0	**2.6**
1289 – 1290	hyp. / hyp.	1.3	**2.5**
2016 – 2021	dltD / dltC1 / dltB / dltA / dltX / pbpX2	1.2	**2.5**
0254 – 0256	cysE / metC1 / cysK	1.3	**2.3**
2931 – 2932	nrdG / nrdC	1.2	**2.2**
0574 – 0577	hyp. / pts9AB / pts9C / pts9D	1.2	**2.0**
2647 – 2651	pts19A / pts19D / pts19C / pts19B / transcription regulator	1.1	**2.0**
0208 – 0209	hyp. / putative cell surface protein	1.1	**1.8**
0396	putative transcripton regulator	1.3	**1.8**
0621 – 0622	rplJ / rplL	1.0	**1.8**
2278 – 2279	rhe3 / hyp.	1.2	**1.8**
0168 – 0171	dak1B / dak2 / hyp. / gplF2	1.4	**1.7**
0230 – 0233	mtlA / mtlR / mtlF / mtlD	1.2	**1.7**
2872 – 2874	oxidoreductase / adh2 / oxidoreductase	1.5	**1.7**
1229	putative cell surface protein	1.1	**1.5**
0172 – 0174	putative regulator / putative regulator / agl1	**1.6**	**1.6**
2189 – 2203	cell devision / cell envelope biosynthesis cluster	**1.4**	**1.6**
0813 – 0816	hyp. / murB / putative antiporter / putative regulator	**1.6**	**1.5**
2525	putative transport protein	**2.1**	1.4
1136	hyp.	**2.0**	1.2
0286	pts6C	**1.6**	1.5

In addition to ANOVA, a Tukey HSD test was applied to identify genes or assumed operons that showed a significantly different expression level (99% confidence interval) between the quenched and the other samples.

In the samples that were harvested by the "standard" method, transcripts from 19 genes or operons showed a significantly decreased expression level, and transcripts from 18 genes showed a significantly increased expression level (Table 2). The "rapid spin" method, which has a shorter time between harvesting and arrest, shows less changes in the 99% confidence interval: 7 transcripts with a significantly lower expression level, and 6 transcripts with a significantly higher expression level.

Discussion

In this paper we describe the validation of a methanol-based quenching method (3), for the reproducible isolation of representative RNA from *L. plantarum*. The effect of quenching on transcript profiles was compared with the effect of two commonly used harvesting methods for which the assumed time between harvesting and arrest of cellular activity was much larger. The results from this study clearly show that harvesting under quenched conditions has positive effects on both the reliability and reproducibility of transcriptomics data.

Principal component analysis revealed two major effects of the harvesting procedures on the transcript profiles: the three groups of the triplicate datasets are clearly classified by the first principal component, indicating that the predominant differences between the independent datasets were caused by the harvesting methods that were applied. Moreover, the variation within the triplicate groups is smaller in the dataset of the quenched cells. The distance between the three method groups in the PCA score plot increased with the assumed time between harvesting and arrest of the corresponding samples, which indicates an important contribution of the time effect. This also indicates that the samples from the cells that were harvested according to the quenching method are most representative for the actual transcription profile during cultivation.

Variation in transcriptome datasets is potentially introduced during several steps in the experiment and can roughly be divided into two classes: Biological variation and technical variation. Biological variation, introduced during culturing, harvesting and RNA isolation, influences the true amount of transcript that is present in the samples. Technical variation is

introduced during labeling, hybridization and detection and influences the perceived amount of transcript, represented by the amount of signal in each spot on the array. Since, in recent years, ongoing efforts have been made in minimization and normalization of the technical variation, the relative contribution of the biological variation to micro-array datasets increases.

Cell harvesting under quenched conditions did not only limit alterations of specific transcript levels, but also variation within the replicate datasets (Fig. 2). During analysis of transcriptome data, such a decrease of the spreading within replicate samples results in an increase of the amount of genes from which reliable conclusions can be drawn.

A more detailed analysis of the transcripts affected by the cell harvesting method (Table 2) showed that several genes have previously been associated with cold shock response in other microorganisms, which is probably the most obvious stress the cells are subjected to during harvesting. The potential impact of environmental changes (such as cold shock) during work-up on transcript profiles has previously been recognized in other prokaryotes (7, 8, 14).

The chaperone encoding genes groES / groEL and the dnaK stress protein encoding gene showed a lower expression level under the non-quenched conditions. Mutations in *E. coli* groES, groEL and dnaK genes have been described to suppress cold sensitivity (4). Moreover, increased expression of groES / groEL coincides with a reduced survival at low temperatures (10). Reduced presence of GroES in response to cold shock has also been observed on 2D-page for *B. subtilis* (6). The metC1 gene of *L. plantarum* showed an increased expression level in response to cell harvesting under non-quenched conditions. In *B. subtilis* metC shows a higher expression in response to a cold shock (9). Furthermore, in *L. plantarum* the rhe3 RNA helicase showed increased expression level upon harvesting under non-quenched conditions. Yamanaka *et al.* showed that RNA helicase plays a critical role in cold shock adaptation in *E. coli* (25).

These results show that part of the differential transcript levels that were observed in this study are involved in specific adaptation responses on the RNA level that occurred after harvesting of the cells, which once more indicates that the quenched samples are most representative for the transcription profile during cultivation.

Three lines of evidence clearly demonstrate the effect of the cell harvesting protocol on the transcription profile: (i) the transcript profiles of the independent samples were grouped according to their harvesting method (Fig. 2); (ii) The order in which the three groups, representing the different harvesting methods, appeared in the score plot of the PCA (Fig. 2) corresponds to the time between harvesting and arrest of cellular activity; (iii) Among

the group of genes that showed increased expression in response to the applied harvesting methods, several have previously been associated with a cold shock response, which is one of the major stress factor during cell harvesting.

The fact that the increased expression of cold shock related genes has previously been associated with harvesting under non-quenched conditions (7, 14) confirms that the observed relative changes in gene expression can not be explained by alterations that are introduced to the transcriptome during harvesting under quenched conditions. This positive effect of the quenching method is to be attributed to the rapid mixture of the cell suspension with the quenching buffer, which results in an immediate chilling of the cells to a temperature below -20ºC.

Although several scientists involved in transcriptomics experiments are currently using quenching protocols for cell harvesting (2, 14, 16), others are still relying on rapid centrifugation (e.g. 17, 18). Moreover, in many cases details about the cell harvesting method are not disclosed at all (e.g. 11, 21, 23).

Recent experiments in which this quenching protocol was applied for the extraction of metabolites from *L. plantarum*, also showed good results (A. Mars and M. Faijes, unpublished data), indicating that the method can indeed be applied for combined RNA / metabolite studies in systems biology-type of studies.

We conclude that the methanol-based quenching method has a significant effect on both the reliability and the reproducibility of micro-array data. Based on metabolomics studies with both prokaryotes and eukaryotes in which the same quenching protocol was applied we expect that this method is applicable for a broad range of microorganisms.

Achnowledgements

We thank Ted van der Lende and Evelyn Wesseling for their contribution in the microarray analysis, Bianca van der Werff-van der Vat for her contribution in the data analysis and Rob Leer, Michiel Kleerebezem and Willem de Vos for their useful suggestions.

Literature

1. **Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen.** 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. **99:**9697-9702.
2. **Boer, V. M., J. H. de Winde, J. T. Pronk, and M. D. Piper.** 2003. The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. J. Biol. Chem. **278:**3265-3274.
3. **De Koning, W., and K. van Dam.** 1992. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal. Biochem. **204:**118-123.
4. **Donnelly, C. E., and G. C. Walker.** 1989. groE mutants of *Escherichia coli* are defective in umuDC-dependent UV mutagenesis. J. Bacteriol. **171:**6117-6125.
5. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA. **95:**14863-14868.
6. **Graumann, P., K. Schroder, R. Schmid, and M. A. Marahiel.** 1996. Cold shock stress-induced proteins in *Bacillus subtilis*. J. Bacteriol. **178:**4611-4619.
7. **Gualerzi, C. O., A. M. Giuliodori, and C. L. Pon.** 2003. Transcriptional and post-transcriptional control of cold-shock genes. J. Mol. Biol. **331:**527-539.
8. **Hinton, J. C. D., I. Hautefort, S. Eriksson, A. Thompson, and M. Rhen.** 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. Curr. Opin. Microbiol. **7:**277-282.
9. **Kaan, T., G. Homuth, U. Mader, J. Bandow, and T. Schweder.** 2002. Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. Microbiology. **148:**3441-3455.
10. **Kandror, O., and A. L. Goldberg.** 1997. Trigger factor is induced upon cold shock and enhances viability of *Escherichia coli* at low temperatures. Proc. Natl. Acad. Sci. **94:**4978-4981.
11. **Khil, P. P., and R. D. Camerini-Otero.** 2002. Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. Mol. Microbiol. **44:**89-105.
12. **Kleerebezem M., J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:**1990-1995.

13. **Lowry, O. H., J. Carter, J. B. Ward, and L. Glaser.** 1971. The effect of carbon and nitrogen sources on the level of metabolic intermediates in *Escherichia coli*. J. Biol. Chem. **246:**6511-6521.
14. **Mangan, J. A., I. M. Monahan, and P. D. Butcher.** 2002. Gene expression during host-pathogen interactions: Approaches to bacterial mRNA extraction and labelling for microarray analysis. Meth. Microbiol. **33:**137-151.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor, NY.
16. **Piper, M. D., P. Daran-Lapujade, C. Bro, B. Regenberg, S. Knudsen, J. Nielsen, and J. T. Pronk.** 2002. Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of Saccharomyces cerevisiae. J. Biol. Chem. **40:**37001-37008.
17. **Price, C. W., P. Fawcett, H. Ceremonie, N. Su, C. K. Murphy, and P. Youngman.** 2001. Genome-wide analysis of the general stress response in *Bacillus subtilis*. Mol. Microbiol. **41:**757-774.
18. **Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner.** 1999. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. **27:**3821-3835.
19. **Saez, M. J., ad R. Lagunas.** 1976. Determination of intermediary metabolites in yeast. Critical examination of the effect of sampling conditions and recommendations for obtaining true levels. Mol. Cell Biochem. **13:**73-78.
20. **Selinger, D.W., R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow.** 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res. **13:**216-223.
21. **Tjaden, B., R. M. Saxena, S. Stolyar, D. R. Haynor, E. Kolker, and C. Rosenow.** 2002. Transcriptome analysis of Escherichia coli using high-density oligonucleotide probe arrays. Nucleic Acids Res. **30:**3732-3738.
22. **Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown.** 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci **99:**5860-5865.
23. **Weber, A., and K. Jung.** 2002. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. J. Bacteriol. **184:**5502-5507.
24. **Yale, J., and H. J. Bohnert.** 2001. Transcript expression in *Saccharomyces cerevisiae* at high salinity. J. Biol. Chem. **276:**15996-16007.

25. **Yamanaka, K, and M. Inouye.** 2002. Selective mRNA degradation by polynucleotide phosphorylase in cold shock adaptation in *Escherichia coli.* J. Bacteriol. **183:**2808-2816.

26. **Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed.** 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucl. Acids. Res. **30:**e15 [online].

Chapter 3

Mathematical design of prokaryotic clone-based microarrays

Bart Pieterse, Elisabeth J. Quirijns, Frank H. J. Schuren, and Mariët J. van der Werf

This chapter has been published as:
Pieterse, B., E. J. Quirijns, F. H. J. Schuren, and M. J. Van der Werf. 2005. Mathematical design of prokaryotic clone-based microarrays. BMC Bioinformatics **6:**238.

Abstract

Clone-based microarrays, on which each spot represents a random genomic fragment, are a good alternative to open reading frame-based microarrays, especially for microorganisms for which the complete genome sequence is not available. Since the generation of a genomic DNA library is a random process, it is beforehand uncertain which genes are represented. Nevertheless, the genome coverage of such an array, which depends on different variables like the insert size and the number of clones in the library, can be predicted by mathematical approaches. When applying the classical formulas that determine the probability that a certain sequence is represented in a DNA library at the nucleotide level, massive amounts of clones would be necessary to obtain a proper coverage of the genome. This paper describes the development of two complementary equations for determining the genome coverage at the gene level. The first equation predicts the fraction of genes that is represented on the array in a detectable way and cover at least a set part (the minimal insert coverage) of the genomic fragment by which these genes are represented. The higher this minimal insert coverage, the larger the chance that changes in expression of a specific gene can be detected and attributed to that gene. The second equation predicts the fraction of genes that is represented in spots on the array that only represent genes from a single transcription unit, which information can be interpreted in a quantitative way. Validation of these equations shows that they form reliable tools supporting optimal design of prokaryotic clone-based microarrays.

List of abbreviations

BP_b	number of base pairs within the operon in front of the specific gene	[bp]
BP_e	number of base pairs within the operon behind the specific gene	[bp]
DIC	predefined minimal insert coverage, i.e. the minimal required representation of the gene on the insert	[%]
Gene	gene size	[bp]
GS	genome size	[Mbp]
IS	insert size	[bp]
$IS_{opt\text{-}MIC}$	optimal value of IS for the MIC equation	[-]
$IS_{opt\text{-}GSI}$	optimal value of IS for the GSI equation	[-]
N	number of clone	[-]
O_b	overlap of the fragment and the beginning of the gene	[bp]
O_e	overlap of fragment and the end of the gene	[bp]
O_{mf}	minimal required overlap of the fragment and the gene (fixed)	[bp]
O_{mv}	minimal required overlap of the fragment and the gene (variable)	[bp]
p	gene specific probability value	[-]
p_{GSI}	predicted fraction of specifically represented genes	[-]
p_{MIC}	predicted fraction of genes represented with a minimal insert coverage	[-]
R	average number of genes per transcription unit	[-]
SSQ	Residual Sum of Squares	[-]
a-l	parameters in MIC- and / or GSI-model	[-]

Introduction

In the past decade, whole transcriptome comparison by microarray hybridizations has proven to be an effective tool for studying genome wide gene responses. The general approaches for the development of microarrays are based on the completely annotated genome sequence of an organism. Usually each spot on the array represents one open reading frame (ORF). Whereas this approach has clear advantages for strains for which the complete annotated genome sequence is available, it is not applicable to strains for which this is not the case.

A method that allows for the rapid construction of microarrays for which the completely annotated genome sequence is not required is by the construction of a clone-based array. In this approach, a chromosomal DNA library is constructed from the strain of interest. From this library the genomic fragments, the inserts, are amplified from the clones by PCR with generic primers and spotted on the array-slide (2, 3).

The major differences between ORF-based and clone-based arrays with respect to the data interpretation are that in case of clone-based arrays the differential signals can only be linked to a specific gene after the DNA fragment from the spot of interest on the array has been sequenced, and that it is beforehand uncertain whether a gene is represented on the array. Moreover, whereas ORF-based microarrays exclusively generate gene specific data, a differential signal within a spot on a clone-based array can originate from multiple genes on the insert that are not necessarily linked at the transcriptional level.

The extent of these limitations can be quantified by estimating the genome coverage by the spots present on the array. The standard formulas for estimating the genome coverage of a DNA library, the Clark-Carbon equation (5) and the Lander-Waterman equation (7), determine this coverage at the nucleotide level. In other words, they consider the genome as a set of nucleotides, which is useful when the library is to be used for genome sequencing. However, these formulas will overestimate the required number of clones for hybridization purposes. The reason for this is that for hybridization purposes small overlapping fragments that allow for specific binding of the labeled cDNA suffice. Akopyants *et al.* (1) developed an equation for the estimation of the fraction of genes that are at least partially represented. This formula is directly derived from classical probability calculations and contains the organism specific variables genome size and average gene size. Due to the fact that Akopyants *et al.* determine the genome coverage at the gene level, and consider a gene represented if a fragment is present that is large enough to hybridize to and large enough to identify the gene, the required number of clones to obtain a certain coverage is reduced.

A general drawback of these three formulas is that they give no insight into the fraction of genes for which specific data can be generated in a transcriptomics experiment. The data from a spot are considered specific if the expression ratios from the quantified signal from that spot can directly be related to the gene(s) represented by the spot. This is not the case if DNA from multiple (neighboring) transcription units is present in one spot, since it would be uncertain which gene is responsible for which part of the total signal from that spot.

In this paper, two formulas were developed that enable for mathematical predictions of genome coverage by a prokaryotic clone-based array at the gene level as a function of genome size, number of clones, insert size, and either the minimal part of the insert that is covered by the gene or the minimal overlap of the gene and the insert: the minimal insert coverage (MIC) equation, and the gene specific information (GSI) equation.

In order to develop equations that are applicable to a broad range of microorganisms, model datasets were generated for 15 prokaryotes originating from several genera (Table 1) that functioned as templates on which the MIC- and GSI-equations were fitted. The resulting formulas were validated on 10 other prokaryotic species.

Description of the developed equations

Minimal Insert Coverage (MIC)-equation

Since the generation of inserts for a genomic DNA library is a random process, a large part of the represented genes may be co-represented with other genes by one spot on the microarray. This complicates data interpretation since it introduces an uncertainty on which gene or genes are responsible for differential signals from these spots. The impact of differential expression of a specific gene on the observed difference of the signal from a spot will be larger when a larger part of the genome fragment in that spot is covered by that gene. Moreover, the larger the part of the insert that is covered by a specific gene, the larger the chance that differential signals for the spot can be attributed to that gene, and the higher the chance that differential expression levels from that gene result in a statistically significant differential signal on the array.

The MIC-equation anticipates to this effect by predicting the number of genes that are (at least partially) present on an insert *and* cover at least a predefined part of the insert (*DIC*). This predefined part is defined as a percentage of the total insert. E.g. if the insert size is 1000 base pairs and the predefined minimal insert coverage (*DIC*) is set at 50%, then at least 500 bp of that gene should be present on an insert to be considered as represented by the

array. Genes smaller than the size of the predefined part of the insert, are considered as not represented on the array.

Gene Specific Information (GSI)-equation

Information on differential expression of a gene can only be quantitative and specific for that gene if it originates from a spot that only represents genes from a single transcription unit, assuming that all genes within one transcription unit are equally expressed. This was the requirement that was set for a gene to be considered represented according to the gene specific information (GSI) equation. The criteria for spots that could generate gene specific information are visualized in Fig. 1. One of the variables in the GSI-equation, the minimal overlap (O_{mf}), allows one to set the minimal number of base pairs that are required for identification of a specific gene or transcript on an insert on the clone-based array.

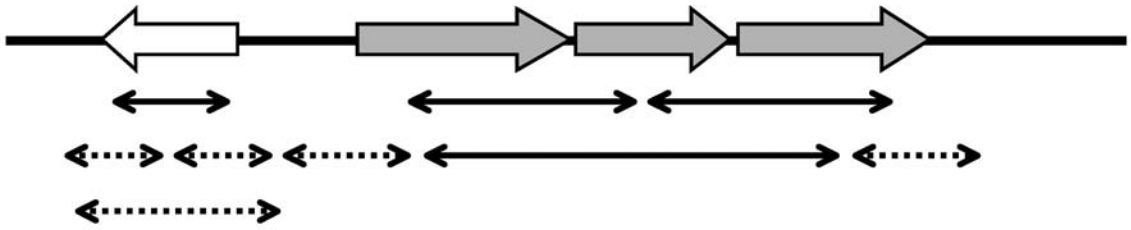

Figure 1. Schematic representation of the criteria that were applied to determine whether gene specific information is generated by a specific insert. The upper line represents a genome fragment in which the block arrows represent genes. Arrows with a gray filling belong to the same transcription unit. The thinner lines represent possible locations of the inserts. The dashed lines represent inserts for which no gene specific information can be generated, since they contain genomic material that possibly belongs to another transcription unit.

Dataset preparation

Fifteen prokaryotes from various genera were selected as model species (Table 1). Genome data from these microorganisms were used for the generation of species-specific values for the expected fraction of represented genes as a function of the genome size (*GS*), number of clones (*N*), insert size (*IS*), and either *DIC* or O_{mf}. Coordinates from all annotated genes from these organisms were obtained from

Table 1. Overview of prokaryotes from several genera with their genes/transcription unit-ratio. Microorganisms that were used for model development (M) or validation (V) of the MIC- and the GSI-equation are depicted in the list.

Genus	***Organism***	Genes/TU (R)	Model (M) or validation (V) strain
Proteobacteria Gammaproteobacteria Enterobacteriales	*Escherichia coli* K-12 MG1655	1.6	
	Escherichia coli O157:H7 EDL933	1.6	
	Escherichia coli CFT073	1.6	M
	Salmonella typhi CT19	1.4	
	Salmonella typhimurium LT2	1.6	
	Yersinia pestis CO92	1.4	
	Shigella flexneri 2a str. 2457T	1.5	
	Buchnera aphidicola Sg	1.5	V
	Wigglesworthia glossinidia	1.5	
Proteobacteria Gammaproteo-bacteria Pasteurellales	*Haemophilus influenzae Rd*	1.7	
	Pasteurella multocida PM70	1.7	V
Proteobacteria Gammaproteo-bacteria Xanthomonadales	*Xylella fastidiosa* 9a5c	1.5	
	Xanthomonas campestris ATCC33913	1.5	V
Proteobacteria Gammaproteo-bacteria Vibrionales	*Vibrio cholerae* El Tor N16961	1.8	M
	Vibrio parahaemolyticus RIMD2210633	1.5	
	Vibrio vulnificus CMCP6	1.5	
Proteobacteria Gammaproteo-bacteria Pseudomonadales	*Pseudomonas aeruginosa* PA01	1.6	M
	Pseudomonas putida KT2440	1.6	
Proteobacteria Gammaproteo-bacteria Legionellales	*Coxiella burnetii* RSA 493	1.6	
Proteobacteria Betaproteobacteria	*Neisseria meningitidis* Z2491	1.6	M
	Ralstonia solanacearum GMI1000	1.6	
Proteobacteria Epsilonproteobacteria	*Helicobacter pylori* 26695	2.3	M
	Campylobacter jejuni NCTC11168	2.7	M
Proteobacteria Alphaproteobacteria	*Rickettsia prowazekii* Nadrid E	1.4	V
	Sinorhizobium meliloti 1021	1.5	
	Agrobacterium tumefaciens C58	1.5	
	Brucella suis 1330	1.5	
	Caulobacter crescentus	1.5	
Firmicutes Bacillales	*Bacillus subtilis* 168	1.6	M
	Oceanobacillus iheyensis HTE831	1.6	
	Stapylococcus aureus MW2	1.6	
	Listeria monocytogenes EGD-e	1.8	M
	Listeria innocua Clip11262	1.8	

Genus	***Organism***	Genes/TU (R)	Model (M) or validation (V) strain
Firmicutes Clostridia	*Clostridium acetobutylicum* ATCC824	1.6	
	Clostridium tetani E88	1.6	
	Thermoanaerobacter tengcongensis MB4T	2.0	
Firmicutes Lactobacillales	*Lactococcus lactis* IL1403	1.5	M
	Streptococcus agalactiae 2603	1.8	
	Streptococcus pneumoniae R6	1.8	
	Lactobacillus plantarum WCFS1	1.6	M
	Enterococcus faecalis V583	1.8	
Firmicutes Mollicutes	*Mycoplasma pneumoniae* M129	2.1	M
	Mycoplasma genitalium G37	3.1	V
	Mycoplasma penetrans HF-2	1.6	
	Ureaplasma urealyticum (serovar 3)	2.1	
Actinobacteria	*Mycobacterium tuberculosis* H37Rv	1.7	M
	Corynebacterium glutamicum ATCC 13032	1.5	
	Streptomyces coelicolor A3(2)	1.4	
	Tropheryma whipplei Twist	1.9	
	Bifidobacterium longum NCC2705	1.3	V
Fusobacteria	*Fusobacterium nucleatum* ATCC25586	2.0	V
Chlamydia	*Chlamydia trachomatis* (serovar D)	1.6	
	Chlamydophila pneumoniae AR39	1.6	
Spirochete	*Borrelia burgdorferi* B31	1.8	
	Treponema pallidum Nichols	1.9	
	Leptospira interrogans 56601	1.5	
Bacteroid	*Bacteroides thetaiotaomicron* VPI-5482	1.8	M
Cyanobacteria	*Thermosynechococcus elongatus* BP-1	1.6	
	Nostoc sp. PCC 7120	1.2	
Green sulfur bacteria	*Chlorobium tepidum* TLS	1.6	
Deinococcus	*Deinococcus radiodurans* R1	1.5	V
Hyperthermophilic bacteria	*Aquifex aeolicus* VF5	2.1	
	Thermotoga maritima MSB8	3.0	V
Archae Euryarchaeota	*Methanococcus jannaschii* DSM2661	1.8	M
	Pyrococcus furiosus DSM3638	2.0	M
	Archaeoglobus fulgidus DSM4304	2.1	
	Thermoplasma acidophilum DSM1728	1.5	
	Methanosarcina acetivorans C2A	1.3	V
	Methanosarcina mazei Goe1	1.3	
	Pyrococcus abyssi	2.1	
Archae Crenarchaeota	*Aeropyrum pernix* K1	2.0	
	Sulfolobus solfotaricus P2	1.6	
	Pyrobaculum aerophilum IM2	1.7	

GenBank, and were used to determine the gene sizes. In addition, information was obtained on the start and stop coordinates from the transcription unit to which the gene belongs, and the position of the gene in this transcription unit. It was assumed that transcription units start at the first base pair of the first gene and finish at the last base pair of the last gene. This information was generated by the combination of intergenic region based transcription unit predictions, generated by Moreno-Hagelsieb and Collado-Vides (8), with gene coordinates from GenBank.

The genome size (*GS*) could be included as a fictitious variable in the datasets, since not the species-specific genome size, but the species-specific gene size distribution and genome organization in genes and transcription units were relevant.

It was assumed that each possible genome fragment of the size of the insert size (*IS*) has an equal chance of being represented. To achieve this, fragments should be generated by physical fragmentation, and not by the use of endonucleases.

Table 2. Overview of the variables that were used for the model datasets on which the MIC- and the GSI-equation are based. Multiple combinations of the mentioned values were applied.

Variable	Values
N	500; 1500; 2500; 3500; 4500; 5500; 6500; 7500; 8500; 9500
IS	100; 300; 500; 700; 900; 1100; 1300; 1500; 2100; 2700; 3000
GS	0.5; 1.5; 2.5; 3.5; 4.5; 5.5; 6.5; 7.5; 8.5; 9.5
O_{mf}	50; 100; 150; 200; 250; 300; 350; 400; 450
DIC	10; 20; 30; 40; 50; 60; 70; 80; 90

Dataset preparation for the fitting procedure for the MIC-equation

For each model organism, the fraction of the represented genes was determined for multiple combinations of the number of clones (*N*), fictitious genome size (*GS*), the insert size (*IS*), and the minimal insert coverage (*DIC*) in the ranges depicted in Table 2. In total, 140 different combinations of values for these variables were tested per strain. This was performed by first calculating the probability value of being represented per gene, and subsequently calculating the average of the probability values from all genes from the organism.

The following formulas were developed for the calculation of the probability value per gene:

$$O_{mv} = \frac{IS \cdot DIC}{100} \quad (1)$$

$$Gene > O_{mv} \Rightarrow p = 1 - \left(1 - \frac{Gene + 1 + IS - 2 \cdot O_{mv}}{GS}\right)^{N} \quad (2)$$

$$Gene < O_{mv} \Rightarrow p = 0 \quad (3)$$

Dataset preparation for the fitting procedure for the GSI-equation

For each organism, the fraction of genes for which specific information could be generated was determined for 114 different combinations of the number of clones (*N*), fictitious genome size (*GS*), the insert size (*IS*), and the minimal required overlap (O_{mf}) in the ranges depicted in Table 2. The represented fraction was determined by taking the average of the probability values per gene. Formulas were developed that describe different situations with respect to the localization and organization of the gene of interest on the insert (eq 4 – 15)

Formulas that were developed to determine the probability value for genes that are transcribed into a single gene transcript:

$$Gene \geq IS \Rightarrow p = 1 - \left(1 - \frac{Gene + 1 - IS}{GS}\right)^{N} \quad (4)$$

$$Gene < IS \Rightarrow p = 0 \quad (5)$$

Formulas that were developed to determine the probability value for genes that are at the beginning of a transcription unit:

$$BP_e \leq IS - O_{mf} \Rightarrow O_e = IS - BP_e \quad (6)$$

$$BP_e > IS - O_{mf} \Rightarrow O_e = O_{mf} \quad (7)$$

$$BP_e + Gene > IS \Rightarrow p = 1 - \left(1 - \frac{Gene + 1 - O_e}{GS}\right)^{N} \quad (8)$$

$$BP_e + Gene < IS \Rightarrow p = 0 \quad (9)$$

Formulas that were developed to determine the probability value for genes that are flanked at both sides by other genes that belong to the same transcription unit:

$$BP_b \le IS - O_{mf} \Rightarrow O_b = IS - BP_b \quad (10)$$

$$BP_b > IS - O_{mf} \Rightarrow O_b = O_{mf} \quad (11)$$

$$BP_e \le IS - O_{mf} \Rightarrow O_e = IS - BP_e \quad (12)$$

$$BP_e > IS - O_{mf} \Rightarrow O_e = O_{mf} \quad (13)$$

$$BP_b + BP_e + Gene > IS \Rightarrow p = 1 - \left(1 - \frac{Gene + 1 + IS - O_b - O_e}{GS}\right)^N \quad (14)$$

$$BP_b + BP_e + Gene > IS \Rightarrow p = 0 \quad (15)$$

Models and fits

The datasets with the expected fractions of represented genes for the various combinations of parameters as presented in the previous section functioned as template for the fitting of the predictive equation for *MIC* and *GSI*.

MIC equation

From equation 2, which was used to determine the probability value per gene, it became apparent that organism-dependent gene size distribution influenced the expected number of represented genes on a clone based array. These organism dependent differences were neglected for the preparation of the MIC equation, which proofed to be justified when validating the MIC-equation (see validation section).

A polynome was developed as MIC model. In the polynome all variables were present in first and second order and in cross terms between two variables. Because of a high expected correlation between *IS* and *DIC* (based on equation 2), this relation was extended with a second order term composed of *IS* and *DIC*, resulting in:

$$\begin{aligned} p_{MIC} = {} & a + b_1 \cdot DIC + b_2 \cdot DIC^2 + c_1 \cdot N + c_2 \cdot N^2 + d_1 \cdot GS + d_2 \cdot GS^2 + e_1 \cdot IS + e_2 \cdot IS^2 + \\ & f \cdot DIC \cdot IS + g \cdot DIC \cdot N + h \cdot DIC \cdot GS + i \cdot IS \cdot N + j \cdot IS \cdot GS + k \cdot GS \cdot N + l \cdot (IS \cdot DIC)^2 \end{aligned} \quad (16)$$

The model datasets for the 15 model species were used together in the regression procedure to estimate the parameters in the MIC model. Linear regression using a standard least squares algorithm (fminsearch) provided by Matlab (The MathWorks) was applied to search the parameters that minimize the sum of squares (SSQ) defined as:

$$SSQ = \sum \left(p_{MIC,\exp} - p_{MIC,\mathrm{mod}} \right)^2 \qquad (17)$$

The resulting parameters are presented in Table 3. The average absolute deviation of the MIC equation from the model dataset was 0.0517.

GSI equation

From the model datasets for the GSI equation it appeared that an organism dependent variable had a strong influence on the calculated number of represented genes (results not shown). Analysis revealed a positive correlation between the number of represented genes and the species-dependent average number of genes per transcription unit, *R*. *R* was determined by dividing the total number of genes (GenBank) by the total number of predicted transcription units (8) (Table 1).

Starting-point for the GSI model was a second order polynome for all variables, extended with the cross terms between two variables. A set of parameters was estimated for each individual model species (results not shown). Parts which appeared to contribute less than 1% to p_{GSI} were not included, which resulted in the following relation:

$$\begin{aligned} p_{GSI} = {} & a + b_2 \cdot O_{mf}{}^2 + c_1 \cdot N + c_2 \cdot N^2 + d_1 \cdot GS + d_2 \cdot GS^2 + e_1 \cdot IS + e_2 \cdot IS^2 + \\ & f \cdot O_{mf} \cdot IS + h \cdot O_{mf} \cdot GS + i \cdot IS \cdot N + j \cdot IS \cdot GS + k \cdot GS \cdot N \end{aligned} \qquad (18)$$

For each prokaryote a set of parameters was obtained by minimizing the SSQ, equivalent to equation 17. The average absolute deviation of the GSI equation from the model datasets was 0.0258.

In order to obtain one generic equation for the organism specific relations for p_{GSI}, the species specific values for the parameters (a – k) in equation 18 were related to the species related variable *R* by a linear relation:

$$parameter(a-k) = q + r \cdot R \qquad (19)$$

in which *R* is species specific (Table 1). Since no dependency of *a* with *R* could be established, *a* was set at the average of all individual *a* values: 0.544. With this value the polynome was fitted again, and the final relations between the other parameters and *R* were determined (Table 3).

Table 3. Values for the parameters in the MIC- and the GSI-equation.

Parameter	MIC equation	GSI equation	GSI equation
		q	*r*
a	4.85E-01	0.544	0
b_1	2.54E-03	*	*
b_2	-1.51E-05	-4.26E-08	-3.05E-07
c_1	1.27E-04	6.13E-05	1.46E-05
c_2	-5.22E-09	0	-1.96E-09
d_1	-1.22E-01	-7.84E-02	-1.06E-02
d_2	3.42E-03	3.31E-03	3.23E-04
e_1	3.95E-04	-5.36E-04	2.08E-04
e_2	-9.57E-08	9.73E-08	-4.62E-08
f	-9.85E-06	1.69E-08	3.42E-08
g	-4.61E-07	*	*
h	3.25E-04	2.55E-06	4.12E-06
I	-1.69E-08	-2.22E-08	5.47E-09
j	2.01E-05	2.42E-05	-6.04E-06
k	2.26E-06	-1.76E-06	1.30E-06
l	2.60E-11	*	*

ad *: this parameter is not present in the GSI equation

Prediction of the optimum value for the insert size (*IS*)

Whereas an increase in *N* will always have a positive contribution to the fraction of represented genes, and an increase in *GS*, O_{mf}, and *MIC* a negative contribution, there may be an optimum *IS* that depends on the values of the other variables. This optimum can be estimated by differentiation of equation 16 and 18 to *IS* (d*p*/d*IS*).

For the determination of the optimal value for *IS* for the MIC-approach this results in the following equation:

$$\frac{dp_{MIC}}{dIS} = e_1 + 2e_2 \cdot IS + f \cdot DIC + i \cdot N + j \cdot GS + 2 \cdot l \cdot DIC^2 \cdot IS = 0 \Rightarrow$$

$$IS_{MIC-opt} = \frac{-e_1 - f \cdot DIC - i \cdot N - j \cdot GS}{2e_2 + 2l(DIC)^2} \quad (20)$$

For the determination of the optimal value for *IS* for the GSI-approach the equation is as follows:

$$\frac{dp_{GSI}}{dIS} = e_1 + 2e_2 \cdot IS + f \cdot O + i \cdot N + j \cdot GS = 0 \Rightarrow$$

$$IS_{GSI-opt} = \frac{-e_1 - f \cdot O - i \cdot N - j \cdot GS}{2e_2} \quad (21)$$

If the indicated values for IS_{opt} are outside the range of 0 to 2000 bp (the range that was applied for validation of the models) no optimum can be identified within the boundaries of the model. In these cases small values of *IS* will give the best results.

Validations

In order to validate the MIC- and the GSI-equation, datasets were generated (as previously described in the "dataset preparation" section) for ten validation species (Table 1). Represented gene fractions were calculated per species for all possible combinations for the variables as presented in Table 4 and distracted from the values as they were predicted by MIC- and the GSI-equations 16 and 18, respectively. The distributions of the residuals, i.e. the difference between predicted and the calculated fraction, for both equations are presented as histograms in Figures 2a and 2b. The residual distributions of both the MIC- and the GSI-equation approach the normal distribution with a slight tendency to underestimate the fraction of represented genes (Fig. 2a and 2b). Moreover, in Table 5 the reliability of the formulas is depicted as the fraction of predictions that differ less than 0.01, 0.05 and 0.10 from the real values. It should be noted that the indicated reliabilities relate to the range of variables as depicted in Table 4.

Table 4. Overview of the variables and the values used for these variables that were used for the datasets that were used for the validation of the MIC- and the GSI-equation. All possible combinations of the mentioned values were tested.

Variable	Values
N	1000 ; 4000 ; 7000 ; 10000
IS	100; 500; 1000; 1500; 2000
GS	1; 3; 5; 7
O_{mf}	25; 50; 75
DIC	100; 200

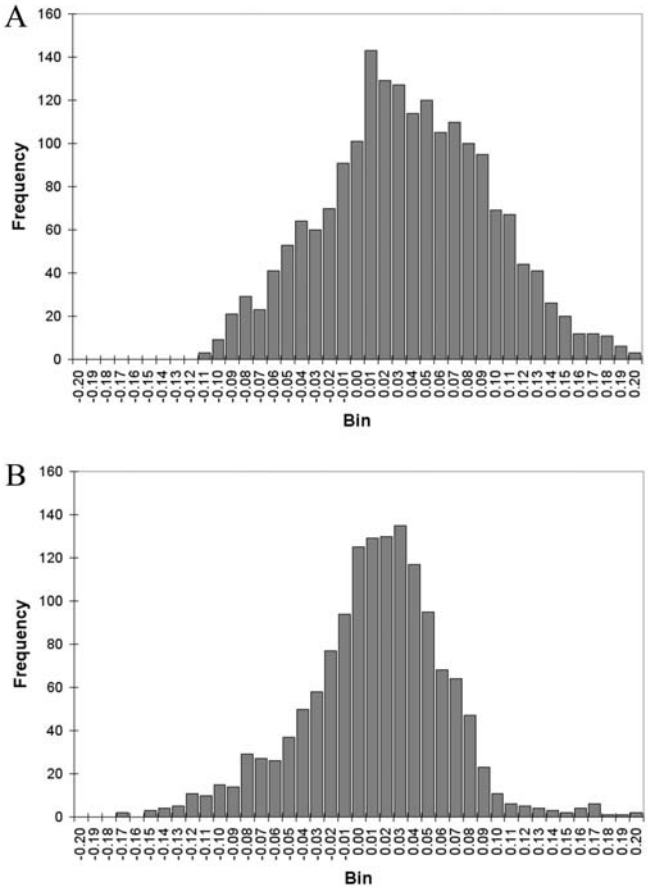

Figure 2. Histogram representations of the residuals from the validation of the MIC-equation (A) and the GSI-equation (B).

Deviations between the predicted fractions by the MIC-equation and the true values as they were determined for the validation species are mainly to be attributed to species-specific gene size distribution. In order to obtain one generic equation, and based on the accuracy of the equation in its current form (Table 5), it was decided not to include a species-specific variable.

Table 5. Reliability of the MIC- and the GSI-equation, depicted as the fraction of predictions that differ less than 0.01, 0.05 or 0.10 from the real values, for the validation sets defined in Table 4.

Abs (Δ predicted vs. real)	Fraction for MIC-equation	Fraction for GSI-equation
< 0.01	0.19	0.24
< 0.05	0.58	0.73
< 0.10	0.87	0.95

Influence of the average number of genes per transcription unit (*R*) on the predicted values

From the input variables for the MIC and GSI formulas, *N*, *IS*, *DIC* and O_{mf} are user-defined, while *GS* and *R* have to be estimated for the specific organism. Whereas current techniques allow for rapid and accurate estimations of *GS* (4, 10, 12), the organism specific value for *R* is difficult to determine for species from which little sequence information is available.

R was determined for 73 prokaryotes from multiple genera, as previously described in the "models and fits" section (Table 1). For 61 of the 73 strains in this list, *R* was within the narrow range from 1.5 – 2.0. Moreover these data indicate that accurate estimations of *R* can be made, based on the genus of the organism, with an exception for the mollicutes, the hyperthermophilic bacteria and the euryarchaeota.

The effect of false estimations of *R* was studied by the generation of validation sets as defined in Table 4 with the exception that higher or lower values for *R* were applied. The resulting values from the GSI-equation were compared with the true values (Table 6). It appeared that an over- or underestimation of 0.2 on *R* had limited effects on the fraction of predictions that differ less than 0.1 from the real values from the validation dataset (0.90 vs. 0.95 for the exact value of *R*). While an overestimation of 0.3 still results in 88% of the predictions that differ less than 0.1 from the real value from the validation dataset. This percentage was 80% in case of an underestimation of the same size.

Table 6. Effect of false estimations of *R* on the fraction of predictions that differ less than 0.01, 0.05 or 0.10 from the real values, for the validation set defined in Table 4.

Applied value for R	Abs(Δ predicted vs. real)	Fraction
R	< 0.01	0.24
	< 0.05	0.73
	< 0.10	0.95
R – 0.1	< 0.01	0.17
	< 0.05	0.65
	< 0.10	0.95
R + 0.1	< 0.01	0.24
	< 0.05	0.75
	< 0.10	0.94
R – 0.2	< 0.01	0.12
	< 0.05	0.55
	< 0.10	0.90
R + 0.2	< 0.01	0.23
	< 0.05	0.69
	< 0.10	0.91
R – 0.3	< 0.01	0.10
	< 0.05	0.38
	< 0.10	0.81
R + 0.3	< 0.01	0.21
	< 0.05	0.59
	< 0.10	0.88

Application

As an example for the applicability of the developed equations, the effect of different combinations of the number of clones and insert size was determined for a prokaryote with a genome size of 4 Mbp and an estimated value for *R* of 1.8 using equations 16 and 18. The effect of multiple combination of *N* and *IS* on p_{MIC} was determined for minimal insert coverage (DIC) values of 25%, 50% and 75%. The results are depicted in the contour plots in figure 3a – 3c. The predicted fractions of represented genes for which gene specific information could be generated (p_{GSI}) with a minimal overlap between the insert and the gene of 100 bp is depicted in figure 4.

Plots like those presented in figures 3a-c and 4 can be used to determine the preferred combination of the number of spots on the array and the insert size. If for instance the number

of spots would be limited to 6000, an insert size of approximately 800 bp would be optimal with respect to the fraction of genes that are represented with a minimal insert coverage of 25% (Fig. 3a). From equation 20 this optimum appears to be 803 bp. With this combination of array parameters the predicted fraction of genes that cover at least 25% of the insert (which equals 803 x 0.25 = 201 bp) is 0.75 (eq. 16). Meanwhile the predicted fraction of genes for which gene specific information can be generated is 0.49 (eq. 18). If the specificity of the data is considered to be more important than the amount of represented genes, it is preferable to have an optimum value for p_{MIC} for higher values of *DIC* (e.g. Fig. 3c) and a high value for p_{GSI} (Fig. 4). These requirements are best fulfilled by combinations with low values for the insert size.

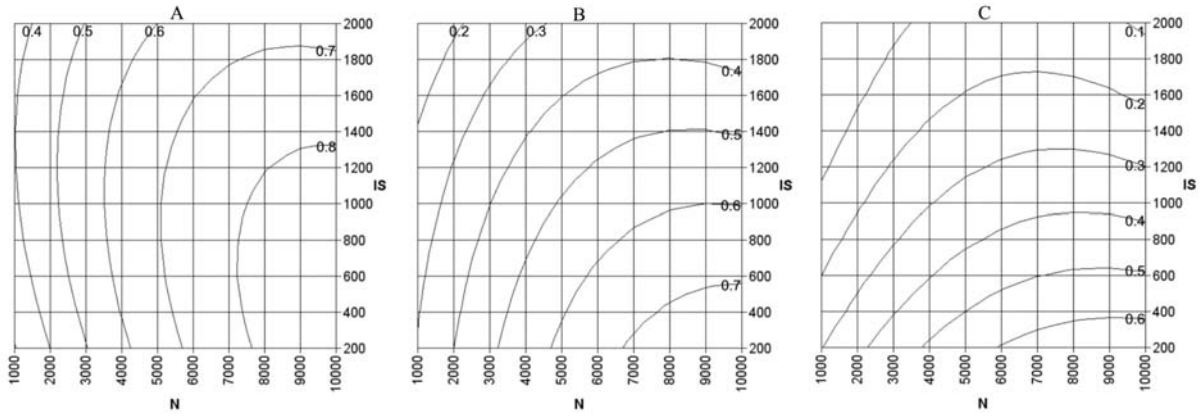

Figure 3. Contour plots of the predicted fractions of represented genes with a minimal insert coverage of 25% (A), 50% (B), or 75% (C) as a function of the number of clones (*N*) and the insert size (*IS*) for a prokaryote with a genome size of 4 Mbp. The predicted fractions are depicted in the plot on top of the lines by which they are represented.

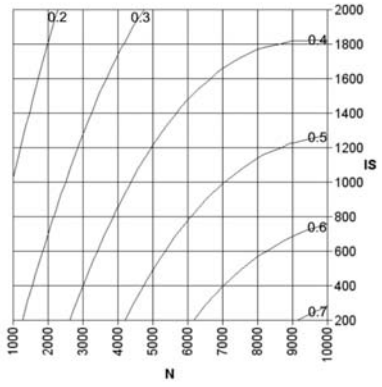

Figure 4. Contour plot of the predicted fraction of represented genes for which gene specific information could be generated as a function of the number of clones (*N*) and the insert size (*IS*) for a prokaryote with a genome size of 4 Mbp, an average number of genes per transcription unit (*R)* of 1.8, and a minimal overlap between the insert and the gene of 100 bp. The predicted fractions are depicted in the plot on top of the lines by which they are represented.

Discussion

Classical approaches for the construction of DNA libraries form a suitable base for the construction of clone-based microarrays. However, as the construction of these libraries is a random process, it is beforehand uncertain whether a gene or transcription unit will be uniquely represented on a separate insert on the array. Genome coverage by a DNA library is usually determined by calculating the expectation that each single nucleotide from that gene is present (5, 7). These formulas will overestimate the number of clones required when the library is to be used for the construction of a microarray, since for this purpose partial representation of a gene is sufficient for hybridization.

To our knowledge, Akopyants *et al.* were the first to estimate genome coverage at the gene level (1). They predicted the fraction of represented genes using equation 22:

$$p_{Akopyants} = 1 - \left(1 - \left(\frac{\text{average transcript size} + \text{insert size - 2 x required overlap}}{\text{genome size}}\right)\right)^{\text{number of clones}} \quad (22)$$

An important variable in this formula is the average transcript size. However, use of this variable is not legitimate for this type of probability calculations since the average probability per gene (the required information) is not *per se* equal to the probability per average gene. When we validated the Akopyants formula on the same dataset that was applied for the validation of the MIC-equation, it appeared that 49% of the predictions deviated more than 0.1 from the real value (calculated as the average chance per gene), with a strong tendency to overestimation. The Akopyants formula therefore appears unreliable for calculating optimal library sizes

None of the previous formulas give insight in the fraction of genes for which gene specific information can be generated, while this is one of the most important features when one is interested in studying differential gene expression. The MIC-and GSI-equations that were developed in this study allow for good estimations of both the genome coverage at the gene level, and the fraction of genes for which gene specific transcription information can be generated.

Whereas the MIC-equation is rather straight-forward with respect to the input variables and interpretation, application of the GSI-equation requires the estimation of the average number of genes per transcription unit for an organism. Although a false estimation of this variable could lead to a wrong prediction of the represented fraction, Tables 1 and 6 indicate that this risk is limited.

The GSI-equation is partially based on operon predictions. For the development of the model and validation datasets we used log-likelihood based transcription unit predictions for adjacent pair of genes to be in the same operon (9). This log-likelihood based prediction method is only applicable to organisms for which at least large parts of the genome have been sequenced, and will therefore not be useful when sequence data from array spots for which differential expression was identified, have to be interpreted. Nevertheless, good predictions can be made on whether or not genes that are co-represented in a single spot on the array belong to the same transcription unit. Strong indications can already be obtained from the physical organization of the DNA fragment of interest, like gene orientation and intergenic distance (8, 11). Other indications are the co-occurrence of genes with a joint function, and the conserved organization of homologous genes in other prokaryotes (6, 11).

Conclusions

The MIC- and GSI-equations that were developed in this study were based on genomes from 15 prokaryotes from different genera, and validated on the genomes of 10 other prokaryotes. These validations show that these equations form reliable tools for optimal design of prokaryotic clone-based microarrays within the ranges that were tested (Table 4), and that they are applicable to a broad range of prokaryotes. Therefore, these equations form a good basis for the design of microarrays for prokaryotes from which the genome sequence is not available.

Acknowledgements

We would like to thank Rolf Boesten, Martien Caspers, Nicole van Luijk and Karin Overkamp for their critical remarks and useful suggestions.

Additional file

The MIC- and GSI-equations (eq. 16 and 18), and the derived equations for prediction of the optimal values for *IS* (eq. 20 and 21), are available as a Microsoft Excel fill in spreadsheet. This spreadsheet can also be applied for the generation of contour plots in which the represented gene fractions are depicted as a function of the number of clones and

the insert size. The file is available online at BioMed Central (http://www.biomedcentral.com/1471-2105/6/238/suppl/S1)

References

1. **Akopyants, N. S., S. W. Clifton, J. Martin, D. Pape, T. Wylie, L. Li, J. C. Kissinger, D. S. Roos, and S. M. Beverley.** 2001. A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. Mol. Biochem. Parasitol. **113:**337-340.
2. **Askenazi, M., E. M. Driggers, D. A. Holtzman, T. C. Norman, S. Iverson, D. P. Zimmer, M. E. Boers, P. R. Blomquist, E. J. Martinez, A. W. Monreal, T. P. Feibelman, M. E. Mayorga, M. E. Maxon, K. Sykes, J. V. Tobin, E. Cordero, S. R. Salama, J. Trueheart, J. C. Royer, and K. T. Madden.** 2003. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. Nature Biotechnol. **21:**150-156.
3. **Cho, J. C. , and J. M. Tiedje.** 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. Appl. Environ. Microbiol. **67:**3677-3682.
4. **Chu, G., D. Vollrath, and R. W. Davis.** 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science **234:**1582-1585.
5. **Clark, L., and J. Carbon.** 1976. A colony bank containing synthetic *Col*E1 hybrids representative of the entire *E. coli* genome. *Cell* **9:**91-99.
6. **Ermolaeva, M. D., O. White, and S. L. Salzberg.** 2001. Prediction of operons in microbial genomes. Nucl. Acids Res. **29:**1216-1221.
7. **Lander, E. S., and M. S. Waterman.** 1998. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics **2:**231-239.
8. **Moreno-Hagelsieb, G., and J. Collado-Vides.** 2002. A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics **18 Suppl. 1:**329-336.
9. **Salgado, H., G. Moreno-Hagelsieb, T. F. Smith, and J. Collado-Vides.** 2000. Operons in *Escherichi coli*: Genomic analyses and predictions**.** Proc. Nat. Acad. Sci. **97:**6652-6657.
10. **Sun, L. V., J. M. Foster, G. Tzertzinis, M. Ono, C. Bandi, B. E. Slatko, and S. L. O'Neill.** 2001 Determination of *Wolbachia* genome size by pulsed-field gel electrophoresis. J. Bacteriol. **183:**2219-2225.
11. **Westover, B. P., J. D. Buhler, J. L. Sonnenburg, and J. I. Gordon.** 2005. Operon prediction without a training set. Bioinformatics **21:**880-888.
12. **Wilhelm, J., A. Pingoud, and M. Hahn.** 2003. **Real-time PCR-based method for the estimation of genome sizes.** Nucl. Acids Res. **31:**(10):e56.

Chapter 4

Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling

Bart Pieterse, Rob J. Leer, Frank H.J. Schuren, and Mariët J. van der Werf.

This chapter has been published as:
Pieterse, B., R. J. Leer, F. H. J. Schuren, and M. J. Van der Werf. 2005. Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling. Microbiol. **151:**3881-3894.

Abstract

The organic acid lactate is the predominant fermentation product of *Lactobacillus plantarum*. The undissociated form of this organic acid is a strong growth inhibitor for the organism. Different theories have been postulated to explain the inhibitory effects of lactic acid, (i) the toxicity arising from the dissipation the membrane potential, (ii) acidification of the cytosol or (iii) intracellular anion accumulation. In general, organic acid stresses are complex to study since their toxicity is highly dependent on their degree of dissociation and thus the pH. In this study transcription profiles of *L. plantarum* grown in steady state cultures that varied in lactate / lactic acid concentration, pH, osmolarity and absolute and relative growth rate, were compared by microarray analysis. By doing so, differential expression of multiple groups of genes could specifically be attributed to the different aspects of lactic acid stress. A highly coherent group of lactic acid responsive, cell surface protein encoding genes was identified to which no function has previously been assigned. Moreover, a group of genes that showed increased expression in response to the combination of lactic acid and a lower growth rate is expected to be involved in the formation of the alternative fermentation end-products malate, acetate and ethanol. One of these pathways is the phosphoketolase by-pass that is typical for bifidobacteria.

Introduction

Lactobacilli are commonly used in fermentation processes for their contribution to the conservation of food and feed products. This feature is attributed to the predominant metabolite biosynthesized during these fermentations, lactate, which is formed as the end product of glycolysis when pyruvate is reduced to lactate with NADH as a cofactor and secreted into the culture medium. The resulting increase in extracellular lactic acid coincides with a decrease of the pH. Lactic acid, as well as other organic acids, is known for its inhibitory effect on bacterial growth (42). Although *L. plantarum* is more resistant towards lactic acid than many other micro organisms, its growth is still strongly inhibited by elevated concentrations of this organic acid reached during fermentation (18, 40). The inhibitory effect of organic acids is mainly caused by the undissociated form of the molecule, which diffuses over the cell membrane towards the more alkaline cytosol (3, 22). This process continuous until an equilibrium has been reached at which the intra- and extracellular concentration of the undissociated, membrane permeable form of the lactic acid molecule are equal. Inside the cell the lactic acid dissociates in accordance with the Henderson-Hasselbach equation. This implies that the inhibitory effect is larger at a larger ΔpH. It also explains why microbes that can lower their intracellular pH values are more resistant towards organic acids (7, 10), as has also been described for *L. plantarum* (31).

Several theories have been postulated to explain the toxic effect of organic acids in more detail. One of these considers organic acids as uncouplers that transport protons towards the inside of the cell, which is a ΔpH driven process. Eventually this influx could lead to a complete dissipation of the proton motive force (3, 24). A second aspect relates to the negative effects of the lower intracellular pH, caused by the influx of lactic acid. However, whereas many organisms aim at maintaining a constant intracellular pH (34, 42), most anaerobic fermenting species avoid a large ΔpH by allowing for a lower intracellular pH, and as a result increase their tolerance against organic acids (6, 7, 10). This is also the case for *L. plantarum* that can grow at intracellular pH values as low as 4.6-4.8 (31). A third factor explaining the inhibitory effect of organic acids, is the intracellular accumulation of anions, which could lead to both end-product inhibition and a loss of water-activity (39, 40). For lactic acid bacteria, end product inhibition by lactic acid could result in a disturbance of the regeneration of the cofactor NAD^+, especially under anaerobic conditions when the cell do not have the possibility of NAD^+ regeneration by NADH oxidase.

The recent sequencing of the genome of *L. plantarum* WCFS1 (26) revealed several genes for which homologues in other prokaryotes were found to play a role in pH control, or in the maintenance of the proton motive force, a proton translocating F_1F_0-ATPase (8); several sodium-proton antiporters (9); amino acid decarboxylases that use an intracellular hydrogen ion for the decarboxylation of an imported amino acid (4, 8, 46); and the genes involved in the malolactic fermentation, during which extracellular malate is imported and decarboxylated. The lactate produced is secreted via an electrogenic uniport which, results in a membrane potential that can be applied for the generation of ATP (33).

The genomic sequence of *L. plantarum* gives insight into potential responses of the cells to counteract the anticipated negative aspects of extracellular lactic acid accumulation. Nevertheless the genomic data give little clue on the exact effects of lactic acid and the response of the cells towards this stress. Functional genomics techniques offer good opportunities for gaining a better understanding of the physiological state and the specific response of the cell. In recent years the acid stress response of several prokaryotes has been studied with both proteomics and transcriptomics approaches (1, 16, 17, 28, 29, 32, 43, 48, 49). A few reports describe the use of these approaches to study the organic acid stress response caused by lactate (19), acetate (e.g. 2, 25, 36), propionate (36), and formate (25). None of these studies, however, allows for the discrimination between the effects of the undissociated and the dissociated form of lactic acid. Moreover, the negative effect that the organic acid could have on the growth rate of the organism, and the thereby resulting secondary effects on gene expression are usually neglected.

The aim of this study was to obtain a better insight into the effect of lactic acid on *L. plantarum*, and to obtain a genome wide view of the transcriptional response towards this stress factor. To this end an experimental design was developed that allows for the discrimination between the effects of pH, lactic acid in its dissociated and undissociated form, and the secondary effect of a diminished maximum growth rate.

Materials and methods

Microorganism

The microorganism used for these experiments is *Lactobacillus plantarum* WCFS1, a single colony isolate from *L. plantarum* NCIMB8826, which is maintained at NIZO food research in Ede, the Netherlands. The complete genome sequence from this organism is available (26; GenBank, AL935263).

Cultivation and cell harvesting

Glucose-limited continuous cultivation of *L. plantarum* WCFS1 was performed in duplicate on 25% MRS medium (13.75 g/l, Difco) with additional sodium chloride or sodium lactate (as specified in Table 1), at 37°C in an Applikon bioreactor with a working volume of 1 liter. Glucose limitation was confirmed by the fact that the addition of a concentrated glucose solution to the steady states resulted in increased cell densities. A constant pH was maintained (pH 6.0 or pH 4.8) by the automatic titration with 10M sodium hydroxide. The culture was maintained anoxic by a continuous overlay with nitrogen gas. The dilution rate was set at 0.3 h^{-1} or 0.05 h^{-1}.

Samples were taken from the steady state cultures and immediately quenched in a -45°C, methanol based buffer as previously described (35). Cell pellets were stored at -80°C until used.

RNA isolation

RNA was isolated from the cells according to the Macaloid / phenol based protocol as previously described (35). RNA purity and concentrations were determined both spectrophotometrically and on agarose gel. The RNA isolates were checked for residual RNase activity by comparing samples that were incubated for 1 hour at 42°C with the initial material on an agarose gel.

Array design

The microarray used is a clone-based array, containing 3714 identified unique fragments from the genomic *L.plantarum* WCFS1 library in pBlueScript SK+ (26). The mean size of the fragments is 1.2 kb with an average deviation of 0.3 kb. The microarray contains approximately 80% of the *L. plantarum* WCFS1 genome (35).

Fluorescent labeling and hybridization

Differential transcript levels were determined by two-color fluorescent hybridizations of the corresponding cDNA's on the *L. plantarum* clone-array. The RNA samples were labeled by *in vitro* reversed transcription with either Cy5- or Cy3 labeled dUTP, using random hexamer primers. Labeling, hybridization and washing were performed as previously described (35). The fluorescent labels were swapped for the labeling of the biological duplicates in order to avoid false positives due to dye specific effects.

Image analysis

The fluorescent signals from the two different labels on the hybridized arrays were quantified with a ScanArray Express scanner (Packard Bioscience) and Imagene 4.2 software (BioDiscovery, Inc.). Spots for which the difference between the mean signal of the spot and the mean signal of the background was smaller than two times the background standard deviation were excluded from further analysis. Spots for which the signal in one or both of the channels exceeded the detection limit of the scanner were also excluded. After removal of the empty spots and the spots for which the signal exceeded the detection limit of the scanner, 2787 spots (75%) remained for further analysis.

Normalization

Within-slide, intensity dependent normalizations were performed with the scatter plot smoother LOWESS from the software program Datafit (Oakdale Engineering). The user-defined fraction of data used for smoothing at each point was set at 20% for all slides.

Significance analysis

Prior to significance analysis, a data transformation was applied on the normalized ratios in order to obtain distribution around zero that approaches the normal distribution (35). Significance analysis was performed by means of 1way-ANOVA. Subsequently, a Tukey HSD test was performed to determine whether a significant differential expression level (99% confidence interval) was observed under a specific condition. If genes or predicted operons were present on multiple spots on the array, these spots were considered as replicates in the significance analysis.

By comparison of overlapping regions of the genomic fragments on the array, we were able to predict which of the specific genes present in the clones is or are the ones affected.

Hierarchical clustering

For the hierarchical clustering and visualization of the results, the programs CLUSTER and TREEVIEW were applied (12). Only those genes or operons were included that fall within the 99% confidence interval of the Tukey HSD test for at least one of the 6 conditions. Average linkage clustering was performed on the natural logarithm of the expression ratios of the genes or operons, based on all replicate spots. The outcome of the clustering was influenced by accounting weight values to the different experiments. On the experiments in which the effect of lactate was studied, a weight value of 1 was applied. The dataset in which the effect of the lower pH value was studied was assigned a weight value of 0.75, the 800mM sodium chloride dataset with 0.5 and the one of 300 mM sodium chloride with 0.25.

***In silico* analysis of cell surface proteins**

The intracellular, extracellular and transmembrane regions of proteins that were annotated as cell surface proteins were predicted using a Hidden-Markov-Model based tool (http, //www.cbs.dtu.dk/services/TMHMM; 27). Multiple protein alignments (gap open penalty of 10; gap extension penalty of 1; GONNET-matrix), and pI predictions were performed in DNAman version 5.2.9. (Lynnon BioSoft). The motif "LPQTxE" is the *L. plantarum* orthologue for the well known cell wall anchoring LPxTG pentapeptide (15, 26).

Scanning electron microscopy

Nucleopore Polycarbonate membranes (Costar Cambridge MA) with 1µm holes were coated with Poly-L-Lysine (30 minutes in 0.01% poly-L- lysine, 0.1 M Tris-HCl buffer). For scanning electron microscopy, several drops of bacteria from duplicate steady state cultures were transferred to the membrane immediately upon harvesting, and stored for 5 minutes at 100% humidity. Fixation was achieved by 30 minutes incubation in the bacterial growth medium with additional 3 % glutaraldehyde. The membranes were subsequently rinsed 3 times in water for 10 minutes. Dehydration was performed by immersing in a series of ethanol washes (30%, 50%, 70%, 90% , 3 x 100%). Membranes were critical point dried in CO_2 (Balzers CPD 020, Balzers Union). The dry membranes were mounted on the sample holders by carbon adhesive tabs (Electron Microscopy Sciences). Sample holders were put inside the preparation chamber (CT 1500 HF, Oxford Instruments) attached to the microscope. Samples were sputter coated with 5 nm platinum and analyzed in a field emission scanning electron microscope (JEOL 6300F) at 5 kV. Images were recorded digitally, and photo processing was done with Adobe PhotoShop 5.5.

Results and discussion

Experimental design

The discrimination between the effects of the dissociated and undissociated form of organic acids on bacterial physiology has been the subject of many studies (e.g. 24, 41). However, so far these studies have not been extended to the gene expression level. In this study, an experimental design was applied that enabled discrimination between the effects of lactic acid in its dissociated and undissociated form, pH, water activity, and absolute and relative growth rate (Table 1; Fig. 1). Cells were grown in steady state cultures, which allows for a reliable view on the long-term effects of the organic acid.

Table 1. Culture conditions of the fermentation with *L. plantarum* for which transcription profiles were compared by microarray analysis.

Code	Effect	Stress condition					Reference condition				
		pH	Addition	μ_{max}	d	rel. μ	pH	Addition	μ_{max}	d	rel. μ
A	mild osmotic stress	6.0	300mM NaCl	0.6	0.3	0.5	6.0	no addition	0.6	0.3	0.5
B	osmotic stress	6.0	800mM NaCl	0.4	0.3	0.8	6.0	no addition	0.6	0.3	0.5
C	low pH	4.8	300mM NaCl	0.5	0.3	0.6	6.0	300mM NaCl	0.6	0.3	0.5
D	lactate	6.0	300mM Na-lactate	0.6	0.3	0.5	6.0	300mM NaCl	0.6	0.3	0.5
E	lactic acid/ higher rel. μ	4.8	300mM Na-lactate	0.06	0.05	0.8	4.8	300mM NaCl	0.5	0.05	0.1
F	lactic acid / lower abs. μ	4.8	300mM Na-lactate	0.06	0.05	0.8	4.8	300mM NaCl	0.5	0.3	0.6

In order to distinguish between the effects of the undissociated (uncharged) and dissociated (charged) form of the molecule, fermentations with 300mM of additional sodium lactate (the lactate concentration resulting from the fermented glucose was approximately 50 mM) were performed at two pH values, pH 6.0 and pH 4.8. In this way, the amount of undissociated lactic acid could be varied (2.1 mM versus 30.7 mM of additional undissociated lactic acid), while maintaining an equal amount of initial total lactate. The effect of the additional sodium lactate at pH 6.0 will be referred to as the lactate effect, while that at pH 4.8 will be referred to as the lactic acid effect. The dilution rate of the cultures with a pH of 6.0 was 0.3 h^{-1}, and that of the cultures at pH 4.8 0.05 h^{-1}. The transcription profiles from these cultures were compared with those of cultures to which an equal concentration of sodium chloride, which causes an equal reduction of the water activity, was added (20). All other parameters were equal (Table 1; row D and E).

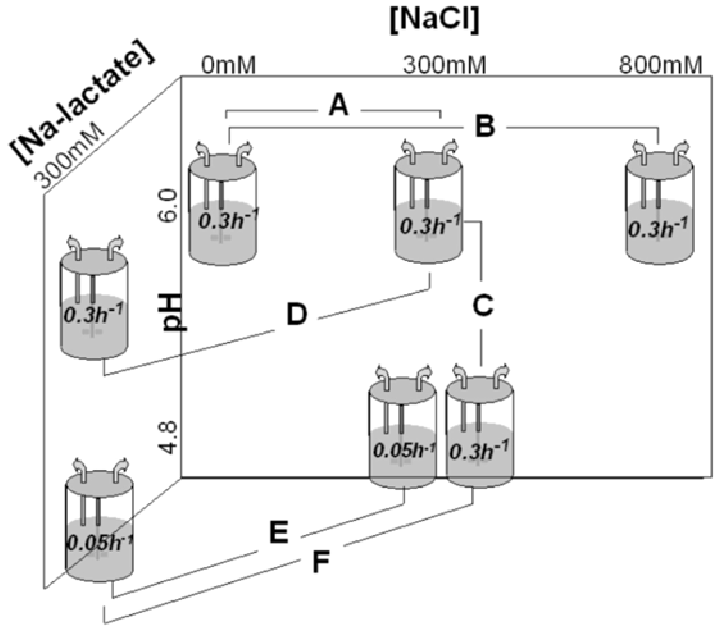

Figure 1. Schematic representation of the experimental design that was applied for discriminating between the effects of lactate, lactic acid, pH and osmotic stress on gene expression of *L. plantarum* WCFS1. Each bioreactor in the scheme represents a steady state fermentation in duplicate. The connectors between the bioreactors represent the hybridizations as they were performed. The letter codes correspond with those in Table 1

The effect of extracellular pH on gene expression was studied by comparing mRNA levels from *L. plantarum* cultured at pH 4.8 with those of *L. plantarum* cultured at pH 6.0 (Table 1; row C).

As the presence of lactic acid also influences the growth rate of the organism, changes in gene expression are expected that are not caused by the primary effect of the sodium lactate, but by the secondary effect of an altered absolute or relative growth rate. In this study, the effect of a different absolute growth rate was avoided by using identical dilution rates for the cultures for which the transcription profiles were compared. In order to determine whether relative growth rate (i.e. the ratio between the set growth rate and the maximum growth rate under the given conditions in a batch culture) also affected gene expression, the transcription profile of the culture with sodium lactate at pH 4.8 and a dilution rate of 0.05 h^{-1} was also compared with that of the culture with 300 mM sodium chloride at pH 4.8 and a dilution rate of 0.3 h^{-1}. The relative growth rates of these cultures were alike (Table1; row F).

Furthermore, transcriptome data sets from two independent duplicate cultures in which the effect of additional sodium chloride (0 mM vs. 300 mM / 0 mM vs. 800 mM) was

studied, were generated (Table 1; row A and B). These were added to study whether osmotic stress mechanisms are initiated due to lactic acid.

Hierarchical cluster analysis

Hierarchical clustering was performed in order to group genes or operons that showed similar expression patterns under the different experimental conditions (Figure 2). The hierarchical cluster clearly showed that the number of genes that were affected by the addition of 300 mM sodium chloride, a lower pH and lactate at a pH of 6, is rather limited.

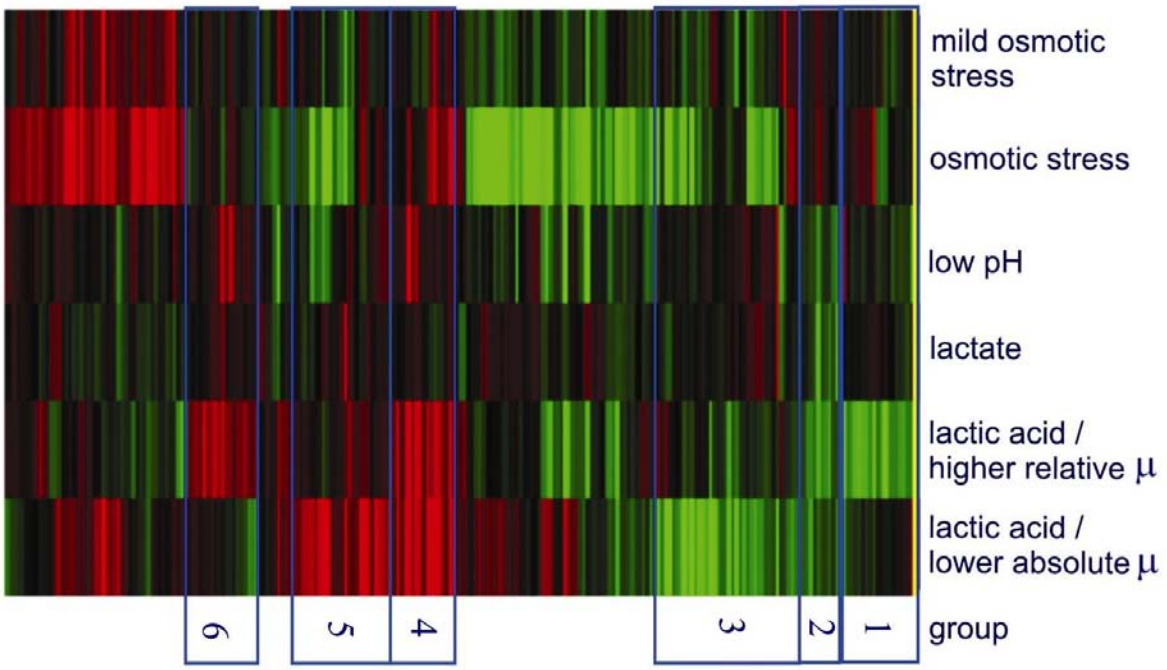

Figure 2. Hierarchical cluster of the normalized, log-transformed data of the average gene expression for all genes for which a significant differential expression (99% confidence interval of the Tukey HSD test) was observed for at least one of the conditions. Red indicates increased expression, green decreased expression, and black unchanged expression. The groups of genes that are marked in the numbered boxes, are listed in Tables 2 - 7.

The limited number of genes that were affected by a lower pH alone is not surprising in view of the limited effect that this pH difference has on growth rate and biomass yield (18). However, according to McDonald *et al.* (31) such a decrease in extracellular pH could coincide with a decrease of the intracellular pH of almost 1 pH unit. Apparently, this decrease in intracellular pH is not itself a strong trigger for gene regulation. The combined facts that a relatively low number of genes was differentially expressed in response to lower pH alone and to additional sodium lactate at a pH of 6.0, support the theory that the effect of lactic acid is mainly caused by the undissociated form of the molecule.

Striking differences occur in the two datasets that both represent the effect of lactic acid, but vary in growth rate / relative growth rate. In order to determine lactic acid specific effects, we concentrated on the genes that are marked with boxes 2 and 4 in Fig. 2. The corresponding genes and their expression data are listed Table 2 to 7.

Table 2. Genes / operons from group 4 of the hierarchical cluster (Figure 2). This group is characterized by the increased expression of genes / operons in response to lactic acid. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nos.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. μ	Lactic acid / lower abs. μ
2109	excinuclease, subunit C	1.0	**1.6**	1.3	1.0	**1.4**	**1.4**
0027-0028	β-phosphoglucomutase; glycosylhydrolase	1.0	**3.1**	1.0	1.3	2.1	2.7
3578	catalase	1.9	**3.6**	1.7	1.5	3.1	3.0
0841-0843	exopolyphosphatase; polyphosphate kinase; exopolyphosphatase	1.4	**1.8**	1.2	1.1	**1.6**	**1.5**
1396	putative transcriptional regulator	1.0	**1.9**	0.9	0.8	1.8	2.1
0313	NADH dehydrogenase	1.4	1.6	1.3	0.8	**2.6**	**2.7**
3128	Dpr-like protein	2.5	**3.1**	1.1	0.7	2.7	**3.5**
3127	cell surface protein	1.7	2.2	1.3	0.8	2.0	**3.2**
0965	transcriptional regulator	0.7	**1.9**	1.3	0.9	**2.5**	**2.9**
2173-2174	extracellular protein; cell surface protein	0.7	0.9	1.2	1.1	1.6	**1.6**
3262-3263	squalene synthase; phytoene synthase	1.0	1.2	1.2	1.3	**2.9**	**2.5**
2975-2977	integral membrane protein ; cell surface protein ; cell surface protein	0.7	1.2	2.1	1.3	**2.6**	**2.7**
1269	Clp protease, ATP-binding subunit clpE	0.8	1.0	1.4	1.1	**1.7**	**1.8**
3678-3679	cell surface protein; cell surface protein	1.0	2.0	4.0	1.2	**14.2**	**12.3**
0946	cell surface protein	0.9	1.1	1.4	1.1	**2.2**	**2.6**
1785	phosphoglycerate mutase	1.1	1.0	1.2	1.0	**1.4**	**1.6**
1697	extracellular protein	1.1	1.1	1.2	1.0	**3.2**	**3.8**
3092	succinate-semialdehyde dehydrogenase	1.2	1.1	1.2	1.1	**1.7**	1.4

Identification of genes with a specifically increased expression in response to lactic acid

The focus of this research was on the identification of genes that showed an increased expression in response to lactic acid, regardless of absolute and relative growth rate. These features are fulfilled by the 18 genes or operons in group 4 (Fig.2; Table 2). Several of these also showed an increased expression in response to other environmental differences (especially the presence of 800 mM NaCl).

Probably the most striking feature from this group of genes was the multiple cell surface protein encoding genes that were found. Analysis of the corresponding regions of these genes on the *L. plantarum* genome indicated that three operons encoding cell surface proteins could be identified, not all genes of which were represented in our datasets. The corresponding proteins showed low similarity to other proteins in database searches. Prediction of transmembrane helices, cell wall anchors, expected pI values, and amino acid lengths from the complete sets of genes, revealed a structural homology between these three operons (Fig. 3). This "structural conservation" might indicate that the proteins within one operon interact. Multiple sequence alignment of the amino acid sequences from the corresponding genes showed a limited number of highly conserved regions (data not shown). The array data indicated that, under inducing conditions, the absolute expression level of the lp_3679-lp_3676 operon was among the highest transcript levels that were measured. This would suggest that the corresponding proteins are abundantly present on the cell surface. Indeed scanning electron microscopy revealed striking morphological differences between the stressed and the non-stressed cells (Fig. 4): whereas the cells that were not subjected to lactic acid stress showed a smooth appearance, the stressed cells had a remarkably rough surface. These structures were mainly situated on the longitudinal sides of the cells. It is tempting to speculate that these cell surface proteins may have a structural, physical function in combating lactic acid stress. To our current knowledge, a specific role for structural cell surface proteins in organic acid stress response has not been implicated before.

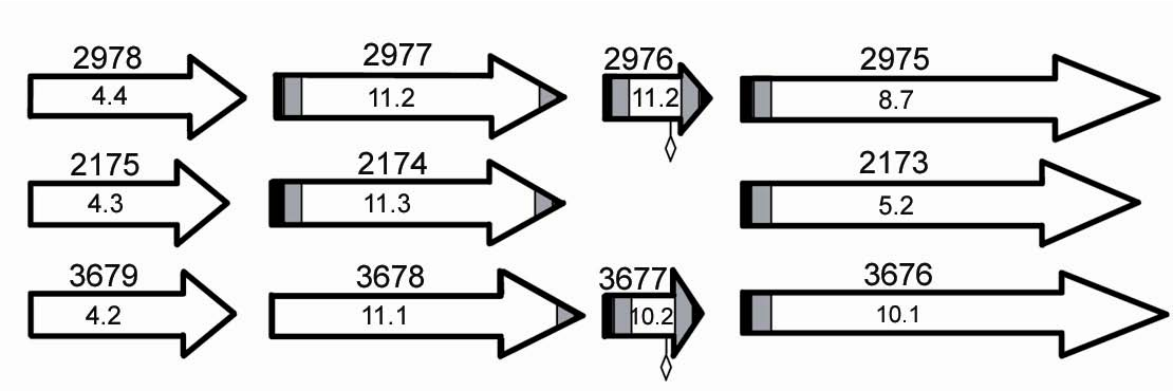

Figure 3. Graphical representation of three cell surface protein operons that showed increased expression upon continuous lactic acid stress. The proteins were vertically aligned in correspondence with the homology between the operons. The size of the arrows is proportional to the length of the proteins in amino acids. Predicted intracellular regions are depicted in black; Predicted transmembrane regions are in grey; Predicted extracellular regions are in white; The orf-number is depicted above the arrows. pI values are depicted within the arrows; The cell wall anchor motif "LPQTxE" is depicted with the diamond symbol.

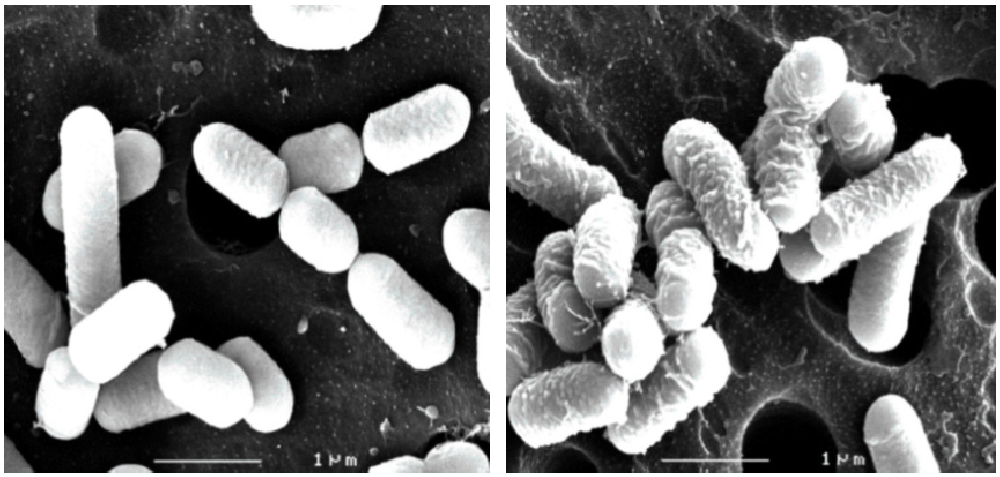

Figure 4. Scanning electron microscopy images from *L. plantarum* WCFS1 cells that were cultured in steady state cultures at pH 4.8 in the absence (-) or presence (+) of 300 mM of additional sodium lactate.

Other genes that showed a growth rate independent lactic acid response are associated with various stress responses in prokaryotes, i.e. the genes encoding Clp protease, an excinuclease (DNA damage), catalase (peroxide stress), and the Dpr-like protein (peroxide stress) (19, 37, 38, 50). Increased expression was also observed for these genes in the cells that were cultures in the presence of 0.8 M NaCl. These results suggest that lactic acid stress in *L. plantarum* WCFS1 also induces a more general stress response. An overlap between the stimulon by lactic acid and those for peroxide and UV-radiation has also been reported for *Lactococcus lactis* (19).

Two other genes that may serve a specific role in the counteracting of lactic acid stress are squalene synthase and phytoene synthase, which are involved in the biosynthesis of sterols. Sterols can increase the rigidity of the membrane (47), which could possibly limit the influx of lactic acid.

Moreover, two predicted regulators were represented in the group of genes that showed increased expression in response to lactic acid. These are specifically interesting because of their potential role in the regulation of an organic acid stress response.

Identification of genes with a specifically decreased expression in response to lactic acid

Group 2 (Fig.2; Table 3) shows 10 genes or operons that showed a decreased expression in response to lactic acid regardless of absolute and relative growth rate, several of which also showed a decreased expression in response to a lower pH itself and lactate at pH 6.0. The putative functions of these genes vary widely.

Table 3. Genes / operons from group 2 of the hierarchical cluster (Figure 2). This group is characterized by the decreased expression of genes / operons in response to lactic acid. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nos.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. μ	Lactic acid / lower abs. μ
0124	cation transporting P-type ATPase	1.0	1.4	**0.5**	**0.4**	0.8	**0.5**
1799	hypothetical protein	1.2	1.0	**0.6**	0.7	0.7	0.8
2929-2930	hypothetical protein; hypothetical protein	1.1	1.0	0.8	**0.5**	**0.4**	0.6
2931-2932	anaerobic ribonucleotide reductase activator protein; anaerobic ribonucleoside-triphosphate reductase	1.3	1.3	0.8	0.7	**0.5**	0.8
2664-2665	hypothetical protein; respons regulator	1.2	**1.3**	0.7	0.8	**0.6**	0.8
2872-2874	hypothetical protein; alcohol dehydrogenase; oxidoreductase	1.6	**1.8**	**0.5**	**0.4**	**0.4**	0.6
2325-2326	thiamine biosynthesis protein; cysteine desulfurase	1.2	0.9	0.8	0.7	**0.6**	0.9
2142-2143	glycosyltransferase, hypothetical protein	1.3	1.3	0.7	0.8	**0.5**	**0.5**
1568	Penicillin-binding protein	0.9	0.9	0.8	0.9	**0.5**	**0.6**
1317-1319	integral membrane protein; pseudouridylate synthase	0.9	0.9	0.8	0.9	0.7	**0.7**

Effect of lactic acid on the expression of genes that are known for their role in (in)organic acid response

In addition to the genes that were identified to show differential expression in response to lactic acid, we studied the expression of a number of genes that have been associated with (organic) acid response in prokaryotes in earlier studies.

Although the role of the F_1F_0-ATPase in acid adaptation has been clearly demonstrated in studies with both mutant lactic acid bacterium and a strain with an increased activity of the F_1F_0-ATPase (11, 46), no differential expression of these genes could be observed in this study. It should be noted that these effects were observed at low pH values in the presence of *inorganic* acids. It is doubtful whether this system play a significant role in stress attributed to lactic acid. The amount of energy that this system would require when the cell has to cope with a massive influx of this organic acid, could ultimately lead to an arrest of growth (5).

If the uncoupling effect would be the main factor responsible for lactic acid stress, one would expect major problems in proton motive force dependent transport processes, whereas the results that are presented in this paper do not indicate a lactic acid-specific adaptation of the expression of genes encoding transport proteins. Furthermore, no lactic acid

specific increased expression was observed of genes that could play a role in restoring the membrane potential, such as cation transporters, and the malolactic fermentation genes. Remarkably enough, a cation transporter (lp_0124) that did respond to lactic acid showed decreased expression (Table 3).

Table 4. Genes / operons from group 5 of the hierarchical cluster (Fig.2). This group is characterized by the increased expression of genes / operons under the combination of lactic acid and a lower absolute growth rate. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nos.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. µ	Lactic acid / lower abs. µ
0898 or 0899	GroEL supressor protein SugE / hypothetical protein	0.9	0.8	1.4	1.0	0.9	**3.0**
3470	transcription regulator of beta-galactosidase gene	1.0	1.2	1.6	1.2	1.3	**2.1**
1303a	cell surface SD repeat protein	0.8	1.0	1.5	1.2	1.5	**2.3**
0779-0780	nucleotide kinase (putative); hypothetical protein	0.9	**1.7**	1.5	1.2	1.2	**2.5**
2668	small heat shock protein	0.9	1.1	1.3	1.1	1.3	**5.4**
1155-1156	cellobiose PTS, EIIC; cell surface hydrolase (putative)	0.8	1.0	1.2	1.1	1.1	**1.6**
3658-3660	Ribose uptake and conversion, rbsU; rbsD; rbsK3	1.3	0.8	1.1	**1.7**	1.6	**2.9**
3416	putative transcription regulator	1.6	2.0	0.7	0.9	1.5	**6.0**
2596	formate acetyltransferase activating enzyme	1.4	1.8	0.8	0.9	1.7	**3.3**
3673 or 3675	cysteine aminopeptidase / signal peptidase I	1.1	1.0	1.0	1.0	1.3	**1.7**
3207	aminotransferase	**0.5**	**0.6**	1.2	1.3	1.5	1.5
0800	cell surface protein	0.7	**0.6**	1.5	1.3	**2.0**	**1.6**
3545-3548	L-iditol 2-dehydrogenase ; galacitol PTS , pts32C; pts32B ; pts32A	0.4	**0.2**	1.0	3.9	1.5	**6.4**
0472-0473	putative integral membrane protein ; hypothetical protein	0.6	0.5	1.1	1.5	**1.2**	1.4
2570	aspartate-semialdehyde dehydrogenase	0.8	**0.5**	1.2	1.1	1.2	1.5
3500-3501	oxidoreductase ; hypothetical protein	0.9	**0.6**	0.9	1.2	1.1	1.4
2643	lipoate protein ligase	0.9	**0.3**	0.7	1.1	1.2	1.9
3418	phosphoenolpyruvate carboxykinase	1.0	**0.2**	0.6	1.3	0.8	3.0
1912	phosphoenolpyruvate synthase	0.5	**0.2**	0.5	1.6	0.7	**7.0**
3554-3557	arabinose uptake and conversion, araD; araD; araB; araT	1.1	**0.6**	**0.6**	1.0	1.2	**2.7**
3520-3524	pts32A; transcriptional regulator; pts32BC; glucokinase regulatory protein; hypothetical protein	1.3	**0.6**	0.7	1.0	1.0	1.9
0498	fucose transport protein	0.9	**0.3**	0.5	0.8	1.1	**2.5**
3538-3539	transketolase; transaldolase	1.0	**0.7**	1.0	1.0	1.2	**2.3**
3540-3542	transport protein; hypothetical protein; hypothetical protein	1.1	0.7	0.8	0.8	1.2	**2.4**
3563-3567	sorbitol uptake and conversion, pts38BC; pts38C ; srlM1 ; srlR2	1.6	0.8	0.8	0.9	1.1	**1.9**
1763	glycosyltransferase	1.0	0.7	1.1	1.0	1.3	1.6
0758	hypothetical protein	0.9	**0.6**	0.8	1.1	1.5	**2.8**
1468-1473	ABC transporter, ATP-binding protein; ABC transporter component (putative); cysteine desulfurase; NifU-like protein; ABC transporter component, iron regulated (putative); iron chelatin ABC transporter, substrate binding protein (putative)	0.8	**0.7**	0.8	1.2	**1.5**	**1.6**

Table 5. Genes / operons from group 3 of the hierarchical cluster (Figure 2). This group is characterized by the decreased expression of genes / operons under the combination of lactic acid and a lower absolute growth rate. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nrs.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. µ	Lactic acid / lower abs. µ
2992	manganese transport protein	1.1	**0.4**	**7.0**	**4.9**	0.7	**0.2**
0230-0233	mannitol uptake and conversion , mtlA ; mtlR ; mtlF ; mtlD	0.9	**0.5**	1.2	1.0	0.9	0.7
0355	putative oxidoreductase	0.8	**0.4**	1.4	1.3	0.8	*0.6*
0197	cell surface protein	0.9	**0.4**	1.4	1.2	0.7	0.7
1497-1500	nitrate reductase, narG ; narH ; narJ ; narI	0.7	**0.5**	1.4	1.4	0.8	**0.7**
2033-2038	aromatic amino acids biosynthesis, aroI ; tyrA ; aroE ; hypothetical protein ; aroF	0.8	**0.5**	1.1	1.6	0.8	**0.5**
1084-1086	aromatic amino acids biosynthesis, aroD1; aroA; aroB	0.8	**0.5**	1.2	1.6	0.9	**0.5**
2082 or 2083 or 2084	hypothetical proteins	0.6	**0.4**	1.1	1.3	0.8	**0.5**
3359-3360	integral membrane protein; integral membrane protein	**0.5**	**0.4**	1.8	1.6	1.1	0.6
2839	hypothetical protein	1.0	1.2	1.4	1.0	0.8	**0.5**
1941	NADH oxidase	1.0	**1.1**	1.2	1.0	0.8	**0.6**
1008	lysine transport protein	0.5	0.9	1.2	1.2	0.6	**0.4**
2136	pyruvate carboxylase	0.8	0.9	1.1	1.0	**0.7**	**0.6**
3184-3185	integral membrane protein; branched chain amino acid transport protein	0.8	0.7	1.2	1.2	**0.5**	**0.4**
1148-1149	glutamyl-tRNA amidotransferase, subunit A; subunit B	0.9	0.9	1.1	1.2	**0.7**	**0.6**
2761-2765	O-acetyltransferase; ADP-ribose pyrophosphatase; hypothetical protein; short-chain dehydrogenase/oxidoreductase; hypothetical protein	1.0	1.0	0.9	0.8	0.8	**0.6**
0977	UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase	1.0	1.0	1.0	0.9	0.9	**0.5**
0881-0883	glutamine ABC transporter, glnH1; glnM; glnP	1.6	0.9	1.0	0.9	0.8	**0.3**
0254-0256	cysteine biosynthesis, cysE; metC1; cysK	1.2	0.6	0.8	1.1	0.4	**0.2**
0190-0192	hypothetical protein; Na^+/H^+-antiporter	0.7	0.8	1.3	0.9	0.9	**0.5**
3042	ABC transporter, ATP-binding and permease protein	0.8	0.8	0.9	0.9	1.2	**0.4**
0113-0116	thiamine biosynthesis, thiM; thiD; thiE; codB	0.6	**0.6**	1.1	1.2	1.1	**0.4**
0571-0572	threonine biosynthesis, hom2; thrB	0.9	**0.7**	0.8	1.0	1.0	**0.5**
3334	adenine deaminase	0.8	**0.2**	0.8	0.9	0.7	**0.1**
2290	intergral membrane protein	0.8	**0.8**	1.0	1.0	0.8	**0.5**
0217-0219	ABC transporter, permease protein; ATP binding domain; putative subunit	0.6	**0.4**	1.2	0.9	0.7	**0.2**
3085	asparagine synthase	0.9	**0.4**	0.9	1.0	1.1	**0.5**
1262-1265	oligopeptide ABC transporter, oppC; oppD; oppE; oppF	1.0	**0.6**	0.9	1.2	1.0	**0.5**
1164-1165	cellobiose PTS EIIC fragment; hypothetical protein	0.7	**0.5**	1.1	1.0	0.9	0.6
0848	permease	1.2	**0.3**	0.9	0.7	0.9	**0.2**
3270	adenylosuccinate synthase	1.3	**0.2**	1.1	0.7	2.0	**0.1**
1729-1731	maltose metabolism, sugar transporter; map3; galM2	1.0	*0.8*	1.1	0.9	1.2	**0.6**
2719-2729	purine biosynthesis cluster	0.8	**0.2**	1.3	1.0	2.1	**0.2**

Metabolic rerouting under the combined effects of lactic acid and a lower absolute growth rate

Groups 3 and 5 (Fig. 2; Tables 4 and 5) contain genes and operons that are differently expressed in cells grown in the presence of lactic acid and a lower absolute growth rate. Within these clusters, several genes involved in energy metabolism showed a differential expression, which could indicate a shift towards alternative fermentation pathways. The complete dataset was studied in more detail to determine whether other genes in these pathways showed similar trends. An overview of these genes and the metabolic routes they are involved in is presented in Table 8 and Fig. 5.

Genes encoding transketolase and transaldolase (lp_3538 and lp_3539), both involved in the pentose phosphate cycle, showed an increased expression (Table 8). Furthermore, an increase of approximately 5 times was observed for the phosphoketolase encoded by lp_3551 (90% confidence interval; Table 8). The co-increased expression of transaldolase with phosphoketolase suggests a shift towards the phosphoketolase by-pass, as is typically performed by bifidobacteria (23). In this pathway the phosphoketolase uses both xylulose 5-P and fructose 6-P as a substrate. Homology searches also suggest that the *L. plantarum* phosphoketolase can convert both substrates (results not shown). As far as we are aware, use of the phosphoketolase by-pass has not previously been described for lactobacilli. It is generally suggested that phosphoketolase only serves a role in the heterolactic fermentation pathway, where it converts xylulose 5-P into acetyl-P and glyceraldehydes 3-P (23).

The genes coding for phosphoenolpyruvate synthase (lp_1912) and phosphoenolpyruvate carboxykinase (lp_3418), responsible for the bypass from pyruvate to oxaloacetate via phosphoenolpyruvate, both showed an increased expression (Table 8). A subsequent conversion to malate by malate dehydrogenase would lead to NAD^+ regeneration. The flux through this pathway seems to be further stimulated by the downregulation of the pyruvate carboxylase (lp_2136) and pyruvate kinase genes (lp_1897; statistically non significant) under the same conditions (Table 8).

The increased expression of lp_2596, the formate acetyl transferase activating enzyme (Table 3), suggests the upregulation of the formate acetyl transferase gene, responsible for the conversion of pyruvate into formate and acetyl-coA. Indeed, an increase in expression (90% confidence interval) of this gene (lp_2598) could be observed (Table 8). A subsequent conversion step in this pathway is that of acetyl-CoA to acetaldehyde and ethanol, during which two NAD^+ are regenerated. This step could be performed by the bifunctional alcohol dehydrogenase / acetaldehyde dehydrogenase (lp_3662) for which a statistically non

significant increase in expression of 150% was observed (Table 8). Alternatively, acetyl-CoA can be converted into acetyl-phosphate and subsequently acetate by phosphotransacetylase and acetate kinase, respectively. However no increased expression was observed for the corresponding genes.

The suggested reroutings make sense in view of the potential effects of lactic acid accumulation on the cell. Conversion of pyruvate into lactic acid is the predominant mechanism for *L. plantarum* to avoid NAD^+ depletion under anaerobic conditions. Other metabolites than lactate have been reported as the end-product of hexose fermentations with *L. plantarum*, but only in cases where oxygen or citrate were present as electron acceptors (14, 30, 44). Therefore, growth inhibition due to pyruvate accumulation and / or due to low NAD^+ levels as a consequence of end product inhibition by lactate, form a plausible explanation of the inhibitory effect caused by lactic acid. Strategies that aim at avoiding pyruvate accumulation and / or regeneration of NAD^+ could be beneficial for *L. plantarum*.

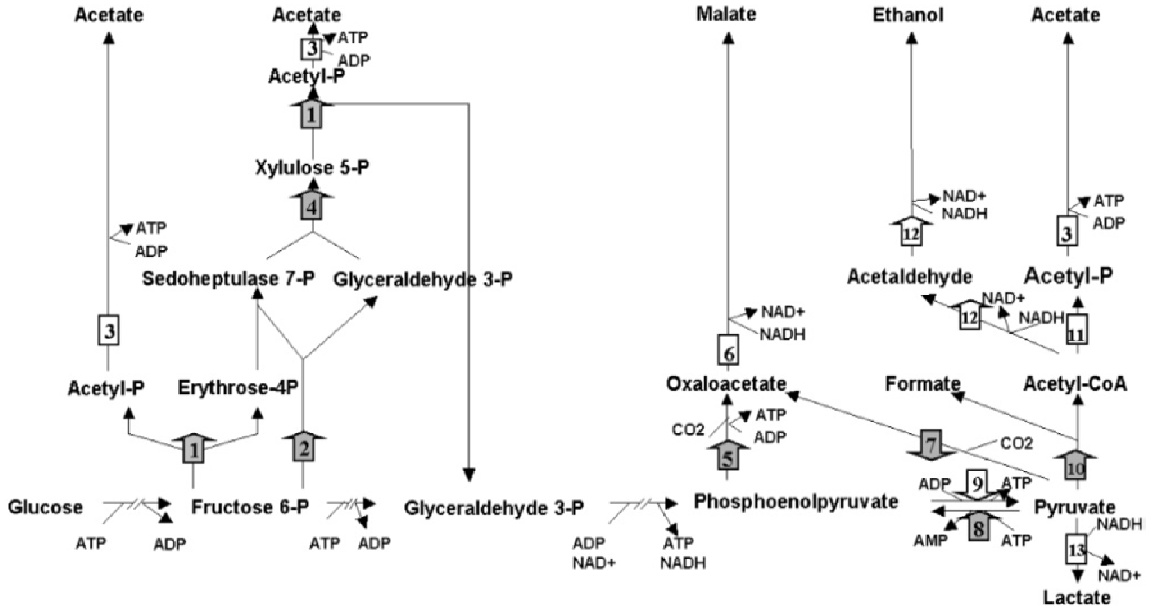

Figure 5. Potential glucose fermentation in *L. plantarum* WCFS1. Block arrows indicate either increased (⇑), decreased (⇓) or unchanged (□) expression of the gene by the presence of lactic acid and a lower absolute growth rate. Arrows with a grey filling represent genes for which the change in expression was within the 90% confidence interval of the Tukey HSD test. (1), phosphoketolase; (2), transaldolase; (3) acetate kinase; (4), transketolase; (5) phosphoenolpyruvate carboxykinase; (6), malate dehydrogenase; (7), pyruvate carboxylase; (8), phosphoenolpyruvate synthase; (9), pyruvate kinase; (10) pyruvate formate-lyase; (11), phosphotransacetylase; (12), bifunctional alcohol dehydrogenase/ acetaldehyde dehydrogenase; (13), lactate dehydrogenase.

The suggested shift towards the phosphoketolase by-pass and the pathways, leading towards malate and ethanol formation, would contribute to a limitation of the pyruvate level. Moreover, these three routes have a clear advantage with respect to the cofactor regeneration. These observations indicate that diminished NAD^+ regeneration due to end product inhibition of the lactate dehydrogenase is an important growth limiting effect of the lactic acid accumulation under anaerobic conditions. This suggestion is confirmed by observations of Ferain and co-workers, who reported that a *L. plantarum* strain that lacked lactate dehydrogenase activity was especially inhibited in its growth under anaerobic conditions (13). Under aerobic conditions the NAD^+ regenerating function of the lactate dehydrogenase could be taken over by NADH oxidase.

The fact that these genes did not show similar trends in response to lactic acid at a higher relative growth rate may indicate that lactic acid is not the only regulatory trigger.

Table 6. Genes / operons from group 6 of the hierarchical cluster (Figure 2). This group is characterized by the increased expression of genes / operons under the combination of lactic acid and a higher relative growth rate. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nos.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. μ	Lactic acid / lower abs. μ
3267	dihydrolipoamide dehydrogenase	0.8	1.0	1.3	1.2	**2.2**	**0.5**
2277	alanine-tRNA ligase	0.8	0.8	1.2	1.2	**1.6**	**0.6**
2187	isoleucine-tRNA ligase	0.9	0.8	1.3	1.2	**1.7**	**0.5**
3059	cell surface protein	1.1	0.8	1.5	1.1	**2.1**	0.7
1980-1981	aspartate-tRNA ligase ; histidine-tRNA ligase	1.0	0.9	1.1	1.3	**1.7**	0.8
1316	leucine-tRNA ligase	0.9	1.1	1.3	1.2	**1.8**	0.8
0793	hypothetical protein	0.8	**1.5**	**1.8**	1.2	1.3	1.0
3412-3414	extracellular protein; cell surface protein; cell surface protein	1.5	1.4	**2.6**	1.2	**1.9**	0.8
0520	ATP-dependent RNA helicase	1.0	0.8	1.4	1.0	**1.6**	0.9
1643-1645	Mucus binding protein; hypothetical protein	0.8	0.6	**2.5**	1.8	**2.5**	0.9
2322	valine-tRNA ligase	0.9	1.1	**1.6**	1.2	**1.8**	1.3
1685	putative transcriptional regulator	1.2	**1.2**	**1.4**	**1.4**	**2.6**	1.1
1446-1450	hypothetical protein; cell surface protein; cell surface protein; hypothetical protein; hypothetical protein	0.9	1.2	1.2	**1.5**	**2.0**	1.3
3174	cyclopropane-fatty-acyl-phospholipid synthase	1.3	**1.5**	1.2	1.1	**2.0**	1.2
1673-1681	fatty acid biosynthesis cluster, fabD; fabG1; fabF; accB2; fabZ	1.4	1.1	0.9	0.9	**18.5**	1.2
0315-0319	spermidine/putrescine ABC transporter, potA; potC; potB; potA	1.0	0.9	1.0	1.0	**1.8**	1.1
1021	DNA-directed RNA polymerase, β subunit	0.7	0.8	1.2	0.9	**2.1**	1.3
0262-0265	trehalose uptake and conversion, treR; treA; pts4ABC; pts5ABC	1.3	0.8	1.3	1.0	**1.8**	1.3
2502	glucose-6-phosphate isomerase	0.9	**0.6**	1.2	0.9	**1.7**	1.1

Table 7. Genes / operons from group 1 of the hierarchical cluster (Figure 2). This group is characterized by the decreased expression of genes / operons under the combination of lactic acid and a higher relative growth rate. The depicted values indicate the expression ratios between the condition of interest and the reference condition. Numbers depicted in bold are within the 95% confidence interval of the Tukey HSD test.

Orf-nos.	Assigned function	Mild osmotic stress	Osmotic stress	Acidic pH	Lactate	Lactic acid / higher rel. μ	Lactic acid / lower abs. μ
3611-3614	sugar transport protein; myo-inositol 2-dehydrogenase; hypothetical protein; oxidoreductase	1.5	0.7	**0.5**	0.8	**0.5**	1.3
3604	myo-inositol transport protein	1.1	0.8	**0.4**	0.7	0.5	1.3
0746-0750	phosphate ABC transporter, pstE; pstD; pstC; pstB; pstA	1.2	1.1	**0.5**	1.1	**0.6**	**1.8**
0856	acyltransferase (putative)	1.1	0.9	0.8	0.7	**0.6**	1.2
3290	transcription regulator	0.9	1.0	0.8	0.8	**0.5**	1.2
1358 or 1359	integral membrane protein / membrane-bound protease, CAAX family	0.9	0.9	0.7	0.9	**0.5**	1.1
3283	hypothetical protein	0.8	**1.3**	0.9	0.8	**0.6**	1.2
0325-0327	transcription regulator; ABC transporter, ATP-binding protein; hypothetical protein	0.7	1.0	0.7	1.0	**0.5**	1.0
2579	hypothetical protein	1.0	1.0	1.0	1.0	**0.5**	0.9
1283	sulfatase	0.7	**0.6**	0.9	1.2	0.3	0.9
2732-2733	oxidoreductase; oxidoreductase	1.4	**0.6**	0.8	1.0	**0.5**	1.0
2075-2077	regulator; nitrate ABC transporter, nrtC; nrtB	1.1	**0.6**	0.8	1.2	**0.6**	1.1
1908	hypothetical protein	0.8	**3.0**	1.3	1.3	**0.2**	0.7
0892-0893	putative regulator; putative transport protein	1.1	**1.5**	1.0	1.2	**0.5**	0.9
0343-0344	teichoic acid ABC transporter, tagG; tagH	1.1	1.1	0.9	1.2	**0.4**	0.8
3281	putative permease	0.9	**1.5**	0.8	0.9	**0.4**	0.9
2580	alkaline phosphatase superfamily protein	1.0	**1.5**	1.0	1.0	**0.4**	0.7
0871 or 0872	hypothetical protein / hypothetical protein	0.9	1.3	0.9	1.0	**0.4**	0.8
0346-0348	integral membrane protein; putative transcription regulator; permease	1.0	1.3	0.8	1.0	**0.5**	0.7
2578	cell surface protein	0.9	1.1	1.1	1.0	**0.6**	0.8
3468-3469	lactose transport protein; β-galactosidase	0.8	0.8	**2.7**	1.0	**0.2**	0.5
3466	branched chain amino acid transport protein	1.0	0.8	1.5	1.0	**0.3**	0.6

Growth rate related effects

From clusters 1 and 5 versus cluster 3 and 6 in the hierarchical cluster (Fig. 2) it becomes apparent that altered growth characteristics (in comparison with the reference condition) form an important parameter in gene expression studies.

The coherence within these clusters clearly indicates that the observed changes are not coincidental: the combined effect of lactic acid and a lower absolute growth rate resulted in, among other effects, the decreased expression of genes involved in transport processes and amino acid uptake and the previously discussed differential expression of genes involved in metabolic rerouting (Table 4 and 5). Among the genes that responded towards the effect of lactic acid and a higher relative growth rate, increased expression was observed for a

remarkable high number of tRNA ligases (Table 6). Based on the fact that the difference in absolute growth rate does not lead to a structural difference in the expression of genes that have been related to glucose limitation (unpublished results), we conclude that this difference does not increase the effect of the glucose starvation on the cells.

Especially in logarithmically growing cultures, it is generally difficult to avoid changes in growth rate as a consequence of the imposed stress. The effect of altered absolute growth rate on gene expression is illustrated by three papers on proteome / transcriptome changes in *E. coli* in response to acetate (2, 25, 35). Transcriptome studies of the effect of a growth inhibiting concentration of this organic acid on logarithmically growing cells, revealed a decreased expression for a large number of genes involved in transcription and translation (2). Other studies recognized this problem and circumvented it by using an acetate concentration that hardly affects growth rate (25, 35). However, for most stress response studies this will not be preferable since the effects on gene expression will also be limited. Experiments with steady state cultures of *E. coli* confirm the effect of absolute growth rate on both metabolome and gene expression (21, 45).

Whereas the use of steady state cultures with an equal dilution rate appears the method of choice to circumvent absolute growth rate related effects, our data indicate that differences in relative growth rate should *also* be accounted for.

Table 8. Genes that were differentially expressed under the combination of lactic acid and a lower absolute growth rate and that suggest metabolic rerouting under this condition. See also Fig. 5.

Orf-nos.	Assigned function	Expression ratio
3538	transketolase	2.3
3539	transaldolase	2.3
3551	phosphoketolase	5.4
1912	phophoenolpyruvate synthase	7.0
3418	phosphoenolpyruvate carboxykinase	3.0
2136	pyruvate carboxylase	0.6
1897	pyruvate kinase	0.5
2596	formate acetyl transferase activating enzyme	3.3
2598	formate acetyl transferase	1.8
3662	alcohol dehydrogenase / acetaldehyde dehydrogenase	2.5

Concluding remarks

The experimental design that was applied in this study allowed the different effects of lactic acid to be distinguished, and resulted in new insights into both the effect of lactic acid stress on *L. plantarum* and the potential environmental response mechanisms of the organism. Especially those genes that showed increased expression towards lactic acid regardless of growth rate related effects may be good target genes for strain improvement. The metabolic reroutings in response to lactic acid at a lower absolute growth rate suggest an inhibitory effect by lactic acid at lower growth rates due to decreased NAD^+ regeneration and / or pyruvate accumulation. The fact that these effects only occurred in combination with the lower absolute growth rate indicates that lactic acid itself is not the regulatory trigger for these reroutings.

The results presented in this paper demonstrate the importance of good experimental design in studies where more than one physico-chemical parameter is affected by the subject of study. A poor distinction between primary and secondary effects, such as growth related effects, will complicate data interpretation and rational selection of genes for further studies and should therefore be avoided.

Acknowledgements

We thank Martien P. M. Caspers for his contribution in the microarray analysis, Adriaan C. van Aelst for the scanning electron microscopy and Robert A. van den Berg, Willem M. de Vos and Michiel Kleerebezem for useful discussions.

Literature

1. **Ang, S., C. –Z. Lee, K. Peck, M. Sindici, U. Matrubutham, M. A. Gleeson, and J. –T. Wang.** 2001. Acid-induced gene expression in *Helicobacter pylori*, Study in genomic scale by microarray. Infect. Immun. **69:**1679-1686.
2. **Arnold, C. N., J. McElhanon, A. Lee, R. Leonhart, and D. A. Siegele.** 2001. Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. J. Bacteriol. **183:**2178-2186.
3. **Axe, D. D., and J. E. Bailey.** 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol. Bioeng. **47:**8-19.
4. **Azcarate-Peril, M. A., E. Altermann, R. L. Hoover-Fitzula, R. J. Cano, and T. R. Klaenhammer.** 2004. Identification and inactivation of genetic loci involved with *Lactobacillus acidophilus* acid tolerance. Appl. Environ. Microbiol. **70:**5315-22.
5. **Bond, D. R., and J. B. Russell.** 1996. A role for fructose 1,6-diphosphate in the ATPase mediated energy spilling reaction of *Streptococcus bovis*. Appl. Environ. Microbiol. **62:**2095-2099.
6. **Booth, I. R.** 1985. Regulation of cytoplasmic pH in bacteria. Microbiol. Rev. **49:**359-378.
7. **Cook, G. M., and J. B. Russell.** 1994. The effect of extracellular pH and lactic acid on pH homeostasis in *Lactococcus lactis* and *Streptococcus bovis*. Curr. Microbiol. **28:**165-168.
8. **Cotter, P. D., and C. Hill.** 2003. Surviving the acid test, Responses of Gram-positive bacteria to low pH. Microbiol. Mol. Biol. Rev. **67:**429-453.
9. **Cotter, P. D., C. G. Gahan, and C. Hill.** 2001. A glutamate decarboxylase system protects *Listeria monocytogenes* in gastric fluid. Mol. Microbiol. **40:**465-475.
10. **Diez-Gonzalez, F., and J. B. Russell.** 1997. The effects of carbonylcyanide-*m*-chlorophenylhydrazone (CCCP) and acetate on *Escherichia coli* O157, H7 and K-12, uncoupling versus anion accumulation. FEMS Microbiol. Lett. **151:**71-76.
11. **Drici-Cachon, Z., J. Guzzo, J. –F. Cavin, and C. Diviès.** 1996. Acid tolerance in *Leuconostoc oenos*. Isolation and characterization of an acid-resistant mutant. Appl. Microbiol. Biotechnol. **44:**785-789.
12. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. **95:**14863-14868.
13. **Ferain T, J. N. Hobbs Jr., J. Richardson, N. Bernard, D. Garmyn, P. Hols, N. E. Allen, and J. Delcour.** 1996. Knockout of the two ldh genes has a major impact on

peptidoglycan precursor synthesis in *Lactobacillus plantarum*. J. Bacteriol. **178:**5431-5437.

14. **Ferain, T., A. N. Schanck, and J. Delcour.** 1996. ^{13}C Nuclear Magnetic Resonance analysis of glucose and citrate end products in an ldhL-ldhD double-knockout strain of *Lactobacillus plantarum*. J. Bacteriol. **178:**7311-7315.
15. **Fischetti, V. A., V. Pancholi, and O. Schneewind.** 1990. Conservation of a hexapeptide sequence in the anchor region of surface proteins from Gram-positive cocci. Mol. Microbiol. **4:**1603-1605.
16. **Fisher, M. A., B. B. Plikaytis, and T. M. Schinnick.** 2002. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. J. Bacteriol. **184 :**4025-4032
17. **Giard, J. -C., J. –M. Laplace, A. Rincé, V. Pichereau, A. Benachour, C. Leboeuf, S. Flahaut, Y. Auffray, and A. Hartke.** 2001. The stress proteome of *Enterococcus faecalis*. Electrophoresis. **22:**2947-2954.
18. **Giraud, E., B. Lelong, and M. Raimbault.** 1991. Influence of pH and initial lactate concentration on the growth of *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **36:**96-99.
19. **Hartke, A., S. Bouché, J. –C. Giard, A. Benachour, P. Boutibonnes, and Y. Auffray.** 1996. The lactic acid stress response of *Lactococcus lactis* subsp. *lactis*. Curr. Microbiol. **33:**194-199.
20. **Houtsma, P. C., J. C. de Wit, and F. M. Rombouts.** 1993. Minimum inhibitory concentration (MIC) of sodium lactate for pathogens and spoilage organisms occurring in meat products. Int. J. Food Microbiol. **20:**247-257.
21. **Hua, Q., C. Yang, T. Oshima, H. Mori, and K. Shimizu.** 2004. Analysis of gene expression in *Escherichia coli* in response to changes in growth-limiting chemostat cultures. Appl. Environ. Microbiol. **70:**2354-2366.
22. **Hutkins, R. W., and N. L. Nannen.** 1993. pH homeostasis in Lactic Acid Bacteria. J. Dairy. Sci. **76:**2354-2365.
23. **Kandler, O.** 1983. Carbohydrate metabolism in lactic acid bacteria. Antonie van Leeuwenhoek **49:**209-224.
24. **Kashket, E. R.** 1987. Bioenergetics of lactic acid bacteria, cytoplasmic pH and osmotolerance. FEMS Microbiol. Rev. **46:**233-244.
25. **Kirkpatrick, C., L. M. Maurer, N. E. Oyelakin, Y. N. Yoncheva, R. Maurer, and J. L. Slonczewski.** 2001. Acetate and formate stress, opposite responses in the proteome of *Escherichia coli*. J. Bacteriol. **183:**6466-6477.

26. **Kleerebezem M., J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:**1990-1995.
27. **Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a Hidden Markov Model, application to complete genomes. J. Mol. Biol. **305:**567-580.
28. **Len, A. C. L., D. W. S. Harty, and N. A. Jacques.** 2004. Proteome analysis of *Streptococcus mutans* metabolic phenotype during acid tolerance. Microbiol. **150:**1353-1366.
29. **Lim, E. M., S. D. Ehrlich, and E. Maguin.** 2000. Identification of stress-inducible proteins in *Lactobacillus delbrueckii* subsp. *bulgaricus*. Electrophoresis **21:**2557-2561.
30. **Lindgren, S. E., L. T. Axelsson, and R. F. McFeeters.** 1990. Anaerobic L-lactate degradation by *Lactobacillus plantarum*. FEMS Microbiol. Lett. **66:**209-214.
31. **McDonald, M. C., H. P. Flemming, and H. M. Hassan.** 1990. Acid tolerance of *Leuconostoc mesenteroides and Lactobacillus plantarum*. Appl. Environ. Microbiol. **56:**2120-2124.
32. **Merrell, D. S., M. L. Goodrich, G. Otto, L. S. Tompkins, and S. Falkow.** 2003. pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*. Infect. Immun. **71:**3529-3539.
33. **Olsen, E. B., J. B. Russell, and T. Henick-Kling.** 1991. Electrogenic L-malate transport by *Lactobacillus plantarum*, a basis for energy derivation from malolactic fermentation. J. Bacteriol. **173:**6199-6206.
34. **Padan, I., D. Zilberstein, and S. Schuldiner.** 1981. pH homeostasis in bacteria. Biochim. Biophys. Acta. **650:**151-166.
35. **Pieterse, B., R. H. Jellema, and M. J. van der Werf.** 2006. Quenching of microbial samples for increased reliability of microarray data. **64:**207-216.
36. **Polen, T., D. Rittmann, V. F. Wendisch, and H. Sahm.** 2003. DNA Microarray analyses of the long-term adaptive response of *Escherichia coli* to acetate and propionate. Appl. Environ. Microbiol. **69:**1759-1774.
37. **Price, C. W., P. Fawcett, H. Cérémonie, N. Su, C. K. Murphy, and P. Youngman.** 2001. Genome-wide analysis of the general stress response of *Bacillus subtilis*. Mol. Microbiol. **41:**757-774.
38. **Quivey, R. G., R. C. Faustoferri, K. A. Clancy, and R. E. Marquis.** 1995. Acid adaptation in *Streptococcus mutans* UA159 alleviates sensitization to environmental stress due to recA deficiency. FEMS Microbiol. Lett. **126:**257-261.

39. **Russell, J. B.** 1992**.** Another explanation for the accumulation of fermentation acids at low pH, anion accumulation versus uncoupling. J. Appl. Bacteriol. **73:**363-370.

40. **Russell, J. B., and F. Diez-Gonzalez.** 1998. The effects of fermentation acids on bacterial growth. Adv. Microbial Physiol. **39:**205-234.

41. **Salmond, C. V., R. G. Kroll, and I. R. Booth.** 1984. The effect of food preservatives on pH homeostasis in *Escherichia coli*. J. Gen. Microbiol. **130:**2845-2850.

42. **Shelef, L. A.** 1994. Antimicrobial effects of lactates, A review. J. Food Protection. **57:**445-450.

43. **Stancik, L. M., D. M. Stancik, B. Schmidt, D. M. Barnhart, Y. N. Yoncheva, and J. L. Slonczewski.** 2002. pH-dependent expression of periplasmatic proteins and amino acid catabolism in *Escherichia coli*. J. Bacteriol. **184:**4246-4258.

44. **Tseng, C. -P, and T. J. Montville.** 1990. Enzyme activities affecting end product distribution by *Lactobacillus plantarum* in response to changes in pH and O_2. Appl. Environ. Microbiol. **56:**2761-2763.

45. **Tweeddale, H., L. Notley-McRobb, and T. Ferenci.** 1998. Effect of slow growth on metabolism of *Escherichia coli,* as revealed by global metabolite pool ("metabolome") analysis. J. Bacteriol. **180:**5109-5116.

46. **van de Guchte, M., P. Serror, C. Chervaux, T. Smokvina, S. D. Ehrlich, and E. Maguin, E. (2002).** Stress responses in lactic acid bacteria. Antonie van Leeuwenhoek. **82:**187-216.

47. **van der Rest, M. E., A. H. Kamminga, A. Nakano, Y. Anraku, B. Poolman, and W. N. Konings.** 1995. The plasma membrane of Saccharomyces cerevisiae, structure, function, and biogenesis. Microbiol. Rev. **59:**304-322.

48. **Wen, Y., E. A. Marcus, U. Matrubutham, M. A. Gleeson, D. R. Scott, and G. Sachs.** 2003. Acid-adaptive genes of *Helicobacter pylori*. Infect. Immun. **71:**5921-5939.

49. **Wilkins, J. C., K. A. Homer, and D. Beighton.** 2002. Analysis of *Streptococcus mutans* proteins modulated by culture under acidic conditions. Appl. Environ. Microbiol. **68:**2382-2390.

50. **Yamamoto, Y., M. Higuchi, L. B. Poole, Y. Kamio.** 2000. Role of the dpr product in oxygen tolerance in *Streptococcus mutans*. J. Bacteriol. **182:**3740-3747.

Chapter 5

Analysis of lactic acid tolerance in *Lactobacillus plantarum* derivatives mutated in lactic acid-stimulon associated functions

Bart Pieterse, Jolanda Lambert, Mariët J. van der Werf, and Michiel Kleerebezem

Abstract

Lactic acid is an effective inhibitor of bacterial growth. This feature is mainly attributed to the undissociated form of the organic acid. In a previous study, multiple *Lactobacillus plantarum* genes were identified that displayed altered expression levels upon exposure of cells to undissociated lactic acid. In this study gene-replacement or overexpression mutants were constructed for two putative regulator encoding genes, a gene encoding the peroxide resistance protein Dpr, and a complete cell surface proteins encoding operon. The growth and survival characteristics of these strains were tested under conditions with high lactic acid concentrations. No significant effects were observed from the replacement or overexpression of the genes. These results may indicate that individual genes of the lactic acid-specific stimulon of *L. plantarum* contribute only marginally to lactic acid tolerance, which could be more dependent on the concerted activity of all or a subset of these genes.

Introduction

Lactobacilli are commonly used in fermentation processes for the production of lactate, which contributes to the conservation of food and feed products. Lactate is the predominant metabolite produced during these fermentations, resulting in extracellular accumulation of this organic acid and a decrease of the extracellular pH. Ultimately, the lactic acid concentrations that are reached during fermentations strongly inhibit the growth of lactobacilli themselves (7, 13). The inhibitory effect of organic acids is mainly caused by the undissociated form, which can easily diffuse over the cell membrane towards the more alkaline cytosol (2, 7). Subsequently, a part of the lactic acid dissociates due to the higher intracellular pH. Different theories that have been postulated to explain the inhibitory effects of this process on the cell have previously been discussed (Chapter 4).

From a previous study on the effect of lactic acid on transcription profiles, 18 transcription units have been identified in *Lactobacillus plantarum* WCFS1 that showed a specific increase in expression in response to lactic acid stress in steady state cultures (11; Chapter 4). Among the genes in this group, two transcriptional regulators (lp_0965 and lp_1396), several genes that have previously been associated with stress response (e.g. the genes encoding catalase, Clp-protease, peroxide resistance protein Dpr, and an excinulease), and three cell surface protein encoding operons (lp_2978-lp_2975, lp2175-lp_2173, and lp_3679-lp_3676) were identified.

Three genes and an operon from this group were selected for the validation of their role under conditions with high lactic acid concentrations: The transcriptional regulators encoded by lp_0965 and lp_1396 were selected for their potential role in the regulation of the expression of lactic acid stress responsive genes; The *dpr* gene (lp_3128), which encodes a peroxide resistance protein on which extensive research has only been performed in *Streptococcus* species; The cell surface proteins encoding cluster lp_3679-lp_3676 was included in this study based on the observation that their increased expression corresponded with morphological changes, and on the relatively high fold-induction observed for these genes (>10 fold induction in response to lactic acid). Replacement and / or overexpression *L. plantarum* mutant derivatives were constructed. The resulting strains were tested for their ability to grow at high concentrations of lactic acid, and for their survival in time at a lethal combination of pH and sodium lactate.

Materials and methods

Bacterial strain and growth conditions

All strains that were constructed are derived from *L. plantarum* WCFS1 (9). Cultivations were performed on MRS broth (Difico). When appropriate either erythromycin or chloramphenicol was added at a final concentration of 10 µg/ml. *Escherichia coli* DH5*α* (3) was used as cloning host and was grown aerobically in TY medium, supplemented with erythromycin or chloramphenicol when appropriate (14).

DNA techniques

For DNA manipulations in *E. coli* standard procedures were applied (14). Plasmid DNA isolations from *E. coli* were performed using Jetstar columns according to the manufacturer's protocols (Genomed GmbH, Bad Oberhausen, Germany). For the isolation and purification of DNA fragments from agarose gel, and for the purification of ligation and digestion mixtures the Wizard SV gel and PCR clean-up system was applied according to the manufacturer's instructions (Promega, Madison, USA). DNA isolation and transformation in *L. plantarum* was performed as previously described (4, 8). Primers were purchased from Isogen Bioschience (Maarsen, The Netherlands).

Construction of the replacement vectors

All gene replacements vectors that were constructed are based on pNZ5317 (Fig 1; J. Lambert, unpublished results).

The 5' and 3' flanking regions (approximately 1 kb) of the target genes were amplified by PCR with a mixture of *Taq* and *Pwo* polymerase (Roche Diagnostics, Almere, the Netherlands) using *L. plantarum* WCFS1 chromosomal DNA as the template and the primer combinations gene-UF and gene-UR, or gene-DF and gene-DR, respectively (Table 1).

For the construction of the replacement constructs for lp_0965 and lp_3676-lp_3676, the 3' flanking region was ligated into pNZ5317 which had been digested with *Ecl*136II. Following ligation the DNA in the mixture was digested again with *Ecl*136II, to digest backligated vector material. Subsequently, DNA was purified from the mixture and transformed to *E. coli* DH5*α*. Chloramphenicol resistant transformants were selected and analyzed by PCR to select clones containing the anticipated 3'-flanking region in the proper orientation. The correct plasmids were isolated and digested with *Swa*I. The 5'-flanking regions of the target genes were amplified and cloned into the *Swa*I digested mutagenesis vectors that already contain the corresponding 3'-flanking region. Analogous to the procedure applied above, the ligation mixture was digested with *Swa*I, prior to purification and

transformation to *E. coli* DH5*α*. Transformants harboring the anticipated double-cross over mutagenesis constructs were selected by PCR using the primer combination gene-UF with RB85, and gene-DR with RB87.

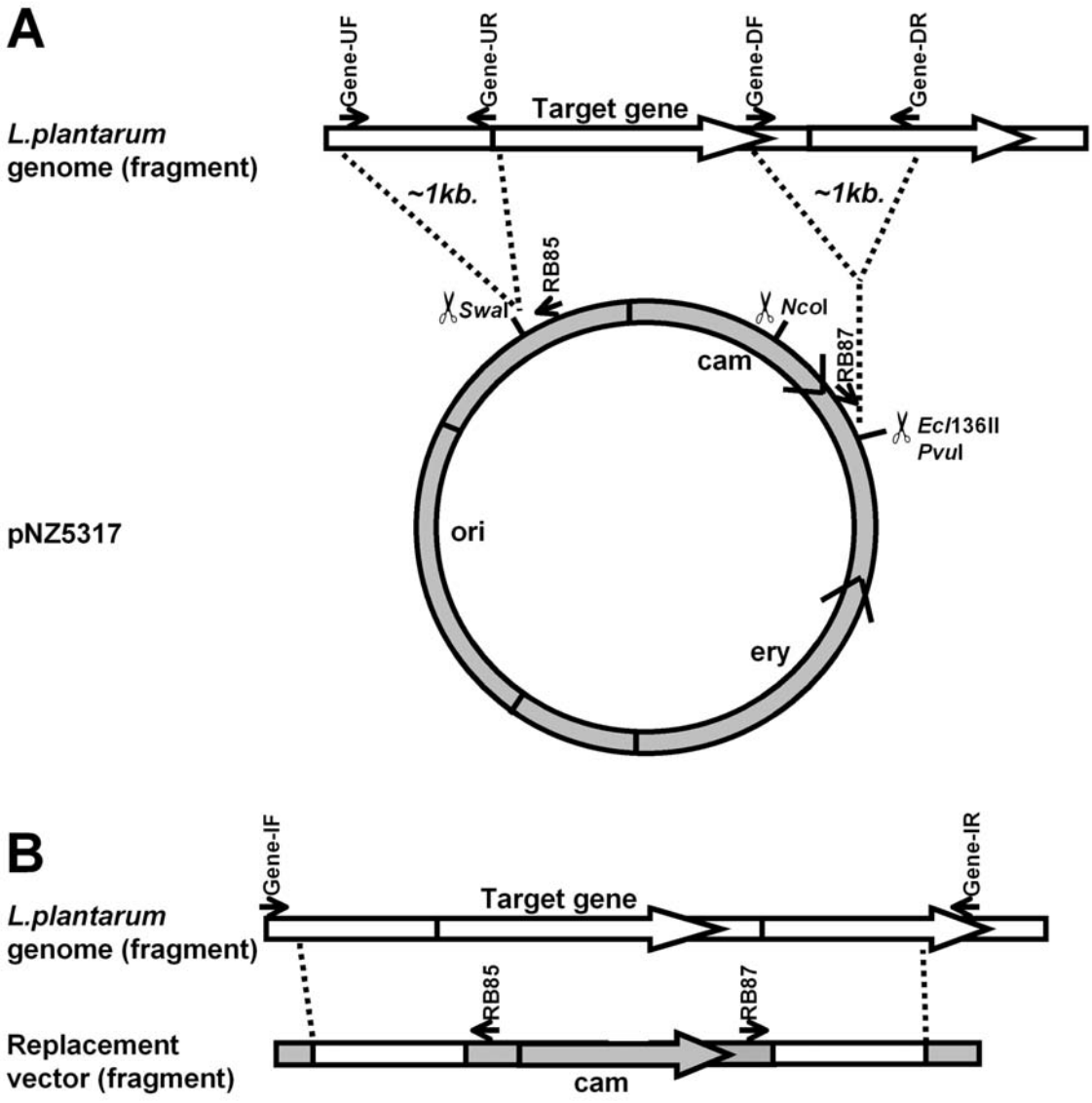

Figure 1. Schematic representation of the construction of the replacement vector (A) and the actual gene replacement by a double cross-over event (B). Only the most relevant features are depicted. DNA of *L. plantarum* chromosomal origin is depicted in white, while DNA derived from pNZ5317 origin is depicted in gray. Insertion and cross over events are depicted by the dotted lines. Primers that were used in these constructions are indicated by the small black arrows. The "gene" part in the primer name refers to the target gene number(s).

Table 1. List of primers that were used in this study. The code "gene" in the primer names in the text refer to the gene number of the specific gene as indicated in the table.

Primer name	Primer sequence
Lp_0965-F	5'-GACACGGCGGCCGCGTGTAAATTGAATTTGAAAGGACGC-3' *
Lp_0965-R	5'-CGTGTCCTCGAGGTTTAGGACCAGTGACCATCCG-3' *
Lp_0965-UF	5'-GCGGAATCAATTGTTGAATCCCC-3'
Lp_0965-UR	5'-GGATTGTTGAAAATTAGCCATGGTG-3'
Lp_0965-DF	5'-CCACTTAGTCAGTCGTAAGCACT-3'
Lp_0965-DR	5'-CAATAATCAAGTTCAAACGGGCG-3'
Lp_0965-IF	5'-CAAGCCCAGCTTGCTAATCAATTG-3'
Lp_0965-IR	5'-GTATGGTGAAACCTTTATGGCGAC-3'
Lp_1396-UF	5'-TCCATCGCCCACTTGACTTATTC-3'
Lp_1396-UR	5'-TCGTGTCGTTTCCATTAGATTTCCC-3'
Lp_1396-DF	5'-CTGGTCATCTTAAAGCAATTACACT-3'
Lp_1396-DR	5'-CTGTATTTGTCGACAAACATATCG-3'
Lp_1396-IF	5'-TGGTTATTTGGCAGTTAGCGTGT-3'
Lp_1396-IR	5'-ACTTGTTGACCATTCGCCTCAG-3'
Lp_3128-F	5'-GGACACGGCGGCCGCCCACTAAGTTATAATAACCATATCG-3' *
Lp_3128-R	5'-CGTGTCCTCGAGGGTTGATTTAACTAATATTTGTTGGG-3' *
Lp_3128-UF	5'-GTATGAGACTAAAGCGCTAG-3'
Lp_3128-UR	5'-GGTTTTCGTGTATTTCATTAG-3'
Lp_3128-DF	5'-TCGGCGCCGGAAATTGATGAA-3'
Lp_3128-DR	5'-CTGGCGAATATGGAAGTGAAACAC -3'
Lp_3128-IF	5'-CCAGACTACGACTTAGTTAAGACTA-3'
Lp-3128-IR	5'-GCTCGGGTATGCCGACAACG-3'
Lp_3679-lp_3676-UF	5'-GGCAAGGATTAATTGCCGGTATC-3'
Lp_3679-lp_3676-UR	5'-TCTTACCAACAATAACTCCCCCTT-3'
Lp_3679-lp_3676-DF	5'-ATACACCACAATCTTGATGAGTGGA-3'
Lp_3679-lp_3676-DR	5'-CCACATTGTTCTCCCTGTAAAGAATG-3'
Lp_3679-lp_3676-IF	5'-GCGGATTTTACATTCCATTAAGGG-3'
Lp_3679-lp_3676-IR	5'-CAGCTTGATTATCAAAACTTGCAGCC-3'
RB85	5'-GTTTTTTTCTAGTCCAAGCTCACA-3'
RB87	5'-GCCGACTGTACTTTCGGATCCT-3'
OSf	5'-GGTACATCATTCTGTTTGTGATGG-3'
OSr	5'-CTACGCTCAAGGGCTTTTACGC-5'

*The underlined sequences indicate restriction sites that were introduced for the cloning procedures.

Plasmids containing the 5'-flanking regions and plasmids containing the 3'-flanking regions from lp_1396 and lp_3128, were constructed following a similar strategy. For obtaining plasmids containing both the 5'- and the 3'- flanking region from these genes, the plasmids with the separate parts were digested with *Pvu*I and *Nco*I for the lp_1396

replacement construct, and with *Sph*I and *Nco*I for the lp_3128 replacement construct. The digestion mixtures were separated on agarose gel after which the fragments of the expected size were isolated and purified. The parts containing the 5'- and the 3'-flanking region were ligated. The ligation mixtures were purified and used for the transformation of *E. coli* DH5*α*. Transformants were checked for the presence of the plasmid of interest as previously described for the other replacement constructs. The vectors for replacement of the genes lp_0965, lp_1396, lp_3128, and lp_3679-lp_3676 were designated pNZ7580, pNZ7581, pNZ7582 and pNZ7583, respectively (Table 2)

Replacement of the target genes in *L. plantarum* WCFS1

For replacement of the target genes, *L. plantarum* WCFS1 was transformed with the corresponding replacement vector. The cells were plated on MRS with 10 µg/ml chloramphenicol. In order to select for strains in which plasmid integration had taken place, chloramphenicol resistant transformants were subsequently checked for erythromycin sensitivity to obtain candidate double cross-over gene-replacement mutants. The selected transformants were checked for the presence of the replacement fragment at the right location in the chromosome by PCR with the primer combinations gene-IF with RB85, and gene-IR with RB87 (Table 1, results not shown). The mutant strains were designated *L. plantarum* Δlp_0965, *L. plantarum* Δlp_1396, *L. plantarum* Δlp_3128, and *L. plantarum* Δlp_3679/6 respectively.

Construction of the overexpression vectors

The two overexpression vectors that were constructed were both based on pNZ5346 (J. Lambert, unpublished results; Fig. 2). This vector contains the promoter region of a DNA-helicase encoding gene of *L. plantarum* WCFS1 (lp_1144). Microarray data and promoter-reporter fusion studies in *L. plantarum* indicate that this promoter is responsible for constitutive expression at a medium expression level (J. Lambert, unpublished observations).

Target genes lp_0965 and lp_3128 were amplified by PCR together with their predicted ribosome binding sites and transcriptional terminators using *L. plantarum* WCFS1 chromosomal DNA as template and lp_0965-F and lp_0965-R or lp_3128-F and lp_3128-R as primer combinations, respectively (Table 1). The resulting amplicons were digested with *Xho*I and *Not*I (sites that were introduced in the primers) and ligated into similarly digested pNZ5346. The ligation mixtures were purified and used to transform *E. coli* DH5*α*. Chloramphenicol resistant colonies were checked for the presence of the target genes by PCR (primers used were gene-F and gene-R; Table 1; results not shown). The vector for the overexpression of lp_0965 and lp_3128 were designated pNZ7584 and pNZ7585 (Table 2).

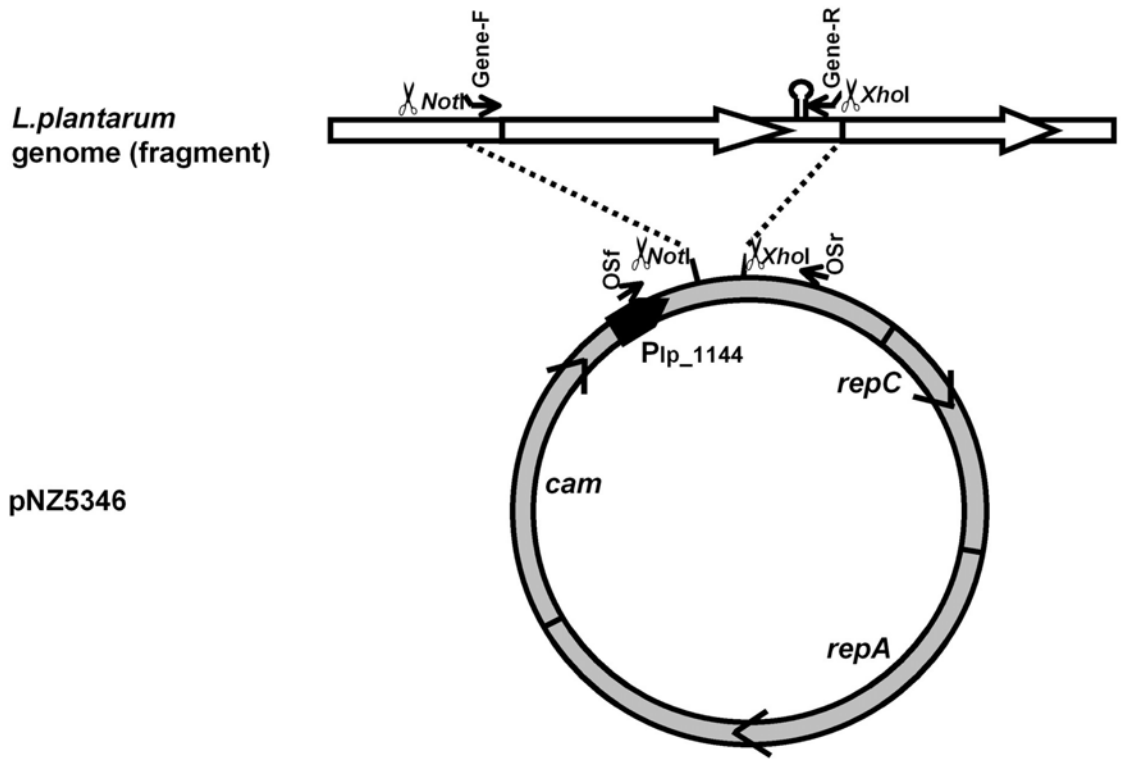

Figure 2. Schematic representation of the strategy applied for the construction of the pNZ5346 based overexpression plasmids.

In order to obtain *L. plantarum* overexpression derivatives of lp_0965 or lp_3128, *L. plantarum* WCFS1 was transformed with pNZ7584 or pNZ7585, respectively. Transformants were selected on 10 µg/ml chloramphenicol and verified by PCR using primer combinations gene-F with OSr or gene-R with Osf (Table 1, results not shown). The overexpression strains were designated *L. plantarum* oe_lp_0965 and *L. plantarum* oe_lp_3128.

Growth experiments

Growth experiments were performed in 0.25x MRS with various combinations of initial pH and lactate concentrations. Cultivations and optical density measurements (OD_{600}) were performed on a Bioscreen C (Thermo Labsystems, Breda, The Netherlands) in multiple well plates containing 200 µl of culture medium, at 37ºC under anoxic conditions. The wells were inoculated with 2 µl of a logarithmically growing culture. Maximal growth rates were determined from cells in the logarithmic growth phase.

Table 2. List of the strains and plasmids that were used in this study along with their relevant features and references.

Material	Relevant features	References
Strains		
E. coli DH5α	Cloning host	3
L. plantarum WCFS1	Wild-type for which the genome sequence is available	9
L. plantarum Hcat	Cm^R, *L. plantarum* WCFS1 containing cam marker in the intergenic region of lp_2681 and lp_2683	R. Bongers and I. I. van Swam, unpublished results
L. plantarum WCFS1::pNZ273	Cm^R, *L. plantarum* WCFS1 containing the vector from which pNZ5346 was derived	12
L. plantarum WCFS1 Δlp_0965	Cm^R, *L. plantarum* WCFS1 from which lp_0965 was removed	This work
L. plantarum WCFS1 Δlp_1396	Cm^R, *L. plantarum* WCFS1 from which lp_1396 was removed	This work
L. plantarum WCFS1 Δlp_3128	Cm^R, *L. plantarum* WCFS1 from which lp_3128 was removed	This work
L. plantarum WCFS1 Δlp_3679/6	Cm^R, *L. plantarum* WCFS1 from which lp_3679-lp_3676 were removed	This work
L. plantarum WCFS1 oe_lp_0965	Cm^R, *L. plantarum* WCFS1 containing pNZ7584	This work
L. plantarum WCFS1 oe_lp_3128	Cm^R, *L. plantarum* WCFS1 containing pNZ7585	This work
Plasmids		
pNZ5317	Cm^R, Em^R, vector for the construction *of L. plantarum* gene replacement mutants	J. Lambert, unpublished results
pNZ5346	Cm^R, vector for the construction of *L. plantarum* over expression strains by fusion of the gene behind the lp_1144 promoter	J. Lambert, unpublished results
pNZ7580	Cm^R, Em^R, lp_0965::cat replacement derivative of pNZ5317	This work
pNZ7581	Cm^R, Em^R, lp_1396::cat replacement derivative of pNZ5317	This work
pNZ7582	Cm^R, Em^R, lp_3128::cat replacement derivative of pNZ5317	This work
pNZ7583	Cm^R, Em^R, lp_3679-lp_3676::cat replacement derivative of pNZ5317	This work
pNZ7584	Cm^R, lp_0965 overexpression derivative of pNZ5346	This work
pNZ7585	Cm^R, lp_3128 overexpression derivative of pNZ5346	This work

Survival experiments

Overnight cultures of the *L. plantarum* strains depicted in Table 5 were cultured in MRS broth pH 6.5. When the cells had entered the log-phase (OD_{600}=1.3 – 1.5) 10 µl was transferred to 1 ml fresh MRS broth, pH 6.5. After 1 hour incubation at 37ºC, dilutions were plated for the determination of the viable cell counts, and 10 µl from the cell suspensions were transferred to 1 ml MRS, 50 mM sodium lactate, pH 2.7. The cells were incubated at 37ºC.

After 30 minutes and 60 minutes of incubation, dilutions were plated to determine the viable cell counts. Relative survival was expressed as the percentage of remaining viable cells as compared to the initial viable cell count.

Results

Construction of mutant strains

Knock-out and / or overexpression strains of the transcriptional regulators lp_0965 and lp_1396, the peroxide resistance protein dpr (lp_3128) and the cell surface protein cluster (lp_3679-lp_3676) were constructed. Growth and survival studies were performed in which *L. plantarum* Hcat functioned as the reference strain for the knock out mutants. This strain contains the chloramphenicol resistance cassette in the non-coding intergenic region of two genes (lp_2681 and lp_2683) that are transcribed in the opposite direction (R. Bongers and I. I. van Swam, unpublished results). The characteristics of the overexpression strains were compared with those of the *L. plantarum* WCFS1 harboring pNZ273, which is the vector from which all overexpression vectors used in this study were derived (12).

Effect of replacement or overexpression of genes from the lactic acid stimulon on growth rate and cell density under several combination of pH and lactate concentration

The effect of various combinations of initial pH and lactate concentrations on growth rate and maximal cell density were determined for the different mutant strains (Tables 3 and 4). Under all combinations tested, the knock out strains appeared to display a higher growth rate and reached a higher final cell density relative to the overexpressing strains. This trend was also observed for the control strains, indicating that this effect could not be attributable to the genes targeted by mutation.

No clear differences in either growth rate or maximal cell density were observed between the strains in which the target gene had been replaced and *L. plantarum* Hcat, the control strain. Similarly, the growth characteristics of the strains in which the target gene was

overexpressed lacked clear differences as compared to the control strain, *L. plantarum* WCFS1 harboring pNZ273 (Table 3).

Table 3. Effect of replacement or overexpression of lp_0965, lp_3128 and lp_3679-lp_3676 on growth rates of *L. plantarum* on media that varied in initial lactate concentration and pH value.

Strain	pH4.0 / 0mM Na-lactate	pH4.0 /50mM Na-lactate	pH4.5 / 0mM Na-lactate	pH4.5/ 150mM Na-lactate	pH5.0 / 0mM Na-lactate	pH5.0/ 150mM Na-lactate
L.plantarum Hcat	0.086 +/- 0.002	0.061 +/- 0.001	0.165 +/- 0.002	0.068 +/- 0.001	0.230 +/- 0.001	0.133 +/- 0.000
L.plantarum Δlp_0965	0.107 +/- 0.003	0.071 +/- 0.001	0.130 +/- 0.001	0.076 +/- 0.000	0.205 +/ 0.000	0.160 +/- 0.000
L.plantarum Δlp_3128	0.101 +/- 0.000	0.063 +/- 0.002	0.159 +/- 0.000	0.073 +/- 0.003	0.203 +/- 0.000	0.120 +/- 0.002
L.plantarum Δlp_3679/6	0.090 +/- 0.002	0.055 +/- 0.002	0.152 +/- 0.012	0.068 +/- 0.000	0.190 +/- 0.003	0.136 +/- 0.000
L.plantarum ::pNZ273	0.040 +/- 0.012	0.030 +/- 0.003	0.128 +/- 0.000	0.047 +/- 0.003	0.202 +/- 0.001	0.117 +/- 0.001
L.plantarum oe_lp_0965	0.057 +/- 0.004	0.033 +/- 0.002	0.118 +/- 0.000	0.033 +/- 0.000	0.185 +/- 0.001	0.103 +/- 0.001
L.plantarum oe_lp_3128	0.079 +/- 0.008	0.033 +/- 0.003	0.170 +/- 0.000	0.054 +/- 0.002	0.200 +/- 0.001	0.122 +/- 0.003

Table 4. Effect of replacement or overexpression of lp_0965, lp_3128 and lp_3679-lp_3676 on maximal OD_{600} values of *L. plantarum* on media that varied in initial lactate concentration and pH value.

Strain	pH4.0 / 0mM Na-lactate	pH4.0 /50mM Na-lactate	pH4.5 / 0mM Na-lactate	pH4.5/ 150mM Na-lactate	pH5.0 / 0mM Na-lactate	pH5.0/ 150mM Na-lactate
L.plantarum Hcat	1.66 +/- 0.25	1.67 +/- 0.20	1.68 +/- 0.15	1.49 +/- 0.03	1.63 +/- 0.00	1.74 +/- 0.09
L.plantarum Δlp_0965	1.50 +/- 0.08	1.70 +/- 0.19	1.80 +/- 0.06	1.28 +/- 0.35	1.65 +/- 0.25	1.77 +/- 0.02
L.plantarum Δlp_3128	1.39 +/- 0.03	1.37 +/- 0.01	1.70 +/- 0.00	1.49 +/- 0.00	1.67 +/- 0.01	1.71 +/- 0.03
L.plantarum Δlp_3679/6	1.51 +/- 0.01	1.35 +/- 0.05	1.73 +/- 0.08	1.52 +/- 0.04	1.58 +/- 0.01	1.68 +/- 0.03
L.plantarum ::pNZ273	1.21 +/- 0.00	0.86 +/- 0.012	1.54 +/- 0.09	1.17 +/- 0.08	1.60 +/- 0.02	1.65 +/- 0.00
L.plantarum oe_lp_0965	1.25 +/- 0.06	0.94 +/- 0.03	1.47 +/- 0.00	0.68 +/- 0.03	1.56 +/- 0.00	1.23 +/- 0.01
L.plantarum oe_lp_3128	1.25 +/- 0.22	0.87 +/- 0.09	1.66 +/- 0.05	1.24 +/- 0.00	1.69 +/- 0.07	1.69 +/- 0.07

Effect of replacement or overexpression of genes from the lactic acid stimulon on survival under a lethal combination of pH and lactate concentration

In order to analyze the potential contribution of the target genes to survival at extreme combinations of lactate and pH, the strains indicated in Table 5 were tested for their survival during incubation at a lethal combination of low pH and high lactate concentration. Logarithmically growing cells were transferred to MRS with 50mM sodium lactate at pH 2.7. After 30 and 60 minutes, the number of remaining viable counts was determined. The percentage of viable cells for the different strains is depicted in Table 5. No significant difference between the killing rates of any of the strains studied was observed.

A similar experiment was performed in which *L. plantarum* Δlp_1396 and *L. plantarum* Δlp_3679/6 were preincubated for 1 hour in 0.25x MRS supplemented with 300mM additional sodium lactate at a pH of 4.8. This pre-adaptation phase was followed by the exposure to the lethal combination of lactate and pH. Despite this pre-adaptation phase, no differences in survival were observed between the mutants and the control strains (results not shown).

Table 5. Effect of overexpression or replacement of lp_0965, lp_1396, lp_3128 and lp_3679-lp_3676 on the survival of *L. plantarum* on 0.25x MRS broth with 50mM sodium lactate at pH 2.7.

Strain	Survival after 30 minutes	Survival after 60 minutes
L. plantarum Hcat	79% +/- 8%	60% +/- 2%
L. plantarum Δlp_0965	78% +/- 0%	62% +/- 3%
L. plantarum Δlp_1396	84% +/- 7%	58% +/- 4%
L. plantarum Δlp_3128	84% +/- 1%	57% +/- 3%
L. plantarum Δlp_3679/6	88% +/- 6%	58% +/- 10%
L. plantarum::pNZ273	70% +/- 10%	48% +/- 6%
L. plantarum oe_lp_0965	80% +/- 1%	54% +/- 7%
L. plantarum oe_lp_3128	82% +/- 3%	61% +/- 3%

Discussion

In a previous study, a select group of genes of *L. plantarum* WCFS1 was identified that specifically showed increased expression in the presence of undissociated lactic acid (11; Chapter 4). Three genes and an operon from this group were selected for more detailed studies in order to validate their role in growth and survival of *L. plantarum* under conditions with high lactic acid concentrations.

The experiments performed, did not indicate a physiological role of the regulatory processes in which lp_0965 and lp_1396 are potentially involved. Since replacement of these genes did not result in decreased lactic acid sensitivity, it appears that these genes can not be designated as master regulators for lactic acid resistance. Nevertheless, a role of these regulators in modulation of gene expression of genes related to lactic acid stress can not be excluded at this stage. Additional (transcriptome) experiments under normal and stress conditions would be required to determine the genes regulated by the products of these genes. However, our experiments clearly establish that individual mutations in these regulator encoding genes do not have a major impact on lactic acid tolerance in *L. plantarum* under the conditions tested.

Previous microarray studies in combination with scanning electron microscopy implicated that the gene products from the lp_3679-lp_3676 operon contribute to lactic acid stress related changes in morphology (11; Chapter 4). Growth and survival experiments using a mutant derivative that lacks these cell-surface protein encoding genes, do not indicate that this operon plays an essential role in the counteracting of lactic acid stress itself: the survival of *L. plantarum* Δlp_3679/6 did not significantly differ from those of the control strain.

The experiments with the Lp_3128 knock out- and overexpression strains do not indicate a specific role for this gene in the counteracting of the negative effects of lactic acid on the cell. The Lp_3128 gene encodes a homologue of the *Streptococcus mutans* peroxide stress resistance protein Dpr (15). Remarkably, the microarray analyses of lactic acid stress response in *L. plantarum* indicated that the gene for the well known peroxide resistance protein catalase also showed increased expression upon lactic acid stress (11; Chapter 4). These results could imply that lactic acid stress evokes a cross adaptation towards peroxide stress. An overlap between the stimulon by lactic acid and that by peroxide stress has also been reported for *Lactococcus lactis* (8).

The microarray experiments in which the transcriptional response of *L. plantarum* towards lactic acid stress (11; Chapter 4) revealed that the genes that were under study in this chapter showed a specific response towards the undissociated form of the organic acid. The results from the current study indicate that the lactic acid stimulon contains genes that may not be directly involved in the counteracting of the negative effects of lactic acid itself. These genes possibly play a role in the anticipation to potential environmental changes under conditions in which the cell does not have the ability to rapidly adapt to environmental changes due to the current stress condition. Cross-adaptations, induced by the one stress factor, counteracting the other, have been observed in prokaryotes both at the level of transcription, protein production and physiological state (e.g. 1, 8, 10).

Furthermore, the adaptation towards lactic acid may require the combined action of multiple genes. In that case clear changes in lactic acid resistance may only be observed if a substantial part of the stimulon would be overexpressed or removed from the genome. The mutagenesis tool applied to construct has been designed to facilitate such multi-locus mutagenesis in a single genetic background. Therefore, additional experiments could focus on the combination of the individual mutations created here and evaluation of the lactic acid stress tolerance in the resulting strain.

References

1. **Arnold, C. N., J. McElhanon, A. Lee, R. Leonhart, and D. A. Siegele.** 2001. Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. J. Bacteriol. **183:**2178-2186.
2. **Axe, D. D., and J. E. Bailey.** 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol. Bioeng. **47:**8-19.
3. **Casadaban, M. J., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138:**179-207.
4. **Ferain, T., D. Garmyn, N. Bernard, P. Hols, and J. Delcour.** 1994. *Lactobacillus plantarum ldhL* gene:overexpression and deletion. J. Bacteriol. **176:**596-601.
5. **Giraud, E., B. Lelong, and M. Raimbault.** Influence of pH and initial lactate concentration on the growth of *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **36:**96-99.
6. **Hartke, A., S. Bouché, J. -C. Giard, A. Benachour, P. Boutibonnes, and Y. Auffray.** 1996. The lactic acid stress response of *Lactococcus lactis* subsp. *lactis*. Curr. Microbiol. **33:**194-199.
7. **Hutkins, R. W., and N. L. Nannen.** 1993. pH homeostasis in lactic acid bacteria. J. Dairy. Sci. **76:**2354-2365.
8. **Josson, K.,T. Scheirlinck, F. Michiels, C. Platteeuw, P. Stanssens, H. Joos, P. Dhaese, M. Zabeau, and J. Mahillon.** 1989. Characterization of a gram-positive broad-host-range plasmid isolated from *Lactobacillus hilgardii*. Plasmid **21:**9-20.
9. **Kleerebezem M, J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:**1990-1995.
10. **Lorca, G. L., and G. F. de Valdez.** 1999. The effect of suboptimal growth temperature and growth phase on resistance of lactobacillus acidophilus to environmental stress. Cryobiology. **39:**144-149.
11. **Pieterse, B., R. J. Leer, F. H. J. Schuren, and M. J. van der Werf.** 2005. Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling. Microbiol. **151:**3881-3894.
12. **Platteeuw, C., G. Simons, and W. M. de Vos.** 1994. Use of the *Escherichia coli* beta-glucuronidase (gusA) gene as a reporter gene for analysing promoters in lactic acid bacteria. Appl Environ. Microbiol. **60:**587-593.
13. **Russell, J. B., and F. Diez-Gonzalez.** 1998. The effects of fermentation acids on bacterial growth. Adv. Microbial Physiol. **39:**205-234.

14. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

15. **Yamamoto, Y., M. Higuchi, L. B. Poole, and Y. Kamio.** 2000. Role of the dpr product in oxygen tolerance in *Streptococcus mutans*. J. Bacteriol. **182:**3740-3747.

Chapter 6

Transcriptome analysis of the effect of continuous NaCl stress on gene expression of *Lactobacillus plantarum*

Bart Pieterse, Michiel Kleerebezem, and Mariët J. van der Werf

Manuscript in preparation.

Abstract

Whereas the prokaryotic response towards acute NaCl and other hyperosmotic stresses is extensively studied, relatively little information is available on the long term effects of hyperosmosis. For this reason the effect of continuous exposure to 0.3 M and 0.8 M NaCl on gene expression in *Lactobacillus plantarum* was studied. No increased expression was observed for genes that had previously been implicated to play a role in the acute response to hyperosmotic stress. However, clear trends were observed over multiple functional classes, which indicates that the long term effects of hyperosmotic stress requires more adaptations than restoration of the turgor alone. Most remarkable were the high number of differentially expressed genes that contain the recognition site *cre* for the carbon catabolite control protein CcpA in their promoter region. The consensus sequence was identified in the upstream region of over one third of the operons that showed decreased expression. Based on the results presented in this study, we hypothesize that their differential expression is possibly related to the lower water activity of the cytoplasm as a consequence of hyperosmosis.

Introduction

The microbial response towards osmotic stress is an important issue in industry. Inhibition of microbial growth by high osmolarity is in many cases a desired effect, e.g. in the food industry where osmotically active agents, such as salt and sugars, are used to suppress spoilage-related and pathogenic microbes (1, 32). In other cases, growth inhibition is an unwanted side-effect, for instance in food-products in which food-grade microorganisms fulfill a positive role and in fermentations where alkaline salts are added for pH control.

For many applications the long term adaptation towards high osmolarity is more relevant than the acute, short term adaptation, e.g. in processes in which the osmolarity of the environment increases gradually due to drying or pH control. In contrast, most studies on the prokaryotic adaptation towards osmotic stress focuss on the effect of osmotic upshock, i.e. a sudden increase in osmolarity of the environment. This type of osmotic shock is characterized by an initial outward flux of water. The consequential loose of turgor hampers cell elongation and thus cellular growth (9). A general response among prokaryotes is to restore this turgor by the uptake or biosynthesis of compatible solutes: osmolytes that accumulate at high concentrations inside the cell without seriously affecting cellular processes (10). The role of these compatible solutes has been extensively studied in multiple organisms (for reviews see: 9, 10, 32, 36, 39, 41, 51), among which *L. plantarum* (16-19, 22-24).

Upon prolonged incubation, compatible solutes also play a role in maintaining the turgor. However, the nature of the stress caused by the high osmolarity has changed: the turgor is restored to a level that allows for cell division, but the cells may still suffer from the high osmolarity due to the long term effects on the cytoplasmic composition (7).

The aim of this study was to obtain an overview on the effects of continuous osmotic stress on *L. plantarum*. For this purpose transcription profiles from *L. plantarum* cultured in the absence or in the presence of NaCl were compared by microarray analysis.

Materials and methods

Microorganism

Lactobacillus plantarum WCFS1, a single colony isolate from *L. plantarum* NCIMB8826, is maintained at NIZO food research in Ede, the Netherlands. The complete genome sequence from this organism is available (27; GenBank: AL935263).

Cultivation and cell harvesting

Glucose-limited continuous culturing of *L. plantarum* was performed on 25% MRS medium (13.75 g/l, Difco) at 37°C in an Applikon bioreactor with a working volume of 1 l (35). Glucose limitation was confirmed by the fact that the addition of extra glucose resulted in an increased biomass yield. The cultures were maintained anoxic by a continuous overlay with nitrogen. All fermentations were performed in duplicate. Either 0 M, 0.3 M or 0.8 M of additional NaCl was added to the culture medium. The pH was maintained at 6.0 by the automatic titration with 10 M sodium hydroxide. Samples were taken from the steady state cultures and immediately quenched in a -45°C, methanol-based buffer as previously described (34). Cell pellets were stored at -80°C until used.

RNA isolation

RNA was isolated from the cells according to the Macaloid / phenol based protocol as previously described (34). RNA purity and concentrations were determined both spectrophotometrically and on agarose gel. Purified RNA samples were checked for residual RNase activity by comparing samples that were incubated for 1 hour at 42°C with the untreated material on an agarose gel.

Array design

The microarray that was used is a clone based array, containing 3714 identified fragments from the genomic *L.plantarum* WCFS1 library in pBlueScript SK+ (27). The average size of the fragments is 1.2 kb with an average deviation of 0.3 kb. The microarray contains approximately 80% of the *L. plantarum* WCFS1 genome (34).

Fluorescent labelling and hybridization

Differential transcript levels were determined by two-color fluorescent hybridizations of the corresponding cDNA's on the *L. plantarum* clone-array. The RNA samples were labeled by random hexamers primed *in vitro* reversed transcription with either Cy5- or Cy3-labeled dUTP. Labeling, hybridization and washing were performed as described previously (34). The labelled samples from the duplicate steady state cultures with 0.3 M or 0.8 M of additional NaCl, respectively, were hybridised on microarrays with RNA from steady state cultures without additional NaCl as the reference. The fluorescent labels were swapped for the biological duplicates in order to avoid false positives due to dye-specific effects.

Image analysis

The fluorescent signals from the two different labels on the hybridized arrays were quantified with a ScanArray Express scanner (Packard Bioscience) and Imagene 4.2 software (BioDiscovery, Inc.). Spots from which the difference between the mean signal of the spot and the mean signal of the background was smaller than two times the background standard deviation were excluded from further analysis. Spots from which the signal in one or both of the channels exceeded the detection limit of the scanner were also excluded. After removal of the empty spots and the spots from which the signal exceeded the detection limit of the scanner, 3176 spots (86%) remained for further analysis.

Normalization

Within-slide, intensity dependent normalizations were performed with the scatter plot smoother LOWESS from the software program Datafit (Oakdale Engineering). The user-defined fraction of data used for smoothing at each point was set at 20% for all slides.

Significance analysis

The total data set on which significance analysis was performed contained 6 duplicate subsets (35), among which the duplicate sets on the effects of 0.3 M and 0.8 M NaCl.

Prior to the significance analysis, a data transformation was applied on the normalized ratios in order to obtain a normal distribution of the data (34). Significance analysis was performed by means of a 1-way-ANOVA. Subsequently, a Tukey HSD test was performed to determine whether a significant differential expression level (99% confidence interval) was observed under a specific condition. If genes or predicted operons were present on multiple spots on the array, these spots were considered as replicates in the significance analysis.

By comparison of overlapping regions of the genomic fragments on the array, it could be predicted which of the specific genes that were present on the clones is or are the ones affected. In the few instances in which this was not possible, these genes were omitted from the final dataset.

Classification of genes

Single genes were classified according to the annotated genome of *L. plantarum* WCFS1 (27). For genes that were grouped in predicted operons, the classification was made for the whole operon. In cases where the separate genes in an operon belonged to different

functional classes, the operon was either classified according to its predominant function, or grouped in multiple classes.

Identification of motifs in the promoter regions of differentially expressed genes

In order to identify conserved sequences in the promoter regions of differentially expressed genes the Hidden Markov Model based tool "MEME" (3) was applied. The non-coding regions upstream of the differentially expressed operons formed the input sequences for these searches. The total number of occurrences of this motif in all the non-coding regions of the *L. plantarum* genome was subsequently determined by a MAST search (4).

Nucleotide analysis

For nucleotide analysis, samples were taken and immediately quenched at -45°C methanol as described previously (34). The protein content of the samples was determined using a modification of the biuret method (47). The intracellular metabolites were extracted from the cell suspensions by chloroform extraction at –45°C as described by Ruijter and Visser (40). The sample was deproteinized by filtration using a Centrisart I, cut-off 10.000 filter (Sartorius, Goettingen, Germany) centrifuged at 3000 g and -20°C for 16 hours. Next, the samples were lyophilized and subsequently dissolved in 0.1 ml methanol/water (1/3 v/v).

Samples (25 µl) were separated on a reversed phase column (Chrompack Inertsil 5µm ODS-3 100*3 mm, Middelburg, The Netherlands) using a 40 min linear gradient from 5 mM hexylamine (pH 6.3) to 100 % of 90% methanol-10 mM ammonium acetate (pH 8.5) at a flow rate of 0.4 ml/min (8). Compounds were detected by electrospray ionization (ESI; negative ion mode) using a Finnigan LTQ linear ion trap mass spectrometer. During data acquisition, the mass spectrometer probe voltage was maintained at 3-4 kV, the heated capillary was kept at 250°C.

For the determination of the concentrations of the compounds analyzed, first the compounds were located in the full scan chromatogram by selecting unique ions from the mass spectrum of each compound and comparing the retention times and mass spectra obtained with those of the standards. Quantification of the metabolites was achieved by means of the peak area of the metabolite in the appropriate reconstructed ion chromatogram of a unique mass of the mass spectrum of that metabolite. The amount of metabolite in the sample was obtained by multiplying the thus obtained peak area with the response factor of that metabolite after correction for the biomass and recovery of the internal standards.

Results

Experimetal design

In order to study the long term effects of osmotic stress on *L. plantarum* cultivations were performed in steady state on complex 0.25x MRS broth in the absence or presence of additional sodium chloride. On this complex growth medium, the impact of the osmolyte concentration is limited as compared to chemically defined medium, which indicates that adaptation processes occur (see discussion). The presence of 0.3 M additional NaCl did not result in a significant decrease in growth rate on 0.25x MRS broth under batch conditions (results not shown). With 0.8 M additional NaCl the decrease in growth rate is 33%. Due to this limited decrease in growth rate large changes in gene expression as a consequence of altered relative growth rate can be excluded (35). Differential gene expression due to growth rate differences was circumvented by the use of steady state cultures with identical dilution rates.

Statistics

In order to guarantee reliability of the data on which the biological interpretations are based, genes and operons were selected that showed a significantly differential expression within the 99% confidence interval of the Tukey HSD test. In total 54 putative operons, containing 132 genes were identified that showed an increased expression in response to 0.8 M sodium chloride, and 97 tentative operons containing 206 genes that showed a decreased expression (99% confidence interval of the Tukey HSD test). Although, in general, the expression changes towards 0.3 M NaCl were limited, the majority of the genes and operons showed the same direction of regulation as they did towards 0.8 M NaCl: 77% for the operons with a decreased expression and 83% for the operons with an increased expression.

Differentially expressed genes per functional class

The long term effects of NaCl on gene expression appeared to cover a wide range of functional classes. Tables 1 and 2 give an overview of the genes that showed decreased (Table 1) or increased (Table 2) expression in response to 0.8 M NaCl. Clear trends were observed for several functional classes:

Amino acids biosynthesis**.** Decreased expression was observed for four operons that are involved in amino acid biosynthesis of the aspartate group, and two operons that are involved in aromatic amino acids biosynthesis. The only amino acid biosynthesis related operon that showed increased expression is involved in the biosynthesis of serine from 3P-glycerate.

Cell envelope. Clear trends were observed in the expression of genes related to (wall)teichoic acids. While the *dlt*-operon, responsible for lipoteichoic acid D-alanylation, showed a decreased expression, genes involved in wall teichoic acid biosynthesis showed an increased expression.

DNA metabolism. The majority of the genes involved in DNA metabolism that showed increased expression is involved in excision repair and repair of double stranded breaks.

Energy metabolism. A relatively large group of genes that were differentially expressed in response to the higher osmolarity, are involved in functions related to primary energy metabolism. A major part of these genes showed a decreased expression (32 operons vs. 5 operons that showed increased expression). From these operons, 10 also contained genes related to sugar transport (mostly phosphotransferase related). Furthermore, decreased expression was observed for genes involved in anaerobic respiration (nitrate / fumarate reduction). Several genes from the glycolysis and fermentation pathways also showed a lower expression at high osmolarity.

Nucleotide biosynthesis. While the expression level of the major operon involved in purine biosynthesis was clearly decreased, a modest but significant increase in expression was observed for the operon involved in pyrimidine biosynthesis.

Transport- and bindingproteins. Decreased expression was observed for 28 operons with involved in transport processes. For only 9 operons in this functional class, increased expression was observed. Half of the operons that showed a decreased expression level concerned PTS-systems or ABC-transporters involved in carbon source transport. The only PTS-systems that showed increased expression was annotated as a N-acetylglucosamine and glucose PTS. Furthermore three cation transporting ATPases, and a glycine betaine / carnitine / choline ABC transporter were downregulated. Genes for the biosynthesis and transport of the bacteriocin plantaricin showed an increased expression.

Table 1. Genes that showed a significant decrease in expression in response to 0.8 M NaCl. Changes are depicted as the expression ratio with the expression level from a culture in the absence of additional NaCl. A "C" in the last column indicates that the motif *cre* was identified in the upstream region of the operon.

Functional class	orf.-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
amino acid biosynthesis					
	lp_0571-lp_0572	homoserine dehydrogenase; homoserine kinase	0.9	**0.7**	
	lp_0822	glutamine-fructose-6-phosphate transaminase (isomerizing)	0.8	**0.6**	
	lp_1083-lp_1086	transketolase; shikimate 5-dehydrogenase; phospho-2-dehydro-3-deoxyheptonate aldolase / chorismate mutase; 3-dehydroquinate synthase	0.8	**0.5**	
	lp_2033-lp_2038	shikimate kinase; prephenate dehydrogenase; 3-phosphoshikimate 1-carboxyvinyltransferase; unknown; chorismate synthase; transport protein	0.8	**0.5**	
	lp_2570	aspartate-semialdehyde dehydrogenase	0.8	**0.5**	
	lp_2830	aspartate ammonia-lyase	0.6	**0.4**	
	lp_3085	asparagine synthase (glutamine-hydrolysing)	0.9	**0.4**	
cell envelope					
	lp_0197	cell surface protein precursor	0.9	**0.4**	
	lp_0472-lp_0473	integral membrane protein; lipoprotein precursor (putative)	0.7	**0.5**	
	lp_0800	cell surface protein	0.7	**0.6**	
	lp_1762-lp_1763	unknown; glycosyltransferase	1.0	**0.7**	
	lp_2016-lp_2021	D-alanyl transfer protein; D-alanyl carrier protein ; D-alanyl transfer protein; D-alanine activating enzyme; D-Ala-teichoic acid biosynthesis protein (putative); serine-type D-Ala-D-Ala carboxypeptidase	0.7	**0.5**	
	lp_2911	membrane-bound protease, CAAX family	1.1	**0.4**	
cellular processes					
	lp_0355	cell division protein SufI	0.8	**0.4**	
central intermediary metabolism					
	lp_0193	alpha-glucosidase	0.7	**0.2**	
	lp_2757	glucan 1,4-alpha-maltohydrolase	0.6	**0.3**	C
cofactor biosynthesis					
	lp_0113-lp_0116	hydroxyethylthiazole kinase; phosphomethylpyrimidine kinase; thiamine-phosphate pyrophosphorylase; purine-cytosine transport protein	0.6	**0.6**	C
	lp_1779	formate--tetrahydrofolate ligase	0.7	**0.5**	
energy metabolism					
	lp_0060-lp_0062	short-chain dehydrogenase/oxidoreductase; acetoacetate decarboxylase (putative); nitroreductase	0.8	**0.4**	
	lp_0184-lp_0185	fructokinase; sucrose PTS, EIIBCA	0.8	**0.4**	C
	lp_0187-lp_0189	beta-fructofuranosidase; sucrose operon repressor; alpha-glucosidase	0.7	**0.5**	C
	lp_0230-lp_0233	mannitol PTS, EIICB; transcription regulator; mannitol PTS, EIIA; mannitol-1-phosphate 5-dehydrogenase	0.9	**0.5**	C
	lp_0435-lp_0440	transcription regulator; cellobiose PTS, EIIC; unknown; cellobiose PTS, EIIC; 6-phospho-beta-glucosidase	1.1	**0.5**	C
	lp_0496-lp_0500	unknown; deoxyribose-phosphate aldolase; fucose transport protein; unknown; ribokinase	0.9	**0.3**	
	lp_0849	pyruvate oxidase	0.8	**0.5**	C
	lp_0952	fumarate reductase, flavoprotein subunit precursor	0.8	**0.5**	C
	lp_1069	NADH dehydrogenase	0.9	**0.5**	C
	lp_1101-lp_1102	L-lactate dehydrogenase; cation transport protein	0.7	**0.2**	C

Functional class	orf.-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
	lp_1105-lp_1109	malic enzyme, NAD-dependent; [citrate (pro-3S)-lyase] ligase; citrate lyase, acyl carrier protein; citrate lyase, beta chain	0.7	**0.2**	
	lp_1112-lp_1113	fumarate hydratase; fumarate reductase, flavoprotein subunit precursor,N-term truncated	0.5	**0.0**	C
	lp_1468-lp_1473	ABC transporter, ATP-binding protein; ABC transporter component (putative); cysteine desulfurase; nifU-like protein; ABC transporter component, iron regulated (putative); iron chelatin ABC transporter, substrate binding protein (putative)	0.8	**0.7**	C
	lp_1497-lp_1506	nitrate reductase complex: narG; narH; narJ; narI; unknown; unknown; integral membrane protein; transcription regulator (putative)	0.7	**0.6**	
	lp_1912	phosphoenolpyruvate synthetase	0.5	**0.2**	C
	lp_2095-lp_2097	transcription regulator of fructose operon; 1-phosphofructokinase; fructose PTS, EIIABC	1.2	**0.5**	
	lp_2151-lp_2154	pyruvate dehydrogenase complex:pdhD; pdhC; pdhB; pdhA	0.9	**0.2**	C
	lp_2502	glucose-6-phosphate isomerase	0.9	**0.6**	
	lp_3009-lp_3011	cellobiose PTS, EIIB; cellobiose PTS, EIIC; 6-phospho-beta-glucosidase	1.0	**0.6**	C
	lp_3054	aryl-alcohol dehydrogenase	0.9	**0.3**	C
	lp_3125	fumarate reductase, flavoprotein subunit precursor,N-term truncated	0.7	**0.6**	C
	lp_3170	phosphoglycerate mutase	1.1	**0.4**	
	lp_3207	Aminotransferase	**0.5**	**0.6**	
	lp_3418	phosphoenolpyruvate carboxykinase (ATP)	1.0	**0.2**	C
	lp_3449	NADH oxidase	**2.3**	**0.6**	
	lp_3480-lp_3482	UTP--hexose-1-phosphate uridylyltransferase; UDP-glucose 4-epimerase; galactokinase	0.9	**0.3**	
	lp_3526-lp_3530	6-phospho-beta-glucosidase; beta-glucosides PTS, EIIBCA; transcription antiterminator; maltose phosphorylase	1.1	**0.3**	
	lp_3538-lp_3539	transketolase; transaldolase	1.0	**0.7**	
	lp_3545-lp_3548	L-iditol 2-dehydrogenase; galacitol PTS, EIIC; galactitol PTS, EIIB; galacitol PTS, EIIA	0.4	**0.2**	C
	lp_3554-lp_3557	unknown; maltose O-acetyltransferase; L-arabinose isomerase; L-ribulose 5-phosphate 4-epimerase; L-ribulokinase (putative); arabinose transport protein	1.1	**0.6**	C
	lp_3560-lp_3567	unknown; lipase/esterase (putative); lipase/esterase (putative); transport protein; Na(+)/H(+) antiporter; unknown; glucokinase	0.7	**0.2**	C
	lp_3662	bifunctional protein: alcohol dehydrogenase; acetaldehyde dehydrogenase	0.6	**0.0**	
fatty acid and phospholipid metabolism					
	lp_0370-lp_0372	glycerol kinase;glycerol-3-phosphate dehydrogenase; glycerol uptake facilitator protein	0.5	**0.2**	
	lp_2643	lipoate-protein ligase	0.9	**0.3**	C
	lp_3504-lp_3506	integral membrane protein (putative); acetylesterase; transcription regulator	0.9	**0.6**	C
protein synthesis					
	lp_0501	serine--tRNA ligase	1.0	**0.5**	
	lp_0578	non-ribosomal peptide synthetase	**3.2**	0.6	
purines, pyrimidines, nucleosides, and nucleotides					
	lp_2719-lp_2729	purine biosynthesis	0.8	**0.2**	
	lp_3270-lp_3271	adenylosuccinate synthase; GMP reductase	1.3	**0.2**	
	lp_3334	adenine deaminase	0.8	**0.2**	
	lp_3480-lp_3482	UTP--hexose-1-phosphate uridylyltransferase; UDP-glucose 4-epimerase; galactokinase	0.9	**0.3**	C

Functio-nal class	orf.-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
regulatory functions					
	lp_3646	transcription regulator	1.0	**0.6**	
transport and binding protein					
	lp_3009-lp_3011	cellobiose PTS, EIIB pts32B; cellobiose PTS, EIIC pts32C; 6-phospho-beta-glucosidase pbg5	1.0	**0.6**	C
	lp_0184-lp_0185	fructokinase sacK1; sucrose PTS, EIIBCA pts1BCA	0.8	**0.4**	C
	lp_0217-lp_0219	ABC transporter, permease protein; ABC transporter, ATP-binding protein; ABC transporter subunit (putative)	0.6	**0.4**	
	lp_0230-lp_0233	mannitol PTS, EIICB mtlA; transcription regulator, mannitol operon mtlR; mannitol PTS, EIIA mtlF; mannitol-1-phosphate 5-dehydrogenase; mtlD	0.9	**0.5**	C
	lp_0286	cellobiose PTS, EIIC pts6C	1.0	**0.3**	C
	lp_0367-lp_0369	glycine betaine/carnitine/choline ABC transporter, substrate binding and permease protein opuB; glycine betaine/carnitine/choline ABC transporter, ATP-binding protein opuB; glutathione reductase gshR0	0.9	**0.6**	
	lp_0435-lp_0440	transcription regulator; cellobiose PTS, EIIC pts7C; unknown; cellobiose PTS, EIIC pts8C; 6-phospho-beta-glucosidase pbg0	1.1	**0.5**	C
	lp_0496-lp_0500	unknown; deoxyribose-phosphate aldolase deoC; fucose transport protein; unknown; ribokinase rbsK0	0.9	**0.3**	
	lp_0567	cation transporting P-type ATPase	0.7	**0.5**	
	lp_0575-lp_0577	mannose PTS, EIIAB pts9AB; mannose PTS, EIIC pts9C; mannose PTS, EIID pts9D	1.3	**0.3**	
	lp_0729	transport protein	0.9	**0.6**	
	lp_0783	lipoprotein precursor, peptide binding protein OppA homolog	1.4	**0.4**	
	lp_0802-lp_0803	glutamine ABC transporter, substrate binding and permease protein glnPH1; glutamine ABC transporter, ATP-binding protein glnQ0	0.6	**0.4**	
	lp_0848	transport protein	1.2	**0.3**	
	lp_1095-lp_1097	manganese ABC transporter, ATP-binding protein mtsC; manganese ABC transporter, permease protein mtsB; manganese ABC transporter, substrate binding protein mtsA	1.1	**0.3**	
	lp_1164-lp_1165	cellobiose PTS, EIIC pts14C; cell surface hydrolase (putative)	0.7	**0.5**	
	lp_1261-lp_1265	oligopeptide ABC transporter, substrate binding protein oppA; oligopeptide ABC transporter, permease protein oppB; oligopeptide ABC transporter, permease protein oppC; oligopeptide ABC transporter, ATP-binding protein oppD; oligopeptide ABC transporter, ATP-binding protein oppF	0.7	**0.5**	
	lp_1468-lp_1473	ABC transporter, ATP-binding protein; ABC transporter component (putative); cysteine desulfurase csd1; nifU-like protein; ABC transporter component, iron regulated (putative); iron chelatin ABC transporter, substrate binding protein (putative) fecB	0.8	**0.7**	C
	lp_2075-lp_2077	transcription regulator; nitrate ABC transporter, ATP-binding protein; nitrate ABC transporter, permease protein	1.1	**0.6**	
	lp_2095-lp_2097	transcription regulator of fructose operon fruR; 1-phosphofructokinase fruK; fructose PTS, EIIABC pts16ABC	1.2	**0.5**	
	lp_2969	N-acetylglucosamine PTS, EIICBA pts22CBA	0.8	**0.3**	C
	lp_3049	amino acid transport protein	0.8	**0.5**	
	lp_3327	cadmium-/zinc-/cobalt- transporting ATPase	**0.6**	**0.6**	C
	lp_3486	lactose transport protein lacS1	0.6	**0.2**	C
	lp_3526-lp_3530	6-phospho-beta-glucosidase pbg10; beta-glucosides PTS, EIIBCA pts33BCA; transcription antiterminator; maltose phosphorylase map3	1.1	**0.3**	
	lp_3545-lp_3548	L-iditol 2-dehydrogenase gutB; galacitol PTS, EIIC pts35C; galactitol PTS, EIIB pts35B; galacitol PTS, EIIA pts35A	0.4	**0.2**	C
	lp_3554-lp_3557	unknown; maltose O-acetyltransferase maa; L-arabinose isomerase araA; L-ribulose 5-phosphate 4-epimerase araD; L-ribulokinase (putative) araB; arabinose transport protein araT	1.1	**0.6**	C
	lp_3560-lp_3567	unknown; lipase/esterase (putative); lipase/esterase (putative); transport protein; Na(+)/H(+) antiporter; unknown; glucokinase	0.7	**0.2**	C

Functio-nal class	orf.-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
hypothetical proteins					
	lp_0063	Unknown	0.7	**0.2**	C
	lp_0152-lp_0158	transcription regulator; unknown; unknown; unknown; inknown	0.8	**0.2**	
	lp_0758	protein containing diguanylate cyclase/phosphodiesterase domain 2 (EAL)	0.9	**0.6**	
	lp_1168	Unknown	0.9	**0.2**	C
	lp_2290	integral membrane protein	0.8	**0.8**	
	lp_2732-lp_2733	oxidoreductase; oxidoreductase	1.4	**0.6**	C
	lp_2787	hydrolase, HAD superfamily, Cof family	0.7	**0.3**	C
	lp_3359-lp_3360	integral membrane protein; integral membrane protein	**0.5**	**0.4**	
	lp_3489	Oxidoreductase	0.8	**0.4**	C
	lp_3500-lp_3501	short-chain dehydrogenase/oxidoreductase; unknown	0.9	**0.6**	C
other					
	lp_1766-lp_1767	unknown; lysine	**1.5**	0.8	

Effect of continuous NaCl stress on the expression of genes that have previously been implicated to play a role in adaptation osmotic stress in lactic acid bacteria

The sequencing of the *L. plantarum* genome revealed the presence of several genes involved in the transport of glycine betaine, carnitine and / or choline (lp_0367-lp_0369, lp_1607-lp_1610, and lp_3324). Whereas lp_1607-lp_1610 and lp_3324 showed no differential expression under the tested conditions (results not shown), lp_0367-lp_0369 showed a significant decrease in expression (Table 1). Also for the three annotated potassium uptake proteins encoding genes (lp_0525, lp_2173, and lp_3279), no differential expression was observed (results not shown).

Other stress proteins that have been associated with osmotic stress response in lactic acid bacteria are GroES (lp_0727), GroEL (lp_0728), DnaK (lp_2027) (25, 37) and HtrA (lp_0043) (15, 50). For none of the corresponding genes in *L. plantarum* differential expression was observed (results not shown). Moreover, increased bacteriocin production was reported for *Lactococcus lactis* when it was exposed to hyperosmotic conditions in the absence of osmoprotectant (glycine betaine) in the growth medium (48). A modest increase in expression of the plantaricin production and transport genes (lp_0423-lp_0431) was observed when *L. plantarum* was grown at high osmolarity (Table 2).

Table 2. Genes that showed a significant increase in expression in response to 0.8 M NaCl. The list is sorted by functional class of the operon. Changes are depicted as the expression ratio with the expression level from a culture in the absence of additional NaCl. A "C" in the last column indicates that the motif cre was identified in the upstream region of the operon.

Functional class	Orf-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
amino acid biosynthesis					
	lp_0202-lp_0204	acetyltransferase (putative); phosphoglycerate dehydrogenase; phosphoserine aminotransferase	1.3	**1.9**	
cell envelope					
	lp_0267-lp_0269	glycerol-3-phosphate cytidylyltransferase; teichoic acid biosynthesis protein; teichoic acid biosynthesis protein	2.1	**2.2**	
	Lp_0510	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	2.1	**2.4**	C
	lp_1816-lp_1819	D-ribitol-5-phosphate cytidylyltransferase (putative); ribitol-5-phosphate 2-dehydrogenase (putative); teichoic acid biosynthesis protein; teichoic acid biosynthesis protein	1.3	**2.0**	
	lp_2843-lp_2844	poly(glycerol-phosphate) alpha-glucosyltransferase; poly(glycerol-phosphate) alpha-glucosyltransferase	0.9	**2.3**	
cellular processes					
	lp_0423-lp_0431	plantaricin biosynthesis and transporter: plnG; plnH; plnS; plnT; plnU; plnV;plnW; plnX; plnY	1.2	**1.7**	
	Lp_3128	Dpr-like protein	2.5	**3.1**	
	Lp_3578	Catalase	1.9	**3.6**	
central intermediary metabolism					
	lp_0027-lp_0028	beta-phosphoglucomutase; maltose phosphorylase	1.0	**3.1**	
cofactor biosynthesis					
	lp_1599-lp_1603	bifunctional protein: methylenetetrahydrofolate dehydrogenase; methenyltetrahydrofolate cyclohydrolase; exodeoxyribonuclease VII, large subunit; exodeoxyribonuclease VII, small subunit; geranyltranstransferase; hemolysin homolog	1.7	**2.1**	
DNA metabolism					
	Lp_0432	DNA helicase (putative)	1.2	**1.6**	
	lp_0698-p_0701	DNA-directed DNA polymerase III, subunit; unknown; recombination protein RecR; unknown	1.2	**1.8**	
	lp_0771-lp_0773	HD superfamily hydrolase (putative); excinuclease ABC, subunit B and A	1.1	**2.1**	
	lp_1548-l_1549	exonuclease SbcC; exonuclease SbcD	1.5	**1.7**	
	lp_1599-lp_1603	bifunctional protein: methylenetetrahydrofolate dehydrogenase / methenyltetrahydrofolate cyclohydrolase; exodeoxyribonuclease VII, large subunit; exodeoxyribonuclease VII, subunit; geranyltranstransferase; hemolysin homolog	1.7	**2.1**	
	lp_1839-lp_1840	topoisomerase IV, subunits A and B	**1.6**	**1.8**	
	Lp_2109	excinuclease ABC, subunit C	1.0	**1.6**	

Functio-nal class	Orf-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
	lp_2168-lp_2170	exodeoxyribonuclease V, alpha chain; unknown; putative phosphoglycerate mutase	1.8	**3.2**	
energy metabolism					
	lp_0841-lp_0843	exopolyphosphatase; polyphosphate kinase; exopolyphosphatase	1.4	**1.8**	
	lp_0900-lp_0903	phosphoglycerate mutase; phosphoglycerate mutase; Na(+)/H(+) antiporter; 7,8-dihydro-8-oxoguanine-triphosphatase (putative)	1.5	**1.7**	C
	lp_1409-lp_1410	amino acid efflux protein; aminotransferase	0.9	**1.8**	
	Lp_2751	aminotransferase	1.2	**2.0**	
	lp_2872-lp_2874	oxidoreductase; alcohol dehydrogenase; short-chain dehydrogenase/oxidoreductase	1.6	**1.8**	
fatty acid and phospholipid metabolism					
	Lp_3174	cyclopropane-fatty-acyl-phospholipid synthase	1.3	**1.5**	
	Lp_3536	choloylglycine hydrolase	4.3	**5.9**	
protein fate					
	lp_0959-lp_0960	dipeptidase pepD3; unknown	1.8	**2.8**	
protein synthesis					
	lp_3119-lp_3120	transcription regulator; GTP-binding translation elongation factor LepA	1.5	**3.9**	C
purines, pyrimidines, nucleosides, and nucleotides					
	lp_0692-lp_0693	ribonucleoside-diphosphate reductase, beta chain; ribonucleoside-diphosphate reductase, alpha chain	1.6	**2.4**	
	lp_0779-lp_0781	nucleotide kinase (putative); unknown; unknown	0.9	**1.7**	
	lp_2696-lp_2703	pyrimidine biosynthesis cluster	0.8	**1.6**	
regulatory functions					
	lp_0036-lp_0037	response regulator; histidine protein kinase; sensor protein	1.3	**2.0**	C
	Lp_0965	transcription regulator	0.7	**1.9**	
	Lp_1396	transcription regulator	1.0	**1.9**	
	Lp_3097	transcription regulator (putative)	1.4	**2.3**	
	lp_3190-lp_3191	histidine protein kinase, sensor protein; response regulator	1.0	**1.7**	
transport and binding protein					
	Lp_0295	transport protein	1.2	**2.2**	
	lp_0423-lp_0431	plantaricin biosynthesis and transporter: plnG; plnH; plnS; plnT; plnU; plnV;plnW; plnX; plnY	1.2	**1.7**	
	lp_0892-lp_0893	transcription regulator (putative); multidrug transport protein	1.1	**1.5**	
	lp_1409-lp_1410	amino acid efflux protein; aminotransferase arcT	0.9	**1.8**	

Functio-nal class	Orf-nos.	Assigned function	0.3 M	0.8 M	*cre* (C)
	Lp_1921	transport protein	0.9	**1.6**	
	Lp_2525	ABC transporter, ATP-binding protein	0.8	**1.9**	
	Lp_2531	N-acetylglucosamine and glucose PTS, EIICBA	1.1	**2.1**	C
	lp_3209-lp_3211	amino acid ABC transporter, substrate binding protein; amino acid ABC transporter, permease protein; amino acid ABC transporter, ATP-binding protein	1.5	**1.8**	
	Lp_3281	transport protein	0.9	**1.5**	
hypothetical proteins					
	lp_0081-lp_0082	intracellular prote60ase/amidase (putative); oxidoreductase	1.3	**2.5**	
	Lp_0793	Unknown	0.9	**1.5**	
	lp_1406-lp_1407	putative D-alanyl carrier protein; putative aminotransferase	1.5	**2.1**	
	Lp_1833	Unknown	1.1	**2.1**	
	Lp_1856	Unknown	1.8	**4.2**	
	lp_1898-lp_1903	pseudouridylate synthase; transcription regulator (putative); unknown; unknown; integrase/recombinase; unknown	1.5	**1.9**	
	Lp_1908	integral membrane protein	0.8	**3.0**	
	Lp_1995	Unknown	2.4	**3.9**	
	lp_2002-lp_2004	hypothetical protein; transcription regulator (putative); integral membrane protein	3.8	**9.0**	
	Lp_2580	alkaline phosphatase superfamily protein	1.0	**1.5**	C
	Lp_2793	Unknown	1.0	**1.4**	
	lp_3404-lp_3405	integral membrane protein; integral membrane protein	1.5	**1.6**	
	Lp_3411	extracellular protein	1.7	**2.8**	
	Lp_3429	hypothetical protein	2.0	**3.4**	C
Other					
	lp_1787-lp_1789	chloramphenicol O-acetyltransferase; unknown; unknown	1.1	**2.3**	
	Lp_3377	prophage P3 protein 12	1.8	**2.5**	
	lp_3382-lp_3385	prophage P3 protein 8, DNA primase/helicase; prophage P3 protein 7, DNA replication; prophage P3 protein 6; prophage P3 protein 4	1.5	**3.0**	
	lp_3389-lp_3392	prophage P3 protein 2, CI-like repressor; prophage P3 protein 1, integrase; unknown; transport protein	1.9	**2.6**	
	Lp_3496	Transposase	1.6	**2.6**	

Conserved promoter regions

In silico analysis of the upstream region of the operons by means of the MEME software tool (3) revealed the frequent presence of a highly conserved sequence of 15 nucleotides that resembles the *cre* motif as identified in front of the *L. plantarum* CcpA gene (Table 3; 31). Sequences homologous (p-value < E-05) to this consensus sequence were identified in the upstream regions of 35 of the operons that showed decreased expression

(Table 1) and 7 of the operons that showed increased expression (Table 2) in the presence of 0.8 M NaCl.

A subsequent MAST search (4) with the multilevel consensus on all the intergenic regions of *L. plantarum* revealed the presence of this motif in 129 intergenic regions.

Table 3. The *cre* motif identified in the promoter regions of 42 differentially expressed operons, and its comparison with the *cre* motif that has previously been identified in the upstream region of the *L. plantarum ccpA* gene (31).

cre motif upstream of the *L. plantarum ccpA* gene (31)	T G A A A G C G A T T T C A
Multilevel *cre* consensus identified in this study	A T G A A A C C G C T T T C A
	T A T T A A A
	A G G
	T

Effect of continuous NaCl stress on adenine nucleotide, NAD(H) and NADP(H) levels

In order to obtain a view on the energy status of the cell under continuous NaCl stress the AMP, ADP, ATP, NAD(H), and NADP(H) levels were determined (Table 4a). Although no significant effect on the energy charge could be observed (Table 4b), the levels of all adenine nucleotides were decreased in the samples from the cells that were cultured in the presence of 0.8 M NaCl. The NAD(H) and NADP(H) levels were also significantly lower in the cells that were cultured in the presence of 0.8 M NaCl.

Table 4. Effect of 0.8 NaCl on NAD(P)H and adenine nucleotide levels.

Metabolite	No additional NaCl	0.8 M additional NaCl
AMP	539 +/- 70 nmol / g protein	228 +/- 16 nmol / g protein
ADP	935 +/- 176 nmol / g protein	378 +/- 5 nmol / g protein
ATP	479 +/- 230 nmol / g protein	98 +/- 1 nmol / g protein
NAD^+	121 +/- 22 nmol / g protein	25 +/- 4 nmol / g protein
NADH	209 +/- 42 nmol / g protein	63 +/- 0 nmol / g protein
$NADP^+$	629 +/- 102 nmol / g protein	208 +/- 11 nmol / g protein
NADPH	140 +/- 26 nmol / g protein	29 +/- 5 nmol / g protein
Energy charge	0.47 +/- 0.08	0.41 +/- 0.01

Discussion

Currently, most studies on prokaryotic response towards hyperosmotic stress have concentrated on the restoration of the turgor immediately after osmotic upshock by means of the accumulation of compatible solutes. For the lactic acid bacteria, the role of compatible

solute accumulation has been studied most thoroughly in *Lactococcus lactis* (30, 44-46) and *L. plantarum* (16-19, 22-24). Although the long term response towards hyperosmotic stress is also relevant for lactic acid bacteria in their natural habitat and with respect to their applications in food industry, limited information is available on the effect of this type of hyperosmotic stress. In order to obtain a global view of the processes that are involved in the long term response to hyperosmolarity of *L. plantarum*, transcriptome analysis was performed.

An important issue in the experimental design of transcriptome studies on the prokaryotic stress response is the level of stress that is imposed on the organism: if the stress is too severe, it will have a large impact on the absolute and / or relative growth rate of the organism, which will lead to secondary effects in gene expression (35). However, if the stress is too mild, the changes in gene expression may be too small to distinguish. In the work presented here, conditions were applied that have limited effect on the growth characteristics of the organism (33% decrease in growth rate under batch conditions as an effect of 0.8 M NaCl). Meanwhile, two factors indicate physiological adaptations by the cell to the high salt concentration: Most obvious is the fact that cell division is taking place which indicates that adaptation of the cytosol has occured in order to restore turgor (9). Moreover, previous experiments with *L. plantarum* have shown that the presence of 0.6 M NaCl causes a decrease in growth rate of 75% on chemically defined medium, whereas this decrease was only 20% on MRS broth (23, 24). The limited decrease in growth rate on MRS broth is presumably caused by the uptake of compatible solutes from the MRS broth. It has previously been shown that yeast extract and beef extract, both components of MRS broth (11), contain betaine and carnitine respectively (14, 23).

From the typical genes that are involved in turgor maintenance in prokaryotes, none showed increased expression. In several prokaryotes, uptake of potassium forms the initial response upon osmotic upshock in order to restore turgor (2). Subsequently, this role is usually taken over by compatible solute transporters. In *L. plantarum*, the uptake of potassium does not seem to play a significant role in the adaptation to osmotic stress (18, 19). Therefore the unchanged expression of these transporters is not unexpected. Although compatible solute transporters play a central role in adaptation to osmotic stress, the only compatible solute transpoter encoding operon from which expression was affected (lp_1607-lp_1610) showed a decreased expression in response to sodium chloride. This corresponds to previous observations by Glaasker *et al*. who reported that increased osmolality of the growth medium did not lead to increased expression of a compatible solute transporter from *L. plantarum* (16). Moreover, they observed feed-back inhibition of the compatible solute transporters

activity occurred when the cells contained large amounts of compatible solutes. The decreased expression of the glycine betaine / carnitine / choline transporter (lp_0367-lp_0369; Table 1) shows that inhibition of compatible solute transport also takes place on the transcription level.

Also for other typical stress related genes that have been associated with osmotic upshock in lactic acid bacteria like *dnaK*, *groES* and *groEL* (25, 37), and *htrA* (15, 50), no increased expression was observed. This may indicate that these genes are only involved in the acute stress response and their expression levels are normalised upon prolonged stress exposure.

A clear trend could be observed in the expression of genes involved in cell wall biosynthesis. A shift was observed from genes involved in lipoteichoic acid biosynthesis to genes involved in wall teichoic acid biosynthesis. Decreased expression of genes involved in lipoteichoic acid biosynthesis has also been observed for *Bacillus subtilis* subjected to long term salt stress (42). High salt concentrations have been suggested to inhibit D-alanine incorporation in lipoteichoic acid in growing *Staphylococcus aureus* by inhibition at the enzyme level. Instead, the D-alanine is incorporated in teichoic acids (28). The data presented in this paper indicate that, for *L. plantarum*, this shift from lipotheichoic acid to teichoic acid in response to salt stress is regulated at the transcription level.

Increased expression of a coherent set of genes involved in DNA repair (Table 2) suggest an important role for these repair processes under high salt conditions. The excinuclease ABC complex, encoded by lp_0773, lp_0772, and lp_2109, is responsible for excision repair, while recombination protein RecR (lp_0700), exonuclease SbcC and SbcD (lp_1548 and lp_1549), and two subunits of the exodeoxyribonuclease VII (lp_1600 and lp_1601), are involved in the repair of double strand breaks. This corresponds to findings by Kultz and Chakravarty who demonstrated that hyperosmolality by increased NaCl concentrations leads to the increased occurrence of DNA damage in the form of double strand breaks (29).

In silico analysis of the upstream regions of the differentially expressed genes revealed the frequent occurrence of the conserved regulatory sequence *cre*, especially upstream of genes involved in transport processes and energy metabolism (Table 1 and 2). The motif was identified in the upstream regions of over one third of the operons that showed decreased expression, which indicates that these findings are not coincidental.

The c*re* sequence forms the binding site for CcpA, the master regulator for carbon catabolite control in various Gram positive bacteria, which is mostly involved in transcription

repression (49). Binding of CcpA to *cre* is enhanced by the phosphotranferase HPr in its seryl-phosphorylated form. The serine residue is phosphorylated by ATP in the presence of the preferred carbon source (in many cases glucose) (12, 38). Instead, the histidyl-phosphorylated form of HPr, is required for the transport of multiple alternative carbon sources via the PTS system. Hence the ratio between seryl- and histidyl phosphorylated HPr may indirectly reflect the metabolic and energetic status of the cell (43). Since, in this study, both the stressed and the unstressed steady state cultures were glucose limited, differential availability of the preferred carbon source could not explain the decreased expression of the CcpA controlled genes.

One might argue that the decreased transcription of CcpA repressed genes indicates a more favourable energetic status of the NaCl stressed cells, resulting in a higher ratio of the seryl- versus the histidyl phosphorylated form of HPr. Experiments with *Escherichia coli* have shown that osmotic stress leads to an increase of the [ATP]/[ADP] ratio in the cell (21, 33). However, this was not the case in *L. plantarum*: no significant effect of the 0.8 M NaCl on the energy charge was observed (Table 4). In contrast, the absolute levels of AMP, ADP and ATP were reduced at high NaCl concentrations which implicates a less preferable energetic status. This decrease in absolute levels of adenine nucleotides corresponds with the lower expression of genes involved in purine biosynthesis (lp_2719-lp_2729). Limitation of the nucleotide levels may serve a role in counteracting the negative effects of high osmolarity on the intracellular water contents.

Observations in *B. subtilis* indicated that the combination of seryl phosphorylated HPr with increased levels of NADP(H), either in its reduced or oxidised form could also account for increased gene repression by CcpA (26). In *L. plantarum* the total amount of NADP(H) was actually lower in the NaCl-stressed cells (Table 4). It should be noted that the absolute NAD(H) levels and the NAD^+/NADH ratio also showed a significant reduction under continuous NaCl stress, which may have a negative effect on the metabolic potential of the cells.

Apparently, under the conditions tested, CcpA mediated gene repression is not triggered by differential glucose levels or the energetic status of the stressed cells. This seems to diverge CcpA from its role in adapting gene expression to the available carbon sources and the energetic needs of the cell. The question is raised whether the observed expression changes of CcpA controlled genes are a truly beneficial for the cell under the current conditions, or an artifact related to changes in the intracellular milieu. Even though the cells are capable of cell division, the high external osmolarity may affect the amount of free available water in the cell (7). The consequential increased concentration of specific

metabolites could enhance binding of CcpA, as has been reported for early glycolytic intermediates (21).

Interestingly, whereas this study is, to our knowledge, the first to relate CcpA mediated gene expression to hyperosmotic stress, increased expression of the master regulator CcpA itself in response to elevated NaCl concentrations has been reported before: in a screen for active promoters in *L. plantarum* during growth on plates with 0.8 M NaCl, the *ccpA* promoter was identified (5). Moreover, CcpA was identified in a proteome study on the effect of osmotic shock on *Listeria monocytogenes* (13).

By focussing on the effect of high NaCl concentrations on gene expression in *L. plantarum* in a steady state culture, the results of this study give insight on the long term effects of this stress factor. The wide range of functional classes over which the genes are distributed that were differentially expressed in response to high NaCl concentrations indicates that the long term effects of hyperosmotic stress requires more adaptations than restoration of the turgor alone. Based on the decreased expression of CcpA mediated genes and the lower amount of adenine nucleotides, we hypothesize that part of these adaptations are required to adapt to the lower amount of free available water in the cytoplasm.

Acknowledgements

We thank Roelie Bijl for extracting the nucleotide samples, Richard Bas for LC-MS analysis and Leon Coulier and Karin Overkamp for quantification of the nucleotide levels.

Literature

1. **Abee, T., and J. A. Wouters.** 1999. Microbial stress response in minimal processing. Int. J. Food Microbiol. **50:**65-91.
2. **Bakker, E.P.** 1992. Cell K^+ and K^+ transport systems in prokaryotes. In: Alkali cation transport systems in prokaryotes. Bakker, E.P. (ed.) Boca Raton, FL:CRC Press pp. 205-224.
3. **Bailey T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. pp. 28-36, AAAI Press, Menlo Park, California.
4. **Bailey T. L. and M. Gribskov.** 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics, **14:**48-54.
5. **Bron, P. A. , S. M. Hoffer, I. I. van Swam, W. M. de Vos, and M. Kleerebezem.** 2004. Selection and characterization of conditionally active promoters in *Lactobacillus plantarum*, using alanine racemase as a promoter probe. Appl. Environ. Microbiol. **70:**310-317.
6. **Catte, M., F. Gancel, F. Dzierszinski, and R. Tailliez.** 1999. Effects of water activity, NaCl, and smoke concentrations on the growth of *Lactobacillus plantarum* ATCC 12315. Int. J. Food Microbiol. **52:**105-108.
7. **Cayley, S. and M. T. Record Jr.** 2004. Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate protein-DNA interactions and growth rate in osmotically stressed *Escherichia coli* K-12. J. Mol. Recognit. **17:**488-496.
8. **Coulier, L., R. C. Bas, M. J. van der Werf, and T. Hankemeier.** 2005. Microbial metabolomics: quantitative determination of metabolites in micro-organisms using ion-pair chromatography-electrospray ionization mass spectrometry. Submitted for publication.
9. **Csonka, L. N.** 1989. Physiological and genetic responses of bacteria to osmotic stress. Microbiol. Rev. **53:**121-147.
10. **Csonka, L. N., and A. D. Hanson.** 1991. Prokaryotic osmoregulation: Genetics and physiology. Annu. Rev. Microbiol. **45:**596-606.
11. **De Man, J. C., M. Rogosa, and M. E. Sharpe.** 1960. A medium for the cultivation of lactobacilli. J. Appl. Bacteriol. **23:**130-135.

12. **Deutscher, J., E. Kűster, U. Bergstedt, V. Charrier, and W. Hillen.** 1995. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to catabolite repression in gram-positive bacteria. Mol. Microbiol. **15:**1049-1053.
13. **Duché, O., F. Trémoulet, P. Glaser, and J. Labadie.** 2002. Salt stress proteins induced in *Listeria monocytogenes*. Appl. Environ. Microbiol. **68:**1491-1498.
14. **Dulaney, E. L., D. D. Dulaney, and E. L. Rickes.** 1968. Factor in yeast extract which relieve growth inhibition of bacteria in defined medium of high osmolarity. Dev. Ind. Microbiol. **163:**260-269.
15. **Foucaud-Scheunemann C., and I. Poquet.** 2003. HtrA is a key factor in the response to specific stress conditions in *Lactococcus lactis*. FEMS Microbiol Lett. **224:**53-59.
16. **Glaasker, E.**, **E. H. M. L. Heuberger, W. N. Konings, and B. Poolman.** 1998. Mechanism of osmotic activation of the quaternary ammonium compound transporter (QacT) of *Lactobacillus plantarum*. J. Bacteriol. **180:**5540-5546.
17. **Glaasker, E., W. N. Konings, and B. Poolman.** 1996. Glycine betaine fluxes in *Lactobacillus plantarum* during osmostasis and hyper- and hypo-osmotic shock. J. Biol. Chem. **271:**10060-10065.
18. **Glaasker, E., W. N. Konings, and B. Poolman.** 1996. Osmotic regulation of intracellular solute pools in *Lactobacillus plantarum*. J. Bacteriol. **178:**575-582.
19. **Glaasker, E., F. S. B. Tjan, P. F. Ter Steeg, W. N. Konings, and B. Poolman.** 1998. Physiological response of *Lactobacillus plantarum* to salt and nonelectrolyte stress. J. Bacteriol. **180:**4718-4723.
20. **Hsieh, L. –S., J. , Rouviere-Yaniv, and K. Drlica.** 1991. Bacterial DNA supercoilig and [ATP]/[ADP] ratio: changes associated with salt shock. J. Bacteriol. **173:**3914-3917.
21. **Gösseringer, R., E. Küster, A. Galinier, J. Deutscher, and W. Hillen.** 1997. Cooperative and non-cooperative DNA binding modes of catabolite control protein CcpA from *Bacillus megaterium* results from sensing two different signals. J. Mol. Biol. **266:**665-676.
22. **Kets, E. P. W., and J. A. M. de Bont.** 1997. Effect of carnitines on *Lactobacillus plantarum* subjected to osmotic stress. FEMS Microbiol. Lett. **146:**205-209.
23. **Kets, E. P. W., E. A. Galinski, and J. A. M. de Bont.** 1994. Carnitine: a novel compatible solute in *Lactobacillus plantarum*. Arch. Microbiol. **162:**243-248.
24. **Kets, E. P. W., M. Nierop Groot, E. A. Galinski, and J. A. M. de Bont.** 1997. Choline and acetylcholine: novel cationic osmolytes in *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **48:**94-98.

25. **Kilstrup, M., S. Jacobsen, K. Hammer, and F. K. Vogensen.** 1997. Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in *Lactococcus lactis*. Appl. Environ. Microbiol. **63:**1826-1837.
26. **Kim, J.-H, M. I. Voskuil, and G. H. Chambliss.** 1998. NADP, corepressor for the *Bacillus* catabolite control protein CcpA. Proc. Natl. Acad. Sci. **95:**9590-9595.
27. **Kleerebezem M., J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:**1990-1995.
28. **Koch, H. U., R. Döker, and W. Fischer.** 1985. Maintenance of D-alanine ester substitution of lipoteichoic acid by reesterification in *Staphylococcus aureus*. J. Bacteriol. **164:**1211-1217.
29. **Kultz, D., and D. Chakravarty.** 2001. Hyperosmolality in the form of elevated NaCl but not urea causes DNA damage in murine kidney cells. Proc. Natl. Acad. Sci. **98:**1999-2004.
30. **Molenaar, D., A. Hagting, H. Alkema, A. J. M. Driessen, and W. N. Konings.** 1993. Characteristics and osmoregulatory roles of uptakee systems for proline and glycine betaine in *Lactococcus lactis*. J. Bacteriol. **175:**5438-5444.
31. **Muscariello, L., R. Marasco, M. De Felice, and M. Sacco.** 2001. The functional *ccpA* gene is required for carbon catabolite repression in *L. plantarum*. Appl. Environ. Microbiol. **67:**2903-2907.
32. **O'Byrne, C. P., and I. R. Booth.** 2002. Osmoregulation and its importance to food-borne microorganisms. Int. J. Food Microbiol. **74:**203-216.
33. **Ohwada, T., and S. Sagisaka.** 1987. An immediate and steep increase in ATP concentration in response to reduced turgor pressure in *Escherichia coli* B. Arch. Microbiol. Biophys. **259:**157-163.
34. **Pieterse, B., R. H. Jellema, and M. J. van der Werf.** 2006. Quenching of microbial samples for increased reliablility of microarray data. J. Microbiol. Meth. **64:**207-216.
35. **Pieterse, B, R. J. Leer, F. H. J. Schuren, and M. J. van der Werf.** 2005 Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling. Microbiol. 151: 3881-3894.
36. **Poolman, B., and E. Glaasker.** 1998. Regulation of compatible solute accumulation in bacteria. Mol. Micrbiol. **29:**397-407.
37. **Prasad, J., P. McJarrow, and P. Gopal.** 2003. Heat and osmotic stress responses of probiotic *Lactobacillus rhamnosus* HN001 (DR20) in relation to viability after drying. Appl. Environ. Microbiol. **69:**917-925.

38. **Reizer, J., C. Hoischen, F. Titgemeyer, C. Rivolta, R. Rabus, J. Stűlke, D. Karamata, M. H. Saier Jr., and W. Hillen.** 1998. A novel protein kinase that controls carbon catabolite repression in bacteria. Mol. Microbiol. **27:**1157-1169.

39. **Roeβler, M., and V. Műller.** 2001. Osmoadaptation in bacteria and archaea: common principles and differences. Env. Microbiol. **3:**743-754.

40. **Ruijter, G.J.G. and J. Visser.** 1996. Determination of intermediary metabolites in *Aspergillus niger*. J. Microbiol. Meth. **25:**295-302.

41. **Sleator, R. D., and C. Hill.** 2001. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. FEMS Microbiol. Rev. **26:**49-71.

42. **Steil, L., T. Hoffmann, I. Budde, U. Völker, and E. Bremer.** 2003. Genome-wide transcriptional profiling analysis of adaptation of *Bacillus subtilis* to high salinity. J. Bacteriol. **185:**6358-6370.

43. **Titgemeyer, F., and W. Hillen.** 2002. Global control of sugar metabolism: a gram-positive solution. Antonie van Leeuwenhoek. **82:**59-71.

44. **van Der Heide, T., and B. Poolman.** 2000. Glycine betaine transport in *Lactococcus lactis* is osmotically regulated at the level of expression and translocation activity. J. Bacteriol. **182:**203-206.

45. **van Der Heide, T, and B. Poolman.** 2000. Osmoregulated ABC-transport system of *Lactococcus lactis* senses water stress via changes in the physical state of the membrane. Proc. Natl. Acad. Sci. **97:**7102-7106.

46. **Van Der Heide, T, M. C. Stuart, and B. Poolman.** 2001. On the osmotic signal and osmosensing mechanism of an ABC transport system for glycine betaine. EMBO J. **20:**7022-7032.

47. **Verduyn, C., E. Postma, , W. A. Scheffers. and J.P. van Dijken.** 1990. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. J. Gen. Microbiol. **136:**395-403.

48. **Uguen P., J. Hamelin, J. P. Le Pennec, and C. Blanco.** 1999. Influence of osmolarity and the presence of an osmoprotectant on *Lactococcus lactis* growth and bacteriocin production. Appl. Environ. Microbiol. **65:**291-293.

49. **Warner, J. B., and J. S. Lolkema.** 2003. CcpA-dependent carbon catabolite repression in bacteria. Microbiol. Mol. Biol. Rev. **67:**475-490.

50. **Wonderling, L. D., B. J. Wilkinson, and D. O. Bayles.** 2004. The htrA (degP) gene of *Listeria monocytogenes* 10403S is essential for optimal growth under stress conditions. Appl. Environ. Microbiol. **70:**1935-1943.

51. **Wood, J. M.** 1999. Osmosensing by bacteria: Signals and membrane-based sensors. Microbiol. Mol. Biol. Rev. **63:**230-262.

Chapter 7

Summary and concluding remarks

Introduction

The increasing number of complete genomes that have recently become available give a wealth of information on the potential of the concerning organisms. Nevertheless, this potential only becomes effective if the specific genes are expressed into a phenotype. The first step in this process is controlled at the level of gene transcription into RNA. Many cellular processes are regulated at this level. This implies that the transcriptome of a cell, which is the total set of RNA under a specific condition, contains valuable information on the biological state of the cell and the genes that play a role under specific environmental conditions. Transcriptomics is the integrated research approach that quantifies genome-wide expression changes in response to one or multiple conditions and extracts relevant information from these data by means of uni- and multivariate data analysis approaches.

In this thesis, transcriptomics approaches were applied to study the stress response of *Lactobacillus plantarum*. This wide-spread lactic acid bacterium plays a role in plant fermentations, dairy fermentations and the production of wine (20). Moreover, probiotic properties have been ascribed to various *L. plantarum* strains (9, 14). Recently, the complete genomic sequence of *L. plantarum* WCFS1 has become available (12). The DNA microarray we applied in this thesis contained selected genomic fragments that were amplified from the *L. plantarum* WCFS1 genome library that was constructed for genome sequencing.

In addition to the studies on gene expression in *L. plantarum*, special attention was paid to several steps in the transcriptomics experiment that influence the reliability of the generated data, and on the design of clone-based microarrays.

Towards relevant and reliable transcriptomics data

One of the strengths of microarray analysis is that there is no need of making a knowledge-based selection of biomolecules, since one simply aims at measuring "everything". Due to the large number of generated data and the unbiased approach, extra care should be taken to minimize the number of false positive observations that would complicate data interpretation and target selection.

In this thesis, special attention was paid to factors that might influence data reliability at several stages in the transcriptomics experiment. In addition, the role of experimental design on data interpretation was studied. The implications and applications of these findings are relevant for transcriptome studies in a broad range of microorganisms.

Validation of a cell harvesting method for increased reliability of microarray data

In many transcriptomics experiments published nowadays, little attention is paid to the effect that the harvesting method has on the quality and representativity of the RNA. Literature on microbial RNA half lives and the rapid transcriptional response towards changes in the environment indicate that potential changes introduced to the transcriptome during the harvesting of microbial cells include both degradation and *de novo* transcription. Significant changes in mRNA levels can already be introduced on a minutes time-scale (5; 21; 22; 23).

In order to overcome this problem, a -45°C methanol-based quenching method was validated for obtaining reliable and reproducible 'snapshot' samples of *Lactobacillus plantarum* cells for transcriptome analyses (Chapter 2). To this end the transcriptomes of cells that were harvested from a single steady state culture according to three different protocols were compared by microarray analysis. The time between harvesting and total arrest of cellular activity for the three methods was a few seconds for the quenching protocol and approximately three and ten minutes for the other two methods.

The results revealed that the quenching method improved the conservation of the transcriptome, and consequently the reliability and representativity of the gene expression data: principal component analysis grouped the replicate transcriptome datasets according to the methods by which the cells were harvested from the steady state culture. The order in which the samples from the three harvesting methods were grouped in the score plot corresponded to the time between harvesting and total arrest of cellular activity, which is a strong indication that the transcript profiles from the quenched samples are closest related to the actual biological condition of interest. Moreover, the expression changes of specific genes indicated the induction of a stress response in the non-quenched cells after the cells were taken from the culture.

Apart from these observations, the results indicated a lower variation between the quenched samples. This implies that harvesting and concentrating the cells under quenched conditions results in a higher reproducibility, which is highly beneficial since this will lead to an increased number of genes that pass the desired significance threshold. Consequently, the number of significant data that are available for the biological interpretation will increase.

Based on other studies in which this protocol was applied for the harvesting of both prokaryotes and eukaryotes, we expect that this method is applicable for a broad range of microorganisms.

Experimental design; Mathematical design of prokaryotic clone based microarrays

Whole transcriptome comparisons by means of microarray analyses have proven to be an effective tool for studying genomewide expression changes. Currently, the vast majority of the transcriptome studies is limited to organisms for which the complete annotated sequence is available. This limitation is inherent to the conventional microarrays in which each spot represents a single gene.

Clone-based microarrays on which each spot contains DNA from a random genome library are a good alternative for the conventional ORF-based arrays, especially for unsequenced strains (19). The construction of a microarray from a random genome library involves the drawback that genes are represented on the microarray by chance. Moreover, a spot on the array may represent multiple genes which could introduce uncertainties on the identity of the gene that is responsible for a differential signal from that spot.

The equations that are generally applied for the prediction of the genome coverage by a genomic library tend to overestimate the required number of clones for a library for hybridization purposes (7; 13), or do not give insight into the fraction of genes for which specific data can be generated in a transcriptomics experiment (3).

For this reason, two complementary equations were developed that allow for the prediction of the number of represented genes on a prokaryotic clone-based microarray for which interpretable results could be generated (Chapter 3). The first equation, the Minimal Insert Coverage (MIC)-equation, predicts the fraction of genes that are represented on the array and cover at least a set part (the minimal insert coverage) of the genomic fragment by which they are represented. The second equation, the Gene Specific Information (GSI)-equation, predicts the fraction of genes that are represented by spots on the array that only represent genes from a single transcription unit, which information can be interpreted in a quantitative way. The two equations will allow for determination of the required insert size of the clones and the amount of spots on the array to obtain an array with the desired genome coverage.

Experimental design; Effect of experimental design on data interpretation and specificity

Whereas some conditions will affect the cell and its transcription profile in a specific way, others will evoke a more complex response that is related to different aspects of this condition. In the latter case, the data will be less specific which will complicate interpretation and target selection. This was illustrated by unraveling the multiple effects of lactic acid on gene expression in *L. plantarum* (Chapter 4). Our main interest was in the effect of the undissociated form of this molecule. Meanwhile this organic acid also influences the cell by

its effects on pH, water activity and (relative) growth rate. Especially the (secondary) effects of the changes in (specific) growth rate on the expression profiles appeared to be massive (Chapter 4). In most experiments where the effect of a stress factor on gene expression is studied, the effect of altered growth rate is ignored. In other cases, absolute growth rate-related effects are avoided by the application of steady state cultures. Our results showed that both absolute and relative growth rate should be taken into account.

The specificity with which changes can be attributed to one certain aspect of a stimulus (e.g. the undissociated form of the lactic acid molecule) will generally increase by pinpointing genes that show similar expression trends under conditions that mimic other aspects of this stimulus (e.g. mimicking the pH effect of lactic acid by hydrogen chloride; Chapter 4). An alternative strategy to increase data specificity is by the selection of genes that show similar expression trends over different conditions that share the feature in which one is interested (e.g. the selection of genes that showed increased expression in response to lactic acid regardless of differences in (relative) growth rate; Chapter 4).

Statistical analyses

Statistical tools serve different, complementary roles in the analysis of transcriptomics datasets: Significance tests are mainly used to judge whether the observation for a single variable (gene) falls within the desired confidence interval. Multivariate methods are suitable for the identification of relations in the expression of different genes over the tested conditions and for the transcription profile-based correlation between different situations. The applicability of the various statistical tools on microarray datasets depends on the experimental design of the experiment and the nature of the input data for the statistical testing.

The microarray approach that was applied in this thesis relied on the co-hybridization of the sample from the conditions of interest with the control sample on the same array. This method has the advantage that the data can be normalized directly to the co-hybridized sample. The normalized ratio between the signals from the two samples, indicates whether a gene shows differential expression under a certain condition. A disadvantage of the ratio approach is that the data are not normally distributed and hence they will not be suitable as input for most statistical tools. For this reason, we transformed the data according to the equations described in Chapter 2. ANOVA and a Tukey HSD test were applied to determine the significance of the observations per spot on the array. Tests that determine a general cut off value for the information from all the spots were avoided since these would lead to increased amounts of both false positive and false negative observations (unpublished observations). This is due to the fact that the amount of technical variation is highly spot

dependent. Most obvious is the effect of the background signal on the variation in spots with a low amount of total signal.

A major part of the information on gene expression and regulation from the pre-omics era is based on developing and testing of new hypotheses that were founded on existing knowledge. Although this existing knowledge is still relevant for the interpretation of functional genomics data, it serves a less important role during earlier stages of the experiment. This is caused by the fact that functional genomics experiments do not select on a limited knowledge-based set of biomolecules, but measure 'everything'. Transcriptomics approaches allow for the mathematical correlation of large sets of genes or biological conditions by means of multivariate statistical tools. The strength of this approach was demonstrated by the unbiased grouping of the biological samples that were harvested according to different protocols by principal component analysis (Chapter 2), and the grouping of genes that showed similar expression trends over multiple conditions by hierarchical clustering (Chapter 4). In both examples mathematical clustering of the correlated variables clearly resulted in coherent groups that allowed for a new and unbiased view of the effect of the environmental conditions on the transcriptome.

Experimental validation of the leads

Significance testing can indicate the probability that a certain observation is a true positive. In the context of a transcriptomics experiment, in which relative expression changes are studied, "true positive" would mean that the expression level of the gene is truly affected by the condition of interest.

In cases where conclusions are to be drawn from the combined differential expression of multiple genes, a proper statistical analysis will usually suffice. This is especially the case if there is coherence in the group of genes that are differentially expressed, like a common biological function (Chapter 4) or the presence of a similar regulatory element in the promoter region of co-expressed genes (Chapter 6). If one wishes to draw definite conclusions from the differential expression of a single transcription unit, experimental validation of the microarray data is desirable.

Validation can be performed at two levels. The differential expression level itself can be validated with alternative techniques such as quantitative RT-PCR. At the level of functionality, physiological testing of knock-out or overexpression mutants is still the method of choice for the validation of the relevance of a differentially expressed gene under a certain condition. It should be noted that differential expression under a certain condition does not necessarily mean that a gene is specifically relevant for that condition. This was illustrated by the experiments with the strains that were mutated in genes from the lactic acid stimulon that

showed no detectable phenotype under the conditions tested (Chapter 6). Moreover, the absence of differential expression of a specific gene does not imply that this gene does not play a role that is specifically essential under the tested condition (e.g. compatible solute transport related genes under osmotic stress; Chapter 6).

Transcriptome analysis of the effects of continuous lactic acid and NaCl imposed osmotic stress on *L. plantarum*

Transcriptomics approaches were applied for the analysis of the effect of two stress factors on gene expression in *L. plantarum*: lactic acid stress and NaCl stress. Whereas the majority of studies on the prokaryotic stress response focus on the initial response after exposure of the organism to the stress condition, we concentrated on the long term effects by constant exposure in steady state cultures. The results from these studies confirmed the strength of transcriptomics for obtaining information on the effect of a condition on the cell. Moreover, by comparing the expression pattern from cells that were cultured under various conditions, highly specific lists of differentially expressed genes could be generated.

Effect of continuous lactic acid stress on gene expression in *L. plantarum*

Lactate is the predominant fermentation product of most lactobacilli and accumulates in high concentrations in the environment of the cells. At the end of the glycolysis, pyruvate is converted to lactate by lactate dehydrogenase, during which NADH is oxidized to NAD^+. This NADH regeneration is essential for the continuation of many reduction processes in the cell. Nevertheless, lactate formation also imposes a burden to the cells (10; 17). The inhibitory effect is mainly attributed to the undissociated lactic acid form of the molecule. This uncharged form easily diffuses from the environment into the cell. Inside the cell, a part of the lactic acid dissociates due to the higher intracellular pH. Different theories have been postulated to explain the inhibitory effects of this process on the cell: (i) dissipation of the membrane potential due to an uncoupling effect by the lactic acid (4), (ii) acidification of the cytosol (18), or (iii) intracellular accumulation of the lactate anions (16).

In order to study the effect of lactic acid on gene expression, an experimental design was developed for unraveling the multiple effects caused by the dissociated and undissociated form of the molecule (Chapter 4). Transcription profiles of *L. plantarum* grown in steady state cultures that varied in lactate / lactic acid ratio, pH, osmolarity and absolute and relative growth rate, were compared by microarray analysis.

The transcriptional response of *L. plantarum* towards lactic acid stress (Chapter 4) was found to differ completely from what may have been expected from other studies on the prokaryotic response towards (organic) acid stress. No differential expression was observed for genes that have previously been associated with pH control and the generation of a membrane potential, such as the F_1F_0-ATPase (24), amino acid decarboxylases (1; 11) and genes involved in the malolactic fermentation (15) .

Among the group of genes that showed increased expression upon lactic acid stress independent of the (relative) growth rate, 2 transcriptional regulators, 3 cell surface protein encoding operons, and genes for which homologues have been associated with stress responses in other micro-organisms were found. This indicates that the lactic acid stimulon of *L. plantarum* contains genes that are involved in the counteracting of multiple stresses. The three cell surface protein clusters showed high similarity with respect to the pI values, sizes and location of predicted transmembrane regions of the subsequent proteins. No specific function has previously been designated to these clusters.

From these genes, two regulators, the peroxide resistance protein *dpr*, and one of cell surface protein clusters, were selected for further research (Chapter 5). Strains were constructed in which the genes were either inactivated by a replacement strategy or overexpressed. No significant effects of these mutations on growth rate and survival under conditions with high lactic acid concentrations were observed. These observations could imply that the adaptation of *L. plantarum* to lactic acid stress, consist of the altered expression of a large number of genes, each responsible for a limited part of the entire response. The specific role of a major part of the lactic acid affected genes is still to be tested.

A second group of genes showed altered expression in response to the growth rate diminishing effect of the lactic acid (Chapter 4). Within this group, multiple genes were identified that suggest a redirection towards three alternative pathways of the central intermediary metabolism: the phosphoketolase bypass, a pathway towards malate, and a pathway towards ethanol. These routes would be beneficial with respect to limitation of the pyruvate and lactate levels and regeneration of NAD^+. These results suggest that diminished NAD^+ regeneration due to end-product inhibition of the lactate dehydrogenase is an important growth reducing effect of the lactic acid accumulation under anaerobic conditions. A possible strategy to avoid NAD^+ depletion in industrial fermentations with *L. plantarum* may be to avoid complete anaerobiosis so that regeneration can be performed by NADH oxidases for which several genes have been identified on the *L. plantarum* genome.

Effect of NaCl stress on gene expression in *L. plantarum*

The bacterial response towards osmotic stress is especially an issue in the food industry, where food-grade as well as spoilage-related and pathogenic bacteria are exposed to matrices with low water activity (2). A sudden increase of the extracellular osmolarity can result in an outward flux of water. The consequential loss of turgor hampers cell elongation and thus cellular growth (6). It has been demonstrated for multiple organisms that the uptake or biosynthesis of compatible solutes can form a successful strategy for the restoration of the turgor. Nevertheless, on the long term the cells may still suffer from the high osmolarity by the negative effects of the altered cytoplasm composition and its consequences for cellular processes (8).

The long term effect of NaCl imposed osmotic stress on gene expression in *L. plantarum* was studied by the comparison of transcription profiles from cells cultured in a steady state culture either in the absence or presence of additional NaCl (Chapter 6). No increased expression was observed for genes that have typically been related to acute osmotic stress in lactic acid bacteria, such as the genes coding for compatible solute transporters, GroES, GroEL and HtrA. Some of the trends that were observed suggested adaptation of the cell wall composition and increased requirement of genes involved in DNA repair. Most strikingly was the high number of differentially expressed operons that contained the carbon responsive element *cre*, the binding site for carbon control protein CcpA, in their upstream region. These observations could not be related to differential glucose levels between the cultures or the energy status of the cells. Therefore, it was suggested that the differential expression of these genes is to be attributed to a decrease in free cytoplasmic water and the resulting increased concentrations of metabolites that influence CcpA mediated gene expression.

The results presented in this thesis contribute to the reliability and applicability of transcriptomics research of microorganisms. Moreover, the power of this research approach for the characterization of the effects of environmental conditions on the cell was illustrated by the studies on the effects of NaCl and lactic acid stress on gene expression in *L. plantarum*. Genes that specifically responded to the tested stress conditions may serve as potential targets for the construction of strains with improved properties for industrial applications.

Literature

1. **Azcarate-Peril, M. A., E. Altermann, R. L. Hoover-Fitzula, Cano, and T. R. Klaenhammer.** 2004. Identification and inactivation of genetic loci involved with *Lactobacillus acidophilus* acid tolerance. Appl. Environ. Microbiol. **70:**5315-22.
2. **Abee, T., and J. A. Wouters.** 1999. Microbial stress response in minimal processing. Int. J. Food Microbiol. **50:**65-91.
3. **Akopyants, N. S., S. W. Clifton, J. Martin, D. Pape, T. Wylie, L. Li, J. C. Kissinger, D. S. Roos, and S. M. Beverley.** 2001. A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. Mol. Biochem. Parasitol. **113:**337-340.
4. **Axe, D. D., and J. E. Bailey.** 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol. Bioeng. **47:**8-19.
5. **Bernstein, J.A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen.** 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. **99:**9697-9702.
6. **Csonka, L. N.** 1989. Physiological and genetic responses of bacteria to osmotic stress. Microbiol. Rev. **53:**121-147.
7. **Clark,L., and J. Carbon.** 1976. A colony bank containing synthetic *Col*E1 hybrids representative of the entire *E. coli* genome. Cell. **9:**91-99.
8. **Cayley, S. and M. T. Record Jr.** 2004. Large changes in cytoplasmic biopolymer concentration with osmolality indicate that macromolecular crowding may regulate protein-DNA interactions and growth rate in osmotically stressed *Escherichia coli* K-12. J. Mol. Recognit. **17:**488-496.
9. **De Vries, M. C, E. E. Vaughan, M. Kleerebezem and, W. M. de Vos.** 2005. *Lactobacillus plantarum* - Survival, Functional and Potential Probiotic Properties in the Human Intestinal Tract. Int. Dairy J. accepted for publication.
10. **Giraud, E., B. Lelong, and M. Raimbault.** 1991. Influence of pH and initial lactate concentration on the growth of *Lactobacillus plantarum*. Appl. Microbiol. Biotechnol. **36:**96-99.
11. **Higuchi, T., H. Hayashi, and K. Abe.** 1997. Exchange of glutamate and gamma-aminobutyrate in a *Lactobacillus* strain. J. Bacteriol. **179:** 3362-3364.
12. **Kleerebezem, M., J. Boekhorst ,R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer *et al.*** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Nat. Acad. Sci. **100:**1990-1995.

13. **Lander, E.S., and M. S. Waterman.** 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics **2:** 231-239.
14. **Molin, G.** 2001. Probiotics in foods not containing milk or milk constituents, with special reference to *Lactobacillus plantarum* 299v. Am. J. Clin. Nutr. **73:**380S-385S.
15. **Olsen, E. B., J. B. Russell, and T. Henick-Kling.** 1991. Electrogenic L-malate transport by *Lactobacillus plantarum*: a basis for energy derivation from malolactic fermentation. J. Bacteriol. **173:**6199-6206.
16. **Russell, J. B.** 1992**.** Another explanation for the accumulation of fermentation acids at low pH, anion accumulation versus uncoupling. J. Appl. Bacteriol. **73:**363-370.
17. **Russell, J. B., and F. Diez-Gonzalez.** 1998. The effects of fermentation acids on bacterial growth. Adv. Microbial Physiol. **39:**205-234.
18. **Shelef, L. A.** 1994. Antimicrobial effects of lactates, A review. J. Food Protection. **57:**445-450.
19. **van der Werf, M. J., B. Pieterse, N. van Luijk, F. Schuren, B. van der Werff-van der Vat, K. Overkamp, and R. H. Jellema.** 2006. Multivariate analysis of microarray data by principal component analysis: Prioritizing relevant transcripts linked to the degradation of different carbohydrates in *Pseudomonas putida* S12. Microbiol. **152:**257-272.
20. **Vescovo, M., S. Torriani, F. Dellaglio, and V. Bottazzi.** 1993. Basic characteristics, ecology and application of *Lactobacillus plantarum*: a review. Ann. Microbiol. Enzimol. **43:**261-284.
21. **Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown.** 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. **99:**5860-5865.
22. **Weber, A., and K. Jung.** 2002. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. J. Bacteriol. **184:**5502-5507.
23. **Yale, J., and H. J. Bohnert.** 2001. Transcript expression in *Saccharomyces cerevisiae* at high salinity. J. Biol. Chem. **276:**15996-16007.
24. **Yamamoto, N., Y. Masujima, and T. Takano.** 1996. Reduction of membrane bound ATPase activity in a *Lactobacillus helveticus* strain with slower growth at low pH. FEMS Microbiol. Lett. **138:**179-184.

Samenvatting

In het afgelopen decennium zijn de volledige genoomsequenties van enkele honderden organismen, waaronder de melkzuurbacterie *Lactobacillus plantarum* WCFS1, bekend geworden. Hoewel de genoomsequentie de blauwdruk vormt voor het organisme en zijn verschillende eigenschappen, worden deze pas geëffectueerd wanneer het genotype wordt omgezet in een fenotype. De eerste stap daarin wordt gevormd door transcriptie van de genen in hun corresponderende (m)RNA's. Aangezien veel cellulaire processen op dit niveau worden gereguleerd, weerspiegelt de samenstelling van het totaal aanwezige RNA in de cel, het transcriptoom, voor een belangrijke deel de fysiologische status van de cel. In transcriptomics onderzoek wordt hier gebruik van gemaakt door de transcriptomen van, onder verschillende condities gekweekte, cellen te vergelijken en vervolgens die genen te identificeren die differentieel tot expressie komen tussen deze condities.

In dit proefschrift wordt beschreven hoe een transcriptomics aanpak is gebruikt in studies naar de respons van *Lactobacillus plantarum* WCFS1 op melkzuur- en zoutstress. Hiernaast is ook onderzoek gedaan dat gericht was op het optimaliseren van verschillende stappen in transcriptomics gebaseerd onderzoek.

RNA, het uitgangsmateriaal voor transcriptomics onderzoek, is doorgaans zeer instabiel. Significante verschillen in mRNA niveaus kunnen reeds binnen enkele minuten optreden. Dit betekent dat wanneer de processen van RNA degradatie en *de novo* synthese niet direct op het moment van het oogsten van de cellen worden stilgelegd, wijzigingen geïntroduceerd kunnen worden in het transcriptoom die niet representatief zijn voor de kweekcondities. In Hoofdstuk 2 wordt de validatie van een quenchingsprotocol beschreven dat als doel heeft dergelijke wijzigingen te minimaliseren. Hierbij worden de biologische processen in de cel binnen enkele seconden na het oogsten van de cellen stilgelegd door ze op te vangen in een buffer oplossing op methanol-basis met een temperatuur van –45°C. Cellen afkomstig uit een enkele steady-state cultures werden geoogst volgens het quenchingsprotol, dan wel volgens twee meer gebruikelijke methoden, voor welke de geschatte tijd tussen oogsten en stilleggen van cellulaire activiteit op respectievelijk drie en tien minuten ligt. De transcriptomen werden vervolgens vergeleken door middel van microarray analyse. De resultaten van dit experiment gaven aan dat het quenchingsprotocol de conservering van het transcriptoom verbeterde en daardoor op de representativiteit en betrouwbaarheid van de transcriptoom data. Principale component analyse liet een duidelijke groepering van de monsters zien naar de verschillende protocollen waarop ze uit de steady-state culture waren geoogst. Een deel van de genen die significant verschillende mRNA niveaus vertoonden tussen de verschillende protocollen duidt op een koude-stress respons in de cellen die niet volgens het quenchingprotocol waren geïsoleerd. Naast het effect van het quenchingsprotocol

op het behoud van mRNA niveaus van specifieke genen, gaven de resultaten van deze studie ook aan dat gebruik van deze methode leidt tot een hogere reproduceerbaarheid.

Terwijl de meeste transcriptomics studies worden uitgevoerd op microarrays waarop iedere spot een specifiek gen vertegenwoordigd, werd dit onderzoek uitgevoerd op kloongebaseerde-microarrays. Het voordeel van dit type arrays is dat ze ook toepasbaar zijn voor organismen voor welke de volledige genoomsequentie niet beschikbaar is. Het is in dat geval niet op voorhand bekend welke genen vertegenwoordigd zijn op de array. Bovendien kan een enkele spot op de array meerdere genen vertegenwoordigen. Om beter in staat te zijn om bij het ontwerp van een prokaryote kloongebaseerde-microarray een nauwkeurige inschatting te maken van het aantal vertegenwoordigde genen, zijn twee complementaire vergelijkingen ontwikkeld die op basis van enkele bekende of te kiezen eigenschappen van microarray en organisme hiertoe de mogelijk bieden (Hoofdstuk 3). De eerste vergelijking voorspelt de fractie aan genen die met voldoende nucleotiden op een fragment op de array vertegenwoordigd zijn en de tweede voorspelt de fractie aan genen voor welke kwantitatieve informatie gegenereerd kan worden.

Melkzuur vormt het belangrijkste fermentatieproduct van *Lactobacillus plantarum* en andere melkzuurbacteriën. De ongedissocieerde vorm van het melkzuur molecuul kan een sterk remmend effect hebben op de groei van het organisme. Het specifieke effect van de ongedissocieerde vorm van organische zuren is doorgaans lastig te bestuderen doordat zij in een pH-afhankelijk evenwicht voorkomen met hun gedissocieerde vorm. In een studie naar het effect van melkzuur op genexpressie in *L. plantarum* (Hoofdstuk 4) is van een proefopzet uitgegaan die het mogelijk maakte onderscheid te maken tussen de effecten van melkzuur in zijn gedissocieerde en ongedissocieerde vorm, pH, osmolariteit en relatieve en absolute groeisnelheid. Dit heeft onder andere geresulteerd in de identificatie van een groep genen die een specifiek hogere expressie vertoonden in respons op de ongedissocieerde vorm van het melkzuur molecuul. Deze groep omvatte onder meer enkele bekende stress-genen, twee regulatorgenen en een coherente groep oppervlakte-eiwitten coderende genen. Er is getracht de rol van enkele van deze genen bij melkzuurstress te bewijzen met een moleculair genetische aanpak, maar dit heeft vooralsnog geen uitsluitsel kunnen geven over de specifieke rol van deze genen onder melkzuur stress (Hoofdstuk 5). Een tweede groep genen vertoonde differentiële expressie in respons op de combinatie van ongedissocieerd melkzuur met een lagere groeisnelheid. Deze zijn waarschijnlijk betrokken bij de vorming van andere fermentatie-eindproducten dan melkzuur en zouden hiermee pyruvaat en melkzuur accumulatie en NAD^+ depletie kunnen beperken.

Hyperosmose is een veel voorkomende stressfactor voor micro-organismen in de levensmiddelenindustrie. Het initiële effect van hyperosmose op de cellen valt grotendeels toe te schrijven aan afname van de turgor. Veel prokaryoten zijn in staat om deze initiële stress te overwinnen door accumulatie van zogenaamde “compatible solutes”. Desalniettemin zal de hyperosmose op de lange termijn ook effect hebben op de cellen. In Hoofdstuk 6 is het lange termijn effect van NaCl op genexpressie in *L. plantarum* bestudeerd. Er werd geen toegenomen expressie waargenomen voor een aantal genen die met de acute hyperosmotische respons in melkzuurbacteriën zijn geassocieerd zoals de genen voor de compatible solute transporters en GroES, GroEL en HtrA. De genen waarvoor differentiële expressie werd waargenomen coderen voor eiwitten uit verschillende functionele groepen. Duidelijke trends werden waargenomen die aanpassingen in de samenstelling van de celwand suggereerden en een toegenomen rol voor genen die betrokken zijn in het herstel van DNA schade. Opmerkelijk was het grote aantal genen en operonen met de consensus sequentie van het koolstof kataboliet responsieve element *cre*, de bindingsplaats voor regulatoreiwit CcpA, in de veronderstelde promoter regio. Differentiële expressie van deze genen kon niet in verband worden gebracht met verschillende glucose niveaus, noch met de energetische status van de cellen. Een mogelijke verklaring is dat de afname van vrij cytoplasmatisch water ten gevolge van de hyperosmose resulteert in een toegenomen concentratie van metabolieten die CcpA afhankelijke genexpressie beïnvloeden.

Dit proefschrift levert een bijdrage aan de betrouwbaarheid en toepasbaarheid van transriptomics onderzoek in, met name, microorganismen. Daarnaast wordt de kracht van transcriptomics voor het karakteriseren van het effect van verschillende (stress)condities op de totale microbiële cel geillustreerd aan de hand van de studies naar de effecten van NaCl- en melkzuurstress op genexpressie in *L. plantarum*. De genen die in deze studies geïdentificeerd zijn kunnen mogelijk een rol spelen bij het optimaliseren van *L. plantarum* en andere melkzuurbacteriën voor industriële toepassingen.

Nawoord

Ondanks dat het al weer een tijdje geleden is dat ik aan het werk dat aan dit proefschrift ten grondslag ligt ben begonnen, zijn de tussenliggende jaren in een sneltreinvaart voorbijgegaan. Er is in die tijd veel, heel veel, gebeurd. Slechts een deel hiervan is terug te vinden in de voorgaande hoofdstukken. De rest zijn wetenschappelijke gedachtespinsels die veilig in mijn hoofd opgeborgen zitten en af en toe nog eens de kop op steken, maar vooral ook ontwikkelingen in mijn privé-leven.

Toen ik nog moest starten met mijn onderzoek was het al duidelijk uit de literatuur dat transcriptomics als onderzoekstechniek veel potentie heeft. Dit was de voornaamste reden dat ik enthousiast was over dit project. Na aanvang bleek al snel dat er nog veel te optimaliseren viel aan de hele transcriptomics aanpak en dat deze methode wezenlijk anders is dan waar iedereen in de moleculaire biologie altijd aan gewend was. Eigenlijk werd mijn aio-onderzoek hiermee nog veel interessanter. Een belangrijk deel van mijn onderzoek heeft zich dan ook gericht op het optimaliseren van transcriptomics gerelateerde procedures. Een bescheiden bijdrage realiseer ik me, want nu ik alweer een jaar bezig ben met mijn huidige baan heb ik alweer aardig wat dingen ontdekt die nog handiger kunnen of succesvoller zijn in het extraheren van de meest relevante informatie uit de data. Je blijft bezig en gelukkig doe je dit meestal niet in je eentje.

Zoals ik al aangaf, is er in dezelfde periode ook heel veel gebeurd in mijn privé-leven. Hele hoge hoogtepunten, maar ook hele diepe dieptepunten. Nu ben ik er altijd een voorstander van om werk en privé tot een bepaalde hoogte gescheiden te houden, maar als het er op aan komt is het toch heel fijn als je een groep goede collega's hebt die met je meeleeft. Andersom geldt ook dat het helemaal niet erg is als je familie en vrienden af en toe eens informeren hoe het met je werk gaat ("wanneer ga je afstuderen?").

Ik heb in de afgelopen jaren van vele kanten hulp en interesse ontvangen op alle mogelijke vlakken. Er zijn dan ook aardig wat mensen die ik mijn dank verschuldigd ben.

Mariët, co-promotor, dank voor je geweldige begeleiding. Hoewel je wel altijd in de gaten had waar ik mee bezig was, heb je me van het begin af aan (uitzonderlijk) veel vrijheid gegeven. Hierdoor heb ik echt mijn eigen onderzoek op kunnen bouwen. Dit wil overigens niet zeggen dat je me maar klakkeloos mijn gang liet gaan: als ik iets graag wilde, mocht ik dit eerst verdedigen. Waren we het eens, dan vond ik je ook tot het eind toe aan mijn zijde.

Willem, promotor, ons contact is met name naar het eind toe intensiever geworden. Ik heb bewondering voor de manier waarop je het steeds weer voor elkaar kreeg om zeer snel en ook nog eens uitermate scherp te reageren op mijn stukken.

Voor mijn ex-collega's op TNO, waar ik het merendeel van mijn onderzoek heb uitgevoerd: ik heb het bijzonder naar mijn zin gehad bij AMGT, een clubje vol uitgesproken karakters. Op de zeldzame dagen dat ik ernstig de balen had van mijn onderzoek, zorgden jullie ervoor dat het toch weer de moeite waard was om 's morgens vol goede moed aan te sluiten in de file naar mijn werk. Ik heb met jullie een uitermate sociale en behulpzame groep collega's getroffen!

Door het onderwerp waar ik aan werkte heb ik inhoudelijk en praktisch met name veel gehad aan Rob Leer, de fermentatieclub, het microarrayteam, het metabolomicsteam en Renger en Bianca. Daarnaast kan ik me eigenlijk niemand indenken van het AMGT-clubje die me niet op enig moment geholpen heeft, dan wel mij gemotiveerd heeft enkel en alleen al door interesse te tonen in mijn bezigheden. En hoewel het al een plezier was om met jullie op het lab te staan, dan wel een serieuze bespreking te voeren, de beste momenten speelden zich meestal af tijdens de koffie!

Mijn studenten, Robert en Suzanne, ik heb veel van jullie geleerd. Robert, ik ben blij dat ik nog een bijdrage heb kunnen leveren in het overhalen van jou om te solliciteren naar de aio-plek die je nu hebt. Ik weet zeker dat dit een hele goede keuze is geweest!

Ik heb na de reorganisatie helaas niet lang genoeg meer op TNO rondgelopen om goed te kunnen integreren met de nieuwe collega's van de micro-club. Vandaar dat ik me hier even beperk tot het bedanken van Ellen en Hakim.

Mijn onderzoek werd uitgevoerd binnen het WCFS-projectteam van Michiel. Daar waar ik me vooral richtte op alles wat met –omics te maken heeft, werd mijn lacto-kennis hier aangevuld door mijn teamgenoten. Het hoogtepunt van het WCFS-samenwerkingsverband was voor mij mijn periode op het NIZO waar ik geweldig ben geholpen door mijn WCFS collega's ter plekke. In dit verband verdienen Michiel, Jolanda, Roger en Peter een speciale vermelding.

Jan en Vief, naast dat jullie mijn collega's waren, zijn jullie ook nog eens mijn paranimfen. Ik ben vereerd en blij dat jullie dit willen doen. Vief, jij was altijd een prettige collega, maar bovenal ben je een waardevolle vriendin. Jan, jij was en bent het sociale gezicht van de afdeling en naturrlijk de persoon om mee in contact te treden als je een kippetje wilt scoren.

En dan mijn familie en vrienden. Zoals ik al eerder zei, is het leuk als mensen belangstelling tonen in je werk. Hier heb ik beslist niet over mogen klagen. Belangrijker nog is de afleiding die jullie me hebben bezorgd, waardoor ik in mijn vrije tijd eigenlijk maar

zelden in mijn hoofd met mijn werk bezig was. Ik weet zeker dat dit mijn proefschrift ten goede is gekomen, want hierdoor heb ik tot het eind toe het plezier in mijn onderzoek gehad.

Speciale dank voor mijn ouders en schoonouders die het afgelopen jaar regelmatig op Merijn hebben gepast terwijl Ilse en / of ik boven achter de compu zaten, en daarnaast enorm hebben geholpen met de verhuizing naar Amersfoort.

Voor mijn ouders, die ook ooit nog hebben mogen meemaken dat ik mijn veterstrikdiploma haalde: dank voor jullie steun om me steeds weer verder te ontwikkelen.

Lieve Ilse, lieve Merijn, ik begon mijn promotie in mijn eentje en nu sluiten we twee promoties af met zijn drieën en een vierde op komst. Merijn, met jouw “papa” en je vette lach heb je me een heel nieuwe kant van mezelf laten ontdekken. Sorry dat ik geen leuker boekje heb gemaakt. Dit haalt het duidelijk niet bij Nijntje! Ilse, wat een wonder dat ik jou heb mogen ontmoeten. Dankzij jouw liefde heb ik het vertrouwen, de kracht en de rust om mezelf te blijven. Dit is een grotere prestatie dan het behalen van een doctorstitel. Je bent heel bijzonder.

List of publications

Pieterse, B., R. H. Jellema, and M. J. van der Werf. 2006. Quenching of microbial samples for increased reliability of microarray data. J. Microbiol. Methods. **64:**207-216.

Pieterse, B., M. Kleerebezem, and M. J. van der Werf. Transcriptome analysis of the effect of continuous NaCl stress on gene expression of *Lactobacillus plantarum.* Manuscript in preparation.

Pieterse, B., R. J. Leer, F. H. Schuren, and M. J. van der Werf. 2005. Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling. Microbiol. **151:**3881-3894.

Pieterse, B., E. J. Quirijns, F. H. Schuren, and M. J. van der Werf. 2005. Mathematical design of prokaryotic clone-based microarrays. BMC Bioinformatics. **6:**238.

Sonneveld, E., J. A. Riteco, H. J. Jansen, B. Pieterse, A. Brouwer, W. G. Schoonen, and B. van der Burg. 2006. Comparison of in vitro and in vivo screening models for androgenic and estrogenic activities. Toxicol Sci. **89:**173-187.

van der Werf, M. J., B. Pieterse, N. Van Luijk, N., F. Schuren, B. van der Werff-van der Vat, K. Overkamp, and R. H. Jellema. 2006. Multivariate analysis of microarray data by principal component discriminant analysis: prioritizing relevant transcripts linked to the degradation of different carbohydrates in *Pseudomonas putida* S12. Microbiol. **152:**257-272.

Curriculum vitae

Bart Pieterse werd geboren op 9 augustus 1975 te Delft. In 1993 behaalde hij zijn VWO diploma aan het Koningin Wilhelmina College te Culemborg. In 2000 werd de studie Bioprocestechnologie voltooid aan de Wageningen Universiteit, voor welke afstudeervakken werden gevolgd in de richtingen industriële microbiologie en moleculaire genetica van industriële micro-organismen. In hetzelfde jaar werd een begin gemaakt met het werk dat beschreven wordt in dit proefschrift. Het onderzoek werd uitgevoerd voor het Wageningen Centre for Food Sciences bij TNO Quality of Life te Zeist onder directe begeleiding van Dr. Ir. M. J. Van der Werf en promotorschap van prof. Dr. W. M. De Vos (Wageningen Universiteit, laboratorium voor Microbiologie). Vanaf 2005 is hij werkzaam als onderzoeker op het gebied van bioinformatica en data-analyse bij BioDetection Systems b.v. te Amsterdam.

printed by: Ponsen & Looijen B.V., Wageningen