CHAPTER 5

THE GROWTH-DEPENDENT TRANSCRIPTOME OF *LACTOCOCCUS LACTIS* MG1363: FOUNDATIONS FOR A BACTERIAL BIOGRAPHY§

§João P. C. Pinto, Rutger Brouwer, Araz Zeyniyev, Oscar P. Kuipers and Jan Kok
Abstract

Whole-genome transcriptional analyses performed on microorganisms are traditionally based on a small number of samples. In order to be able to map transient expression variations and thoroughly characterize gene expression throughout the entire growth curve of the widely used model organism *Lactococcus lactis* MG1363 we collected gene expression data with unprecedented time resolution using DNA microarrays. Thus, we globally analyzed the resulting gene expression patterns in a number of different ways to demonstrate the richness of the data and the ease with which novel phenomena can be discovered. We show that, when the culture moves from one growth phase to another, gene expression patterns change to such an extent that we suggest that those patterns can be used to unequivocally distinguish growth-phases apart from each other. Also, we were able to distinguish, within the classically defined growth-phases, sub-growth-phases with a distinct biological signature. Apart from the global expression pattern shifts seen throughout the growth-curve, several cases of short-lived transient gene expression patterns were clearly observed. These could *e.g.*, help explain the gene expression variations frequently observed in biological replicates. We have devised a method to compute unnormalized/absolute gene expression levels and used it to determine how global transcription patterns are influenced by nutrient starvation or acidification of the medium. For example, we inferred that *L. lactis* MG1363 produces proteins with on average lower pI’s and lower molecular weights as the medium acidifies and nutrients get scarcer.

This dataset is in itself a rich resource for microbiologists interested in common mechanisms of gene expression, regulation and in particular the physiology of *L. lactis*. Thus, similarly to the common use of genome sequence data by the scientific community, the dataset constitutes a very rich repository for mining and an opportunity for bioinformaticians to develop novel tools for in-depth analysis.
Introduction

With the availability of eight genome sequences of the industrially important and well-characterized bacterium *Lactococcus lactis* (1–8), post-genomics experimental approaches such as transcriptomics, proteomics and metabolomics analyses can robustly be applied to this organism. Also, from the resulting sets of data and aided by specifically designed bioinformatics and modeling tools, systems biology holistic approaches have greatly improved our understanding of this lactic acid bacterium. Altogether, these methods have already allowed improved genome annotation and yielded valuable information for the construction of complete metabolic network models (see for example (9)). However, the dynamics of cellular processes, specifically the rapid fluctuations in gene expression necessary for keeping homeostasis during growth and adaptation to natural stresses caused by intrinsic and extrinsic factors such as changing nutrient availability, oxygen pressure, or pH, have not yet been studied in depth.

Whole-genome transcriptional analyses are traditionally based on a small number of samples. However, some transitions in gene expression could in principle be fast and transient, as indeed is shown in this report, resulting in "spiky" expression patterns or fast oscillations throughout the growth curve. Moreover, transition points between the classical growth phases could be sharp and might be missed when sampling is not frequent enough. As recently demonstrated for *Bacillus subtilis* (10), a detailed profile of transcriptional fluctuations in expression of each gene during growth of a bacterium that is normally grown in batch fermentation is clearly desired in order to obtain a rich repository of the expression profile of each gene in the genome. Such information would allow extensive and in-depth mining for correlations or anti-correlations between genes with regard to their temporal expression. These kinds of analyses could result in better operon definitions, help explain temporary physiological behavior and assist in identifying functions of genes with unknown functions. Moreover, they may help to clearly define growth phases in terms of signatures of gene expression and possibly identify sub-phases within and between the growth phases.

Here, we have developed an experimental design that charts the often rapid transcriptional changes of genes of *L. lactis*. We have sampled a batch-grown culture of the widely used model organism *L. lactis* MG1363, grown in GM17 broth at 30°C, every 15 minutes after OD$_{600}$ reached approximately 0.1 and then until 12 hours after inoculation. Triplicate samples corresponding to 24, 36 and 48 hours after inoculation were also obtained to assess whether further changes took place after such prolonged incubation. In this way we obtained 45 samples in time, the transcriptomes of which were analyzed using a specific hybridization scheme yielding 6 transcription readings per gene per time point (see Materials and Methods).
The results show, among others, that gene expression patterns can be used to unequivocally distinguish the different growth-phases in *L. lactis* batch fermentation. In addition, we have identified several short-lived transient gene expression patterns that could e.g., help explain gene expression variations frequently observed in biological replicates in microarray experiments.

To the best of our knowledge this is the first report with such a resolution on gene expression during growth of any lactic acid bacterium. The entire data is available at http://webserver.molgenrug.nl/publication/MG_TS as we believe it to be a rich resource for the microbial research community.

**Results and Discussion**

*Growth of Lactococcus lactis ssp. cremoris MG1363 in GM17*

A 12-liters GM17 batch fermentation was initiated with a 1% inoculum of an exponentially growing culture of *L. lactis* MG1363 growing in the same medium (see Materials and Methods). The absence of an adaptation (lag) phase led to a well-defined and reproducible growth curve, with minimal interference of the history of the inoculum on the growth of cells during the fermentation (Figure 1). The reproducibility of growth using this procedure was confirmed by repeating the fermentation procedure 3 times (data not shown). Growth arrest occurred at an optical density measured at 600 nm (OD$_{600}$) of 2.8 as a result of depletion of glucose from the medium (Figure 1). This metabolic constriction and its consequent arrest of the production of lactate also coincided with the stabilization of the pH of the medium at a value of 5 (Figure 1). It should be noted that glucose is limiting when its initial concentration is below approximately 0.75%. When higher concentrations of glucose are used, a further drop in pH limits growth if the lactate produced is not neutralized (data not shown). At around 1.5 h after inoculation the culture became dense enough to allow sampling of sufficient cell material for the DNA microarray procedure. Samples were taken every 15 min during the first 12 h, after which three samples were then taken at 24, 36 and 48 hours after inoculation, after which those from 1 fermentation were examined by DNA microarray analysis (see below).

Exponential growth was characterized by two discernible phases, each with a distinct growth rate (Figure 1). Optical density remained constant between 6 and 24 h after inoculation. However, whole transcriptome analyses suggest the existence of three distinct periods in the stationary phase (see below). After 24 h the optical density began to decrease, corresponding to the canonic death phase in bacterial growth. At the same time, medium pH increased marginally suggesting that, since intracellular pH is likely to be higher than that of
the medium even during this stage (11), the observed lysis of part of the culture slightly neutralized the medium.

**Figure 1 - Growth of *L. lactis* MG1363 in GM17.** OD$_{600}$, pH and dissolved glucose in the medium were recorded throughout the batch fermentation. Grey and white bands indicate the different growth phases as indicated on the top axis. Exp., exponential; stat., stationary.

**Dimensioning-Noise-Amplitude (D-N-A) Scaling**

We observed that, using the standard normalization and scaling of the DNA microarray data (see Materials & Methods), the method improperly transformed values in a way that made the average of all signals to remain roughly constant in all time points. This artificially pushed the values of the last time points (namely in the death phase) towards a plateau not consistent with e.g., cDNA synthesis yields (typically 10 times lower in the death phase than the average cDNA yield of all other samples; data not shown). Even genes that were least expressed throughout the entire growth, with microarray signal intensities in the order of magnitude of what should be considered noise, showed high expression levels towards the end of the experiment (Figure 2).
To correct for this artifact we employed the assumption that noise, since it is of technical origin, should have a constant dimension throughout all data points and used that measure to recalibrate the entire set of data of all genes in all time points. Thus, the genes for which the expression was always below the median in all the time points were initially screened. In this way, the signals of a set of 368 genes (a representative selection of which is given in Figure 2) were set aside as an initial measure of what should be considered “noise”. Due to the stochastic nature of noise, the entire set of 368 gene expression profiles was not used, but rather the 30th and 90th percentile signals from each time point of genes within this group. Also, for each time point, these values were averaged with the ones from the two adjacent time points, thus producing a robust proxy of the dimension of noise. This information was then used to transform signals of all genes in the following manner: the mentioned 3-time-points moving average of the 30th percentile was subtracted from all signals for each time point and then the resulting value was divided by the 3-time-points moving average of the 90th percentile. The outcome of this transformation on the so-called background gene expression signals is illustrated in Figure 2. In this new scale, gene expression is not entirely arbitrary but a relative measure against noise (an expression value of 2 represents roughly twice the average value of noise).

Figure 2 – Dimensioning-Noise-Amplitude Scaling. A) A sample of 5 out of 368 genes that, despite being the least expressed throughout growth, displayed a significant signal increase as a result of the standard inter-slide scaling. B) The signal from the same 5 genes after D-N-A scaling.
This method does not merely lead to a minor correction of the data, it has a fundamental influence in the interpretation of that data. For example, one could be led to believe that the prophage gene *ps342* is highly expressed towards the end of the stationary phase, at a level that compares *e.g.*, to the peak in *pmrB* expression just before cells stop growing exponentially (Figure 3). According to data obtained after D-N-A scaling, although *ps342* indeed starts to be transcribed during stationary phase, the magnitude is certainly not comparable to the peak expression of *pmrB* (Figure 3).

**Whole Transcriptome Analyses**

As described above, a novel method was developed to recapture a sense of the absolute levels of expression of genes throughout growth (although there is a technical difference between level of transcription and measured cDNA signal, such distinction was generally not made in this text for the sake of clarity). As such, it is possible to infer that the global transcription activity remains roughly constant until the moment that coincides with a swift decrease in the growth rate around time points (tp) 12 and 13, around 4.5 h after inoculation (Figure 1 and Figure 4). In a second stage, here discriminated as exponential phase 2 since the optical density of the culture is still increasing logarithmically (although at a lower rate), the overall transcription level initially decreases but quickly rises as the culture approaches the stationary phase, around tp 18 and 19, a little over 6 h after inoculation.

Glucose depletion and the consequent growth arrest were accompanied in a first stage by a peak in expression around 40% greater than what was globally observed during exponential
growth (Figure 4). This reveals that cells undergo a substantial adaptation process in the initial period of growth arrest. Although the stationary phase, in the standard sense that optical density does not change, lasted until 24 h after inoculation, three distinct sub-stationary phases were discernible on the basis of the overall transcription level of the cells (Figure 4). In the initial stationary phase 1 (tp 19 to 24) global transcription peaks but later decreases to a plateau of transcription, similar to the one measured during the exponential growth period, in the subsequent stationary phase 2 (tp 25 to 38). Some 5 h after glucose depletion (from tp 38 onward), global transcription levels suddenly drop to what likely represents an energy-saving mode as well as the onset of a phase in which cells fail to maintain essential processes that ultimately precludes viability. As global transcription levels decrease, culture optical density is initially maintained (stationary phase 3; tp 39 to 42) but ultimately the cells start lysing. This final period, 24 h after inoculation and onwards, constitutes the death phase (tp 43 to 45). As expected, the total culture in this phase shows a reduced transcriptional activity that is 4 times lower than the culture’s peak activity in transcription.

It is also interesting to notice that, at any given moment in growth, the variable set of the top 10% of most expressed genes accounts on average for 50% of the transcriptome of L. lactis at that point in time (Figure 4). The top 20% of most expressed genes at any given moment corresponds to roughly 80% of all transcription, which is yet another example of the seemingly universal Pareto principle (also called the 80-20 rule).

Of the 2308 genes spotted on the DNA microarray slides, only 368 genes were not at least once more expressed than the median at any given moment during growth. As the expression of some of these genes might not have been measured accurately due to inefficient DNA microarray probes, this estimate may be overstated. This set represents less than 20% of all L. lactis genes. Thus, most genetic networks are active at least at some stage during growth of L. lactis in batch culture in a complex medium, where cells go through different stages of growth as a consequence of varying conditions, such as nutrient starvation and medium acidification.
Figure 4- Inferred global transcription (A) and expression percentile (B) levels throughout the growth of *L. lactis* MG1363 in GM17. Grey and white bands: growth phases as presented in Figure 1.
Principal Component Analysis (PCA) was used to qualitatively assess the variation in the DNA microarray samples. Remarkably, PCA supports the partitioning of growth phases, as described above, according to growth rates and global levels of transcription. The first two principal components (PCs) account for approximately 60% of all variation in the dataset and enable the separation of the samples according to the growth phase from which they were taken (Figure 5). Samples 1 to 12, which were taken during exponential growth phase 1, cluster close together in this analysis using the first two PCs. The great resemblance of the transcriptomes throughout these initial time points, due to the absence of a detectable lag/adaptation phase, is probably the result of using an inoculum that was kept growing exponentially for approximately 40 h. After this phase the profile of gene expression, as mapped with PCA, drifts continuously as a result of the adaptations that cells have to undergo. Later, a few time points into stationary phase 2, cells again seem to stabilize their transcription profiles since tp 27 until 38 cluster close together (Figure 5). Afterwards, and while the culture is still able to keep a constant optical density, the collective transcription pattern is further modified (tp 39 until 42). The last 3 time points are characterized both by a decrease in the OD600 and a drift regarding the gene expression profile.

Figure 5 – Principal Component Analysis. PCA was applied to all gene expression data. All time points were mapped in relation to the two most relevant principal components (PC1 and PC2).
The Time-Series as a Reference Curve

An immediate application of the transcriptome data reported here follows from the PCA described above. Genes whose expression contributed strongly to separate the various time points according to their expression profiles via PCA are characterized by well-defined periods of expression (Figure 6). This information allows designing, e.g., growth-phase markers or validating and interpreting other DNA microarray data using this time-series (chrono)-transcriptome data set as a reference.

Using the expression profiles of the 45 growth-phase marker genes selected from the PCA analysis, one is be able to validate the growth stage from which an *L. lactis* culture was sampled. Most often, samples from bacterial cultures are taken during mid-exponential or mid-stationary phases of growth. We show here that those periods are characterized by major changes in gene expression patterns. Thus, two samples taken from approximately the same growth phase may in fact not be comparable with respect to their transcription profiles. This problem may be exacerbated if one of the cultures is treated in such a way that the growth rate is even slightly affected, as that could perturb the synchronicity of sampling.

Such matching using growth-phase marker genes enables the current chrono-transcriptomics experiment to be used as a background model for other more conventional single time-point experiments. By identifying the chrono-transcriptomics sample most identical to such an experimental sample, a researcher can determine which changes in gene expression are likely to have been caused by growth effects and which changes are due to e.g., the condition applied. Furthermore, when examining an *L. lactis* strain that is severely retarded in growth, this analysis may be used to more accurately determine the approximate growth-phase of the culture than what visual inspection of the growth curve would allow to.

Three DNA microarray datasets (12) were matched to samples from the chrono-transcriptomics experiment presented here (Figure 7). In all of these experiments an *L. lactis* MG1363-derived strain was compared to the same strain that was induced to overproduce a membrane protein. The *L. lactis* cells overproducing the homologous protein OpuA, but sampled just prior to induction, greatly resemble cells around tp 10 in the *L. lactis* chrono-transcriptome. This is as expected since cells were induced around that period of growth and optical density. Approximately 64 min after induction the culture became slightly affected; it lost a bit of the identity of an exponentially growing culture and its overall gene expression profile slightly increased to that of a culture later in growth (Figure 7). This observation was even more pronounced in a culture of cells overproducing the human protein Presenilin, as indeed expected since these cells display a significant growth defect (12) (Figure 7).
Figure 6 – Genes that contributed the most to PC1 and PC2 constitute optimal markers for growth-phase discrimination since they are expressed only during specific periods of the growth curve.
Use of the *L. lactis* chrono-transcriptome data set as a reference curve. *L. lactis* 0 min (blue) or 64 min (red) after induction of OpuA overproduction, and *L. lactis* overproducing the human protein Presenilin, 64 min after induction (violet), were matched with the 45 time points of the chrono-transcriptome.

As most of the obtained transcriptional datasets are highly comparable to the data of the chrono-transcriptomics experiment described here, since *L. lactis* cells are typically grown under roughly the same conditions (standing GM17 batch cultures at 30°C), gene expression information from the latter can be transposed to the experiments described in literature. This is especially relevant when only a small number of biological replicates are described in the published experiments. This high-density chrono-transcriptomics study can serve as a reference to clarify whether observed expression changes are due to 1) the treatment applied to cells, 2) expression fluctuations that characterize the sampling period (as discussed below) or 3) indirect effects such as those that derive from a growth rate change.

**Abrupt Transient Changes in Gene Expression**

An important observation from this high-resolution chrono-transcriptome is that a fraction of the genes of *L. lactis* display abrupt and transient changes in gene expression, even during the middle of the seemingly monotonous exponential phase 1. Notably, genes involved in purine and pyrimidine biosynthesis (13) and those involved in arginine metabolism (14) belong to this group (Figure 8).
Figure 8 – Expression profiles of genes representative of operons/regulons involved in purine (A), pyrimidine (C) and arginine metabolism (B and D).

The behavior of these genes could possibly lead to misinterpretations in the analyses of single time-point transcriptome experiments where, generally, a variant strain in which a gene is deleted or overexpressed is compared to its cognate parental strain. Classically, these cultures are harvested around the mid-exponential growth phase where a minor deviation in the growth of both strains could result in significant differences in the amounts of specific mRNA’s. Thus, genes showing such transcription bursts could easily be falsely enlisted as being differentially expressed. Indeed the reason that \textit{arg, arc, pyr} and \textit{pur} genes are
frequently observed as being affected in these types of experiments could be due to the fact that these genes show such an irregular expression behavior especially during the exponential phase of growth. By overlapping such expression patterns one can conclude that there are only very short-lived intervals where the overall expression pattern remains constant (Figure 9).

Figure 9 – Expression of certain genes of *L. lactis* MG1363 is characterized by abrupt transient changes.

**Expression Weighted Properties**

By plotting gene expression levels on a map of the genome of *L. lactis* MG1363 conspicuously silent regions were highlighted; these corresponded to places where the genome has undergone recent rearrangements, notably areas rich in transposable elements or carrying bacteriophage genes (8) (Figure 10). It should be noted that while most genes stop being transcribed during stationary sub-phase 3, some phage genes show an opposite behavior, since they were only transcribed when growth ceased. There is also a constant tendency throughout growth for genes close to the origin of replication to be expressed at higher levels than the average throughout the genome (Figure 10). This could be a consequence of genes closer to the origin of replication having on average a higher copy number per cell, simply because they are replicated sooner than the others during the cell cycle.
Figure 10 – The chrono-transcriptome plotted onto a chart of the *L. lactis* MG1363 genome. The six innermost circles indicate the level of expression of genes in the different growth phases (exponential sub-phase 1 being the closest to the center; the intensity of the red color is proportional to the averaged level of transcription in each phase). The seventh circle depicts prophage remnants (blue), IS elements (green) and pseudogenes (black). The outermost circle depicts the orientation of genes (red, when coded on the forward strand; green, when present on the reverse strand).

We also observed that throughout the entire chrono-transcriptome a bias exists towards the expression of genes encoding proteins with predicted pl’s lower than that of the average value of all proteins coded in the genome of *L. lactis* MG1363 (Figure 11A). This predisposition is further accentuated during growth and correlates with the decrease of the
pH of the medium. Possibly, *L. lactis* has evolved mechanisms to cope with lactate, its main metabolic by-product, by producing comparatively low-pI proteins when environmental pH decreases.

By extrapolating our data to translation, assuming that protein synthesis is roughly proportional to mRNA levels, one can infer that the average size of proteins predicted to be synthesized at any given moment is higher than the average size of all proteins coded in the genome. During the stationary sub-phase 3, cells seem to greatly diminish the average size of synthesized proteins (Figure 11B). This might represent a mechanism accounting for nutrient scarcity and a way to optimize resources.

We also observed a tendency of *L. lactis* to express genes with a better-than-the-average codon adaptation index (cai) throughout growth (Figure 11C). This trend, at the level of transcription, is quite constant during growth, being only slightly alleviated in the death phase, presumably because translation levels are so low that codon usage is not likely to be a constraint anymore.

The *L. lactis* chrono-transcriptomics results reveal a striking correlation in time between gene-strand bias and transcription intensity. Genes coded on the leading strand (transcription sense strand is the same as the replication leading strand) are on average more expressed than those coded on the lagging strand (transcription sense strand is the same as the replication lagging strand). One could hypothesize that this represents an adaptation to deal with the interference between DNA replication and RNA transcription, in particular with respect to the direction of one process relative to the other (15, 16). Indeed, the observed bias is more pronounced in fast growing cells, in the exponential phase, but minimal during stationary phase and onwards when both replication and transcription levels diminish (Figure 11D). It has been disputed whether essentiality or expressiveness could drive gene-strand bias (see for example (17)) although, very likely, both are relevant.

**Conclusions**

The transcriptomics data obtained in this work represents a huge mining repository for the LAB research community, similarly to that comprised by LAB genome data. We performed a global analysis on the whole dataset and a more in-depth survey into major fields of interest of *L. lactis* research. All raw data and the normalized gene expression profiles can be easily retrieved and viewed from the dedicated website at http://webserver.molgenrug.nl/publication/MG_TS and should allow rapid progress in both fundamental and application-oriented research on *L. lactis*. 
Chapter 5

Figure 11 - Expression Weighted Properties. A) Average inferred pI of all proteins putatively produced at a given time point (blue) and culture pH (red). B) Average inferred molecular weight of all proteins putatively produced at a given moment relative to the genomic averaged inferred value (32.6 KDa). C) Average inferred weighted codon adaptation index (cai; genomic average is 0.695). D) Relative transcription from the leading strand, normalized for the overrepresentation of genes coded on that strand.

Acknowledgements

We would like to thank Siger Holsappel for printing the microarray slides and both dr. Anne de Jong and dr. Evert-Jan Blom for the fruitful discussions regarding methods and data presented in this chapter.

Materials and Methods

Bacterial Strain and Growth Conditions

*Lactococcus lactis* ssp. *cremoris* MG1363 (18), from the same -80°C aliquot as the one that was used for the sequencing of the genome of this strain (8), was grown in M17 medium (Difco Laboratories - Detroit, MI, USA) supplemented with 0.5% w/v glucose (Acros Organics, Geel, Belgium) at 30°C. A chemostat of 16 L (working volume of 12 L) was used and
homogeneity of the culture was maintained through mild stirring (30 rpm). The medium pH was monitored with a built-in probe and growth was followed by measuring the optical density at 600 nm (OD$_{600}$). Pre-warmed medium was inoculated with a 1/100 volume of an *L. lactis* MG1363 GM17 culture maintained in the exponential phase of growth by means of consecutive dilutions, during a period of at least 48h. Reproducibility of the procedure was verified by repeating the above fermentation procedure 2 times. Cultures and inoculums were grown in media from the same batch. Samples from all inoculums and from the cultures at the end of the experiment were plated, examined under the microscope and re-grown in microtiter plates to check for potential contamination. Glucose concentration was measured via an enzymatic assay using the D-Glucose Kit (R-Biopharm, Darmstadt, Germany) according to the manufacturer's instructions. Measurements were done in duplicate on cell-free supernatant samples obtained during fermentation.

**Sampling and RNA Isolation**

The equivalent of 10 OD units of volume (Volume [mL] * OD$_{600}$) was taken in duplicate every 15 min during 12 h. Additionally, triplicate samples (15 min apart) were taken at each time point corresponding to 24, 36 and 48 h after inoculation. Cells were pelleted in 50-mL Greiner tubes in an Eppendorf 5810R centrifuge (Eppendorf AG, Hamburg, Germany) (1 min, 10000 rpm, 4°C) and resuspended in 0.5 mL of T$_{10}$E$_{1}$ (10 mM Tris-HCl, 1 mM Na$_2$-EDTA, pH 8.0), prepared with diethylpyrocarbonate (DEPC)-treated deionized water before transfer to 2-mL screw cap tubes. The resuspended cells were immediately frozen in liquid nitrogen and then kept at -80°C.

For RNA isolation, 0.5 g of glass beads (~ 100 μm in diameter), 50 μl of 10% Sodium dodecyl sulfate (SDS), 500 μL of premixed phenol:chloroform:isoamylalcohol (25:24:1) and 175 μL of macaloid suspension were added to the thawed cells in the screw cap tube. Cells were disrupted using 2 cycles of 45 sec of bead beating with a 1 min interval on ice. The cell lysate was cleared by centrifugation in an Eppendorf 5417R centrifuge (Eppendorf AG) (10 min, 10000 rpm, 4°C) after which the upper phase was extracted with 500 μL of chloroform:isoamylalcohol (24:1). The two phases were resolved by centrifugation (10 min, 10000 rpm, 4°C) and total RNA was isolated from the aqueous phase using the High Pure RNA Isolation Kit (Roche Molecular Biochemicals, Mannheim, Germany), according the manufacturer's instructions. RNA concentration was determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA); RNA quality was assessed using an Agilent Bioanalyzer 2100 with RNA 6000 LabChips (Agilent Technologies Netherlands BV, Amstelveen, the Netherlands).
**DNA Microarray procedure**

Synthesis of copy DNA (cDNA) was performed using the Superscript III Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA). Incorporation of amino allyl-modified dUTPs during cDNA synthesis allowed Cy3/Cy5 labelling with the CyScribe Post Labelling Kit (Amersham Biosciences, Piscataway, NJ, USA), according to the supplier's instructions. All intermediate and final purifications of either labeled or unlabeled cDNA were performed with a NucleoSpin Extract II Kit (Clontech Laboratories, Mountain View, CA, USA), according to manufacturer's instructions except while purifying unlabeled cDNA, where 80% ethanol was used in a second washing step and 0.1 M sodium bicarbonate, pH 9.0, was used as the elution buffer. The hybridization of all samples followed the scheme depicted in Figure 12. Thus, the transcript levels for every gene were measured at least 6 times for every time point (3 slides per time point and at least duplicate spots for each gene).

**Figure 12 - Hybridization scheme.** Synthesized cDNA obtained from the purified total RNA from each sample, labeled either with Cy3 or Cy5 (according to the green or the red hedge of the arrows, respectively), was hybridized according to the indicated dye-swap strategy on 3 DNA microarray slides.

Hybridization to the probes spotted on the *L. lactis* MG1363 mixed amplicon and oligonucleotide DNA microarray slides, covering 2308 of the 2435 predicted ORF’s, was done using the SlideHyb #1 hybridisation buffer (Ambion Biosystems, Foster City, CA, USA), during 16 h at 45°C. After hybridization, slides were washed for 5 min in 2xSSC (150 mM NaCl and 15 mM trisodium citrate) with 0.5% SDS, 2 times for 5 min in 1xSSC with 0.25% SDS and 5 min in 1xSSC with 0.1% SDS. All washing steps were performed at 30°C with preheated buffers. The washing buffers were removed via centrifugation (Eppendorf 5810R, 2 min, 2000 rpm). The DNA microarray slides were scanned using a GenePix Autoloader 4200AL confocal laser scanner (Molecular Devices Corporation, Sunnyvale, CA, USA).

The resulting images were analyzed using the ArrayPro Analyzer 4.5 software (Media Cybernetics, Silver Spring, MD, USA). Signal intensities were quantified for each spot on the DNA microarray slides after subtracting the background intensities, which were determined for each spot by reading the signals in the regions that separated diagonal spots. Signals
were initially normalized and scaled via LOWESS using the MicroPrep software (19), after which a Dimensioning-Noise-Amplitude (D-N-A) Scaling was performed, as described above.

**Output Visualization**

Gene expression profiles were studied and depicted using R 2.15.1 software packages (The R Project for Statistical Computing - http://www.r-project.org/). The gene expression plots illustrate the range of expression signals obtained for each gene at each time point from the 6 measurements, while the darker line within each range represents the median value. PCA data was equally obtained using R 2.15.1 software packages. The genome circular plot in Figure 10 was obtained using DNAPlotter (20).

**Matching Microarray Experiments against the Chrono-Transcriptome Reference Curve**

Microarray signals from three DNA datasets (12) were modified according to the D-N-A procedure described above. Signals from the 45 genes considered to be markers of growth-phase (see above) were chosen, plotted against those from each time point from the chrono-transcriptome and the correlations were used as a measure of the relative identity of gene expression between samples. For clarity purposes, results were normalized so that 1 would be the correlation of any sample against itself and 0 the correlation of the two most different samples in this analysis.

**Expression Weighted Properties**

The whole genome expression data obtained through this study was used to infer macroadaptation trends throughout the growth curve. Thus, we were interested in the genome position and direction of genes being mostly transcribed, their codon optimization, and whether there was any general trend regarding the proteins being (putatively) translated. Since nutrient availability/starvation and pH played critical roles in this experiment, we set on to infer whether the first might modulate the molecular weight and the latter the pl of the proteins being produced at each time point. Clearly, the assumption is that transcript levels correlate well with the level of translation of the respective protein at that moment.

The general position of the genes being most transcribed was examined using the genome plotting, as described above. The other whole-cell properties (average protein MW and pl, average Codon Adaptation Index, and Relative Sense Transcription) were calculated for each time point by weighting the associated value of each gene with its expression level. The weighting formulas for each time point are as follows.

Inferred Average pl:

\[
\text{Inferred Average pl} = \frac{\sum_{i=0}^{n} \text{Expression Level}_i \times pl_i}{\sum_{i=0}^{n} \text{Expression Level}_i}
\]
For example, a value of 6.5 means that, at that point in time, cells produce proteins with an average pI of 6.5.

Inferred Relative Average Protein Weight:

\[
\text{Inferred Relative Average Protein Weight} = \frac{\sum_{i=0}^{n} \text{Expression Level}_i \times MW_i}{\sum_{i=0}^{n} \text{Expression Level}_i}
\]

Where \( n \) is number of genes spotted on the microarrays. A value of 1.10 means that, at that time point, cells produce proteins with a Mw that is on average 10% higher than the average of all proteins encoded in the genome of \( L. \text{lactis} \) MG1363.

A gene CAI varies between 0 and 1, depending on whether the gene uses rare codons or frequent codons, respectively (21). On average, genes in the \( L. \text{lactis} \) MG1363 genome have a CAI of 0.695. An expression weighted CAI of, e.g., 0.715 means that the organism is transcribing genes with a better than average codon usage.

Relative Transcription from Leading Strand:

\[
\text{Relative Transcription from Leading Strand} = \frac{\sum_{i=0}^{n} \text{Expression Level}_i}{\sum_{i=0}^{n} \text{Expression Level}_i}
\]

Where \( i \) represents genes for which the transcription sense strand is the same as the replication leading strand and \( n \) the total number of those genes and where \( j \) represents genes for which the transcription sense strand is the same as the replication lagging strand and \( m \) the total number of those genes. Only genes spotted on the DNA microarrays were considered. A value of, e.g., 1.20 means that genes coded on the leading strand at a certain time point are on average 20% more expressed that the other genes, even after normalizing for their genomic overrepresentation (1863 versus 439).
References


