Chapter 5

The chronotranscriptome of Lactococcus lactis reveals extensive reprogramming of gene expression during growth

Rutger W.W. Brouwer, João P.C. Pinto, Araz Zeyniyev, Sacha A.F.T. van Hijum, Jan Kok, and Oscar P. Kuipers

Abstract

The lactic acid bacterium Lactococcus lactis has been the subject of numerous gene expression studies. Most of these have focused on determining the effects of specific growth conditions or mutations on the gene expression in this bacterium. The natural variations in gene expression during growth of L. lactis have thus far not been thoroughly investigated. Here, we present an unprecedented densely sampled DNA microarray time-course of L. lactis subsp. cremoris MG1363 grown in batch culture in the complex medium GM17.

The resulting dataset was analyzed using various bioinformatics approaches. Correlations between the expression of genes throughout growth were investigated using Pearson’s correlations. Within the exponential and stationary growth phases, sub-phases were distinguished in which the samples exhibited highly correlated gene expression. Genes differentially expressed between these sub-phases were identified and used in COG, GO and metabolic overrepresentation analyses, which yielded novel insights into the transcription patterns during growth of the widely studied L. lactis strain MG1363. This dataset provides a valuable resource to researchers studying gene expression in this organism and in related bacteria.
Introduction

Lactic acid bacteria (LAB) are of high industrial relevance as they are used in the production of a host of fermented foods and feed, among which many dairy products such as yoghurts and cheeses. The LAB comprise several genera of Gram-positive bacteria, including the Lactococci, Streptococci, and Lactobacilli. They derive their name from the fact that they all produce lactic acid as the main end-product of sugar metabolism. The production of lactic acid lowers the pH of the environment, thus preventing food spoilage by other bacteria and by fungi. In the laboratory, several model LAB organisms are used, such as Lactococcus lactis, Streptococcus thermophilus and Lactobacillus plantarum.

Many different strains of the two subspecies of L. lactis have been isolated over the years and the genome sequences of four of these are publicly available 6–9. Two of these strains, L. lactis subsp. lactis IL1403 and L. lactis subsp. cremoris MG1363, are used worldwide as model organisms. Efficient genetic tools have been established, such as gene knock-out 91 and (over) expression systems 92. Sequencing the genomes of these two bacteria has enabled the expansion of this repertoire of tools with transcriptomics, proteomics and metabolomics techniques 93–95. The genome sequences also allowed the development of a genome-scale metabolic model for L. lactis MG1363 and L. lactis IL1403 96,97. These techniques allowed elucidating many aspects of the cellular biology of L. lactis, including the kinetic parameters of a number of enzymatic pathways 98–100, the proteins involved in specific regulons 101–103 and stimulons 104,105, as well as the stability of messenger RNAs 17.

L. lactis, being a model for the LAB, has been the subject of many transcriptomic studies and we now have a clear picture of the regulons of the major transcriptional regulators operative in this bacterium. Most regulons have been elucidated using genome-wide DNA microarray studies in combination with genetic perturbations of the regulators involved 101,102,106–111. In these studies, the differences in gene expression were determined at a single time-point during growth, mostly in the exponential phase. These studies did not determine during which time-points in growth the regulon members were actually expressed. Zomer et al. performed a short time-course transcriptomics (chrono-transcriptomics) experiment on carbon catabolite repression in L. lactis MG1363 103. Samples of a wild-type and a CcpA deficient strain were taken at four points in time and compared: two in the exponential phase of growth, one at the transition point between the exponential and stationary phases and one approximately 6 hours into the stationary phase 103. It was observed that the effects on gene
expression of the global regulator (CcpA) differed between growth-phases. Differences were observed in the expression of genes involved in carbohydrate, amino acid and nucleotide metabolism. From a total of 422 genes only 3 genes were identified to be differentially regulated at all four time-points\(^\text{103}\). More recently, de Jong \textit{et al.} performed a chrono-transcriptomics analysis on \textit{L. lactis} MG1363 growing in milk\(^\text{112}\). Gene expression was measured at 12 points in the growth showing substantial differential expression during this period. Using the temporal gene expression and other data, they were able to reconstruct parts of the active genetic network of \textit{L. lactis} growing on milk.

Due to cost and time considerations, performing chrono-transcriptomics to study the regulon of each regulator based on comparison of a knock-out / overexpression with the wild-type strain in \textit{L. lactis} would be near to unfeasible, also because it is impossible to predict the required number of samples and their optimal timings, as these factors are dependent on the biological role of the particular transcriptional regulator \(^\text{10}\). Transcriptional regulators directing the expression of many genes such as CcpA or CodY can have different roles throughout the growth \(^\text{102,103}\). The times at which they are active are not known before measuring the expression of the genes they regulate. A more effective way to determine the changes in gene expression is to perform a transcriptomics time-course experiment with a high temporal resolution \(^\text{113}\). Such studies provide invaluable insights into both the biology of an organism and the protein-encoding potential during growth \(^\text{113}\). Furthermore, this information may help to extend known pathways by determining genes not previously associated with a pathway that have similar gene expression patterns to those that are part of a pathway through the guilty-by-association rule \(^\text{29}\).

Here, we present a densely-sampled DNA microarray time-course in which transcription of genes of \textit{L. lactis} MG1363 was followed during growth under standard laboratory conditions, namely as a standing batch culture at 30 °C in rich M17 medium containing 0.5% w/v glucose. Samples were taken every 15 minutes for 12 hours, during which the culture did not reached the stationary phase. Samples were taken at 24, 32 and 48 hours to characterize the late stationary phase. The data obtained from this chrono-transcriptomics experiment furthers our understanding of the gene expression patterns in \textit{L. lactis} MG1363 during growth in a complex medium. We have analyzed the gene expression data using correlation and functional overrepresentation analyses. The dataset generated in this study is a rich resource for the LAB research community as it can be used to determine the timing of gene expression of vital processes in \textit{L. lactis}
MG1363 and might be used to predict gene expression timing in other related bacteria.

**Results**

**The growth of L. lactis MG1363**

*L. lactis* MG1363 was grown as a batch culture at 30°C in rich M17 medium with 0.5% (w/v) glucose (GM17) in a 12-L fermentor under modest stirring at 30 RPM, to prevent settling of the cells. During the first 12 h of growth, 45 samples of 50 ml each were taken at 15-min intervals. Additionally, samples were taken at 24, 36 and 48 h after inoculation to monitor gene expression in the culture during the late stationary phase. For each time-point (tp), both the optical density at 600 nm (OD₆₀₀) and the pH of the culture were recorded (Fig. 1). For selected samples taken during the exponential growth phase, the concentration in the medium of free glucose was determined (Fig. 1). During the 48-h monitoring period, the culture proceeded through all of the classical growth phases (Fig. 1). Gene expression was not determined in the lag phase as cell densities were very low during this period of growth and sample volumes in excess of 0.5 L would have had to be processed to obtain sufficient RNA. We strived to minimize the lag phase by inoculating the medium with a culture of exponential-phase cells grown in the same batch of GM17 that was used for the fermentation. The cells in the inoculum thus needed minimal adjustments to their new environment.

It is evident (Fig. 1) that the culture enters the stationary phase at tp 19, placing the transition from the exponential to stationary growth phase between tps 18 and 19, 6 h and 15 min after inoculation. Interestingly, 2 periods are observed in the exponential growth phase based on the growth-rate of *L. lactis*. Up to tp 12, the cells in the culture grow exponentially, as one would expect in this phase of growth. After tp 12, the growth rate steadily decreases until the culture enters the stationary phase (Fig. 1). This trend in the OD₆₀₀ is mirrored by the development of both the pH and the glucose concentration in the medium. These observations suggest that (a subset of) the *L. lactis* cells have sensed the trigger(s) that ultimately lead the entire culture to enter into the stationary phase approximately 1.5 h prior to the transition to the stationary phase. This period from tp 13 to 19 seems to represent a transition phase between the exponential growth and stationary phase.
After the transition phase, the OD$_{600}$ of the culture is maintained at the same level for at least 6 h (Fig. 1). In the first sample taken in the late stationary phase (tp 43, 24 h after inoculation), the OD$_{600}$ had significantly decreased. The decrease in OD$_{600}$ continued until the end of the experiment at 48 h after inoculation (Fig. 2) and was accompanied with a slight rise in culture pH.

At each of the tps indicated in Fig. 1, samples were taken from the culture and the genome-wide expression of genes was assessed using two-dye DNA microarrays. Total RNA of each sample was assayed on three different DNA microarray slides and dye swaps were taken to reduce technical bias. The entire procedure yielded 6 expression signal values per gene per tp, all of which were subsequently analyzed using the approaches described below.

Fig. 1 Growth of *L. lactis* MG1363 in GM17 under standard laboratory conditions.

The optical density at 600 nm (OD$_{600}$), glucose concentration in the medium and extracellular pH of the *L. lactis* MG1363 were followed in time. Samples in the right panels: samples taken at 24, 36 and 48 h after inoculation.
The vertical dashed lines indicate the boundaries between the exponential growth phase, transition phase (bold dashes), and the first and second stationary phase. The late stationary or death phase is shown in the separate panels at the right of the figure. The x-axes show the time after inoculation.

**Experimental design and technical replicates**

The gene expression signals obtained from the DNA microarrays were normalized and scaled as described earlier. In this procedure, intra-slide normalizations were performed using the LOWESS method while PreP scaling was employed for normalization between slides over the complete dataset. A large difference between this and previous studies is that the current study measures gene expression at multiple points in time. To be able to use the above-described data processing methods, a comprehensive experimental design was devised (Suppl. Fig. 1); a labeled c-DNA sample from each tp was hybridized with that of the previous tp on a DNA microarray slide and on another slide with the labeled c-DNA of the next tp samples (loops). As an internal control, each c-DNA was also differentially labeled and hybridized on a DNA microarray slide with the c-DNA of samples taken 1 h earlier or later (hops). This hybridization scheme enabled using of both regular analysis methods, LOWESS and PreP, and is more cost-effective than a design in which a common reference is used. Correlation analyses were performed between the gene expression levels for each tp from the various slides (technical replicates; Fig. 2). Replicate gene levels of the same tp samples are expected to be highly similar to each other and thus to have a high correlation. Indeed, this expectation is met in many cases, showing that the normalization and scaling procedures were appropriate for this experimental design (Fig. 2).

Especially for the tps up to the transition point, high correlations are observed between the technical replicates. From tp 18 onwards, the gene expression levels of the replicates become less comparable to each other and a clear difference is observed between the datasets originating from a loop comparison and those from a hop comparison. This difference is likely caused by the LOWESS normalization, which assumes that the expression of approximately 50% of the genes does not change when comparing gene expression in two samples. This premise is most probably not true when comparing samples from different growth phases, which could happen in the hop comparisons as
these samples are taken 1 h apart. As this may lead to the introduction of artifacts and false trends, the gene expression levels obtained from hop comparisons were not taken into account in the further analyses, unless explicitly stated. For tps 30 and 31, low correlations were observed for all cross-slide replicates (Fig. 2). This is most likely caused by a hybridization issue that occurred on a single slide (slide no. 188275), containing the data for both tp 30 and 31. This dataset did not correlate well to that obtained from the other slides on which these samples were hybridized nor did they show sufficient resemblance (correlation > 0.9) to any of the other samples taken in the stationary phase (data not shown).
Visualization of Pearson's correlations between the replicate datasets. The color of the dot indicates the type of comparison: loop-to-loop, loop-to-hop or hop-to-hop (Suppl. Fig. 1). Correlations below a value of 0.8 or on the sample diagonal were omitted for clarity. Each sub-graph contains the expression levels obtained for a single tp (grey box). The correlations in the expression levels per replicate have been mapped to the color intensity of the tiles and the comparison type is indicated with the points in the tiles. A black square with a red central dot indicates a correlation of 1 between 2 replicates for the same tp. The diagonals of the sub-plots indicate the correlation between the same replicate datasets and have been left blank.

Fig. 3 Signal distribution over all the datasets. Signal distributions over the “loop” datasets are represented as boxplots per tp. Individual points are the 5% outlying values from the distribution; bars indicate the
5 and 95% quantile values. The boxes represent the 25 and 75% values of the distribution. Median values are indicated in the boxes.

**Gene-expression throughout growth**

From earlier experiments it is known that many genes are differentially expressed between different growth phases\textsuperscript{103}. On the other hand, only little change in gene expression is expected to occur within a particular growth phase. Genes of whom the expression changes within a certain growth phase are likely to be part of distinct metabolic or regulatory pathways. The dataset presented here allows employing an alternative method to test these assumptions on a genomic scale. To this end, the similarity in gene expression between the samples was determined using the Pearson’s product moment correlation method. This analysis clearly shows that there is little similarity between gene expression in samples taken from the exponential growth and that in samples from the stationary phase (Fig. 4). Substantial variation in gene expression is also observed within both of these growth phases, allowing defining several growth sub-phases with highly similar gene expression patterns.
Fig. 4 Correlation in gene expression between the time-points
Average Pearson's product moment correlations between the gene expression measurements. The tps are indicated in the row and column labels. Each of the large squares contains the correlations for all the replicates for a single time-point. The fill-colors in the squares show the correlations of each measurement, as defined to the right of the figure. Correlations below 0.80 are shown in gray.

The correlation between tps in the exponential growth phase was at least 0.8, which is indicative of the co-expression of many genes throughout this phase of growth (Fig. 4). When the correlation matrix is inspected in more detail, several sub-phases, formed by tps 1-4, 5-8,
and 9-12, are observed in which the correlation exceeds 0.9 (Figs. 4 and 5). The exact boundaries of these periods of highly similar gene expression are not known. The gene expression patterns in the exponential growth are similar to those of samples in the transition phase (tps 13-18).

![Comparison between adjacent timepoints](image)

**Fig. 5** Number of differentially expressed genes during the time-course

Up- and down-regulated genes are plotted as a function of the tps. Differential gene expression was determined by calculating the ratio between the mean gene expression in one tp and that in the next tp. A ratio larger than 1 will thus mean that the gene expression signal at tp_n is greater than that of tp_{n+1}. Genes with at least a 2-fold average ratio were considered to be differentially expressed. The red points indicate the number of down expressed genes of which the expression was down regulated, the blue points show the number of up regulated genes between the time-points.

Between the transition and stationary phases (between tps 18 and 19) many genes are differentially expressed, as can be seen in both the correlation analyses (Fig. 4) and the numbers of genes of which the expression changed over 2-fold (Fig. 5). However, the number of differentially expressed genes between these phases is smaller than that between the first and second sub-phases of the exponential growth (Fig. 5). From the start of the stationary growth (tp 19) to tp 25, gene expression remains relatively stable. However, gene expression in the samples taken after tp 24 shows little correlation with that in earlier samples (Fig. 4) and many genes are down-regulated between tps 24
and 25 (Fig. 5). A period with highly similar gene expression follows and extends to tp 40 (Fig. 4 and 5). After tp 40, many genes are down-regulated and fewer genes are expressed than in earlier tps (Fig. 3). This period extends up to the end of the experiment, 48 h after inoculation.

The differential gene expression patterns throughout this chronotranscriptomics experiment are most probably caused by the changes in the environment as a consequence of bacterial growth. Based on this assumption, it is expected that genes part of a certain biological pathway to behave similarly in time. To gain insight into these pathways, overrepresentation analyses were performed on various gene classifications: clusters of orthologous genes, (COG 85), gene ontology (GO 114) and Kyoto encyclopedia of genes and genomes metabolic (KEGG 86). These classification schemes are based on different data sources and should provide complementary information on the transcriptional reprogramming of L. lactis MG1363 throughout growth.
Fig. 6 Time-resolved overrepresentation of COG functional classes.
Overrepresentation of COG classes is plotted amongst the up- and down-regulated genes throughout this time-course (Fig. 5). Overrepresentation analysis was performed using Fisher’s exact tests. The resulting p-values from these tests were log transformed using –log10 transformation for plotting purposes (tile fill color). All classes with a p-value below 0.01 are highlighted with a blue border.

The COG classification encompasses a total of 20 classes, offering a broad overview of the biological processes that are differentially expressed throughout growth of *L. lactis* MG1363 in GM17. From statistical overrepresentation analyses on both the up- and down-regulated genes per tp it is clear that only a few processes are differentially expressed between any two subsequent tps (Fig. 6). Most differentially expressed genes are associated with only 7 COG classes, namely inorganic ion metabolism (P), energy production and conversion (C) post-translation modification and chaperones (O), transcription (K), translation and ribosomal structure (J), nucleotide (F) and amino acid transport and metabolism (E). These classes are very broad; they encompass most of the processes that are expected to change during growth. Due to the breadth of the COG classification, it is unclear whether the same or different pathways are differentially expressed at different points in time. To answer these questions, GO and KEGG analyses were also performed (see below). Another interesting observation was made for the period 12 to 24 h after inoculation (tps 42 - 43); at 12 h after inoculation (tp 42) the expression of 50 genes are over 2-fold higher than at 24 h after inoculation (tp 43) (Fig. 5). Among these are many that encode ribosomal proteins, indicating that the translation machinery undergoes changes during this time. Adaptations in the ribosomal content in *L. lactis* MG1363 in the stationary phase, were not observed in previous studies.

To further pinpoint the biological processes differentially expressed during *L. lactis* MG1363 batch fermentation, a GO overrepresentation analysis was performed on the up- and down-regulated genes. A total of 290 overrepresented categories were obtained without multiple testing correction among the 606 down-regulated genes, while 258 GO classes were overrepresented among the 751 up-regulated genes (for both analyses: *p*-value < 0.05) (Fig. 7). The reason for these high numbers lies at least partly in the nature of the GO classification: it contains many classes that represent essentially the same process, making GO annotation a less suitable tool for initial analysis of high-
density chrono-transcriptomics datasets. Nevertheless, combining the GO annotation and COG classification results yielded valuable insights into the differentially regulated processes occurring in *L. lactis* MG1363, some of which will be detailed below.

Fig. 7 Number of overrepresented GO categories per time-point. The number of overrepresented GO categories present among the 2-fold up- or down-regulated genes (Fig. 5). A threshold of 0.05 was used for the *p*-values. The boundaries of the growth-(sub)phases are indicated with the dotted lines.

**Nucleotide metabolism**

Nucleotide transport and metabolism were overrepresented among both the up- and down-regulated genes throughout exponential growth (Fig. 6; group F). In the time-point comparisons where this class was overrepresented, 58 and 57 GO categories were overrepresented amid the up- or down-regulated genes, respectively. To determine what GO categories co-expressed with nucleotide transport and metabolism in *L. lactis* MG1363, GO terms which were overrepresented multiple times were filtered, leaving 23 and 21 GO categories, respectively, among the up- or down-regulated genes that were primarily differentially expressed between tps 7-8, and tps 13-14. Of these GO categories, 19 overlapped (Fig. 8). It is immediately clear that the COG class nucleotide transport and metabolism associates with a diverse set of GO categories including very broad ones such as “primary metabolic process” as well as with very specific classes e.g., “pyrimidine nucleotide biosynthesis..."
process”. Due to tree-like relations between the GO classes, overrepresentation of a broad class will often be caused by the overrepresentation of a more specific sub-category. The more specific GO categories associated with nucleotide transport and metabolism are biological processes centered around the four compounds pyrimidine, purine, glutamine and arginine, the latter two of which share precursors with both pyrimidine and purine, which explains the co-regulation between these pathways (Fig. 8).

![Fig. 8 Overrepresented GO-categories relevant to nucleotide metabolism.](image)

Indicated are the GO categories that were overrepresented in the up and down regulated genes (Fig. 5) on more than one occasion within the COG class nucleotide metabolism. The boundaries of the sub-phases are indicated with the dotted lines. The bold dotted line indicates the end of the transition phase. All classes shown here have p-values below 0.05. The p-values are indicated with the fill intensity of the tiles and their direction with the color.

A clear trend is observed between genes associated with purine and pyrimidine metabolism (Fig. 8). When pyrimidine biosynthesis genes are overrepresented among the down-regulated genes, purine biosynthesis genes are up-regulated. This inverse relation holds for
most of the exponential and the transition phases and is even seen at the very short 15-min time span covered by tp 7 to tp 8. At tp 8 when the expression of pyrA, carA, pyrDA, llmg_1089, pyrK, pyrDB, pyrF, pyrC, pyrE is more than doubled compared to the expression levels at tp 7. In the same period the expression levels of purC, purQ, purM, purH, purD, and purK are at least halved (Fig. 8; Suppl. Fig. 2). The nucleotide-transport-and-metabolism-associated GO classes clearly show these fast changes in gene expression. The processes involving arginine and glutamine do not follow such an apparent tendency other than that they are often overrepresented in those samples in which the purine- or pyrimidine-associated genes are also overrepresented (Fig. 8). Another interesting observation is that the statistical overrepresentation of the broader GO classes, such as nitrogen compound metabolism (GO:0006139) and the nucleotide biosynthesis process (GO:0009165), behave similarly in time to the pyrimidine nucleotide biosynthesis process instead of mimicking those processes associated to purine biosynthesis (Fig. 8). This is easily explained when looking at the genes contained in the category nitrogen compound metabolism (GO:0006139); 10 of the 13 genes therein are actually pyr genes. The others are guaB, add, and rdrB. Of these three genes the products of guaB and add are also linked to nucleotide metabolism. The rdrB gene is a transcriptional regulator of which the regulon is unknown.

The GO categories associated with nucleotide transport and metabolism are overrepresented among both the up- and down-regulated genes in the exponential growth phase (Fig. 8). This observation indicates that expression of the genes underlying these categories shows most variation during this phase of growth, in which the demand for nucleotides is the largest. Tight regulation of the pathways involved seems to be required to ensure that the cells have sufficient nucleotides to continue to rapidly grow and divide. After exponential growth, the nucleotide-associated processes are no longer overrepresented indicating that these processes are less important in the ensuing phases.

**Amino acid transport and metabolism**

Among the other COG processes overrepresented in the exponential growth-phase is that of amino acid metabolism (COG E) (Fig. 6). Overrepresentation of amino acid metabolism partially overlaps with that of nucleotide transport and metabolism (COG F) due to the arginine and glutamine metabolic processes (Fig. 8 and Fig. 9). The GO overrepresentation analyses clearly show that the pathways for serine (GO:0009096), cysteine (GO:0019344), glutamine (GO: 0009084) and
arginine (GO: 0006525) biosynthesis are down-regulated after the early exponential growth phase and those associated with glutamine and arginine are then again up-regulated at the start of the transition phase (Fig. 9; tps 13-15).

More GO terms are overrepresented together with the COG E class of amino acid transport and metabolism. The genes associated to these GO categories, such as the carboxylic acid biosynthesis (GO:0046394) and cellular nitrogen biosynthesis (GO:00444271) processes, form proteins that supply amino acid metabolism with the compounds it requires. After the down-regulation of genes related to amino acid transport and metabolism (COG E; Fig. 6) at tp 5, the expression levels of the genes encoding the components of these supporting pathways also stay stable. This observation suggests that, during growth of L. lactis (in GM17), expression of these pathways is coupled to that of amino acid metabolism.
Fig. 9 GO classes associated with amino acid transport and metabolism
GO classes overrepresented with the COG class E for amino acid transport and metabolism (Fig. 6). All GO categories depicted here are overrepresented with a p-value smaller than 0.05. The numbers of differentially expressed genes associated to each pathway are indicated by the color of the border of the tiles. The intensities of the tiles indicate the p-value while their colors indicate whether the class was overrepresented amongst the up or down expressed genes. The borders of the tiles indicate the number of genes.
GO classification is not the only source with fine-grained annotations of amino acid metabolism. KEGG contains metabolic pathways including those concerning amino acids. It also includes the pathways for amino acids such as proline and methionine, which are not present in the GO classification. The KEGG overrepresentations are in line with the GO analyses (Fig. 10). The pathways for arginine and cysteine are overrepresented in the up and down regulated genes (Fig. 5) in the same time-points as in the GO analyses. More amino acid pathways were identified with the KEGG overrepresentation analysis. Amid the up-regulated genes in the transition phase (tps 13-18), the overrepresentation of KEGG pathways additionally identified the pathways for alanine, aspartate and glutamate, and lysine biosynthesis.

Fig. 10 Overrepresented KEGG pathways associated with amino acid metabolism
Indicated are those KEGG pathways that are overrepresented at the same time-points as the COG class E of amino acid metabolism and transport. All overrepresented pathways depicted here have a $p$-value below 0.05. For figure legends see Fig. 9.

By combining the results of the COG classifications with the GO and KEGG annotations, a more complete picture is obtained as to which amino acids are limiting at which point in time during the growth of $L$. 74
lactis as a batch culture in GM17. A demand for a broad range of amino acids seems to be present at the onset of the exponential phase (tp 1-5). Near the end of exponential growth, as the growth rate decreases, amino acid biosynthesis is up-regulated (Fig. 9 and 10). Similar to nucleotide biosynthesis, the expression of genes related to the amino acid production is highest in the transition phase, where amino acid and nucleotide availability seems to become limited. The results also suggest that at tps 6, 8, and 17 the uptake of from the complex GM17 medium was unable to fully meet the demand for these nutrients. The expression levels for the genes encoding the secondary amino-acid transporters 116 support this explanation. The genes for 7 of the 9 transport systems were continuously expressed during the exponential growth indicating that amino-acids were taken up from the medium at this time (Suppl. Fig. 2). Only the gene for lysQ (now known as hisP) was expressed throughout the transition phase.

**Processes overrepresented in the stationary phase.**

The only COG class overrepresented in the stationary phase was that of transcription (COG K) (Fig. 6 and 7). A total of 117 GO terms were overrepresented among the up- and down-regulated genes in the stationary phase. As expected, most of these GO classes involve down-regulated genes, since growth has ceased and, as a consequence, only few pathways are apparently expressed during this phase.

Most classes that are overrepresented in the exponential growth phase are also overrepresented at the boundary between the stationary and the transition phases (Fig. 11). Even-though these processes are statistically overrepresented with a p-value below 0.05, only one or a few genes in these categories are actually differentially expressed between the transition and stationary phases (Fig. 11). The GO categories of which larger numbers of genes are up-regulated seem to encompass many other processes. The categories include cellular biosynthetic process (GO:0044249) and cellular metabolic process (GO:0044237; Fig. 11). Their statistical overrepresentations may indicate that the culture is performing renewal processes for specific pathways.
Fig. 11 GO categories overrepresented among genes up-regulated in the stationary phase. For figure legends see Fig. 9.

**Ribosomal gene expression**

The expression of 20 of the 56 genes encoding ribosomal proteins lowered at least two-fold after the exponential growth (tps 12-13; Fig. 5 comparison 12, up). In this study, no large changes in the expression of the ribosomes and their related genes were expected. The ribosomal proteins are required for the assembly of the ribosomes via one of several pathways (for review see117). Not all of these proteins are essential as *Escherichia coli* is known to be viable even when certain ribosomal proteins have been disrupted (for review see117). The small
30S subunit of the 70S ribosome consists of 1 ribosomal RNA (16S) and 21 proteins, while the large subunit (50S) is formed by 2 ribosomal RNAs (5S and 23S) and 34 proteins. It is possible that the protein composition of the ribosome changes during growth. By clustering the expression profiles of the ribosomal genes, we can gain insights into which ribosomal proteins are predominantly transcribed in which growth phase (Fig. 12).

To cluster the expression profiles of the genes encoding ribosomal proteins, hierarchical clustering was performed using average linkage and a Pearson’s correlation based distance measure. The results were visualized using dendograms in which the distance between nodes is indicated on the y-axis (Fig. 12). Clusters were obtained by imposing restrictions on the maximum distance between nodes. When the maximum distance between nodes in the hierarchical clustering is set to 0.2 (Fig. 12 top, dotted line), the ribosome-associated genes form 9 clusters of which 4 contain only a single gene (Fig. 12). The largest cluster (cluster 4) is characterized by a steady expression of the constituting genes up to the mid-exponential growth phase (Fig. 12; *rplV*). Afterwards, gene expression declines to a relatively stable level in the stationary phase and a final decline after 24 h. The genes in cluster 2, of which *rpmB* is presented as an example, show a profile similar to that of *rplV* but without the decline in expression level after 24 h. This cluster could be expanded with the *rplQ* and *rplI* genes as these show similar expression profiles in all growth phases except for the late exponential phase. The genes in cluster 3 have less consistent gene expression patterns (Fig. 12; *rplE*). All of these genes show an increase in expression in the stationary phase. In some cases, the peak of expression in this phase is higher than that in the exponential growth phase (e.g. for *rplE* and *rpsR*).
Fig. 12 Ribosomal gene expression
The top panel contains a hierarchical clustering of the gene expression profiles of ribosomal protein encoding genes. The ribosomal complex in which the products of these are present are added to the gene name. The proteins of which the localizations are not known are indicated with a *. This clustering was performed with average linkage and the Pearson’s distance measure. The bottom panel contains expression profiles of the clusters.
Not much is known about the specific physiological function(s) of individual ribosomal proteins and therefore we can only speculate on the effect of the differential expression of the encoding genes on the ribosome composition. The expression profiles of the ribosomal protein genes suggest that the protein composition of the \textit{L. lactis} ribosomes may change during growth. The effect is observed for genes of proteins of both ribosomal subunits. Changes in ribosome composition may play a role in protein translation, and/or in growth rate. Previous studies performed in the gram-negative bacterium \textit{Escherichia coli} have implicated ribosomal proteins in ribosome hibernation in the stationary phase \textsuperscript{118,119}. Similar processes are also likely to occur in \textit{L. lactis} MG1363. Further genetic experiments involving genetic knock-downs and gene over-expression strains combined with proteomics studies should shed further light on these findings.
Fig. 13 Expression weighted pIs, codon adaptation indices and molecular weights.
The expression weighted isoelectric points (pI), codon adaptation index (CAI) and molecular weight (MW) were determined using BioPerl. The end of the transition phase is indicated by the dotted vertical line. The variation in the sample is indicated with vertical lines. The horizontal lines show the averages of the factors in the genome.

**Expression-weighted gene properties**

The chrono-transcriptomics data were also used to uncover possible trends in the physical properties of the protein products of the genes expressed throughout growth. To this end, the codon adaptation indices (CAI) were determined for all genes, as were the iso-electric points (pI) and molecular weights (MW) of their putative proteins. The values were weighted using the gene expression levels at a given time-point (Fig. 13).

As is clear from figure 13, the three weighted properties show different trends in time. The expression-weighted pI decreases from approximately 6.5 to 6.2 just after the transition phase and seems to follow the culture pH (Fig. 1). The values for the weighted pIs are much lower than the mean pI of all proteins encoded in the genome of *L. lactis* MG1363 (6.77). The apparent congruity between both parameters is quite striking and may suggest physiological importance. However, without accurate measurements of the cytoplasmic pH the importance of these findings cannot be judged.

The codon adaptation index (CAI) is used gene-wide to determine whether a gene contains frequently used codons (~1), or codons that occur only rarely in the organism (~0). The expression-weighted CAI is mostly constant from the exponential growth phase up to the mid-stationary phase (Fig. 13). After tp 25, it quickly declines to a new plateau of 0.692, a value very close to the average CAI (0.691). After tp 38, the expression-weighted CAI steadily declines to a value of approximately 0.680. The expression-weighted CAI profile seems correlated to the number of differentially expressed genes (Fig. 5). Throughout the exponential growth and the early stationary phase, up to tp 24, the expression-weighted CAI shows little variation. After the down-regulation of many genes at tp 24 (Fig. 5; ~120 genes over 2 times down-regulated), the expression-weighted CAI reaches a lower and stable plateau. At tp 39, another peak in down-regulation of gene expression is observed (Fig. 5) while at the same time the expression-weighted CAI gradually decreases further.
The expression-weighted molecular weights are quite constant up to the end of transition phase. After that, a slow but steady decline is observed. After 24, 36, and 48 h after inoculation, the expression-weighted molecular weights remain stable at a slightly higher level than at 12 h after inoculation. The findings suggest that, on average, the culture expresses more genes that encode smaller proteins in the stationary growth-phase than in the exponential growth.

The profiles of each of the expression-weighted properties differ considerably. Each individual property shows gradual changes as a function of time, suggesting that the fluctuations are not random and that *L. lactis* might tune the physical properties of its proteins to the changing environment.

**Discussion**

The high density of sampling accomplished in this chronotranscriptomics experiment, combined with in-depth bioinformatics analyses, allowed clearly distinguishing the various growth phases that *L. lactis* MG1363 undergoes during batch fermentation and determining the transient expression of large numbers of genes and the regulation of cellular processes.

Analysis of technical replicates revealed that the experimental design used allowed reliably detecting changes in gene expression throughout growth. Although the accuracy in the samples after tp 26 may be somewhat less than in the exponential growth phase, as the correlation between replicates decreased, it was still sufficient for determining changes in gene expression. The slight decrease in accuracy is probably caused by fewer genes being expressed and lower variations in gene expression at these later stages in growth. The loop design allowed doubling the number of sampling time-points in comparison to what would be needed in a common reference design. The “hop comparisons” did not contribute greatly to this dataset as they proved to be unreliable at later time-points.

Correlation analysis disclosed the existence of various periods of highly similar gene expression during *L. lactis* growth. Upon transit from these periods, many genes were differentially expressed. By performing functional overrepresentation analyses on these transits, large numbers of differentially expressed pathways and biological processes were discovered. Two processes that were clearly differentially expressed throughout exponential growth were the purine and pyrimidine biosynthesis pathways. The direction of differential expression of one of the two pathways is always opposite to that of the other. The time-points at which the purine and pyrimidine
pathways are differentially expressed overlap with the differential expression of the glutamine and arginine biosynthesis pathways. However, aside overlap in timing of differential expression between these pathways, there is not a correlation in the direction of expression. A relation between the purine, pyrimidine, glutamine and arginine pathways is not unexpected as the purine and pyrimidine pathways produce and require metabolites that are also necessary for the production of these two amino-acids. The observed transient expression of the pur and pyr pathways seems rather striking as pyrimidine, purine as well as amino acids are supposedly in high demand throughout exponential growth. Thus, one might expect these pathways to be continuously expressed throughout the exponential phase. From the chrono-transcriptomics data it is clear that the expression of these pathways is tightly controlled and limited to specific intervals during growth. The window of expression of most purine and pyrimidine genes is at most 30 min. In this time frame apparently sufficient nucleotide biosynthesis capacity is provided to allow the culture to reach the stationary phase. Due to this spiky expression pattern of the pur and pyr genes, a small difference in the timing of sampling of two cultures to be compared by DNA microarray analysis could easily result in a many-fold difference in the expression of these genes and explains why the members of these two pathways are often reported to be differentially expressed in single time-point perturbation studies.

Using functional analyses, several processes were statistically overrepresented amongst up-regulated genes in the stationary growth phase. Most of these overrepresentations were based on only one or a few genes. The processes that were overrepresented with larger numbers of genes were general GO categories containing large numbers of genes, such as cellular biosynthetic process (GO:0044249) and cellular metabolic process (GO:0044237). The absence of more specific GO categories in combination with the general GO categories might suggest that there are groups of co-regulated genes that are not yet present in the GO annotation for L. lactis MG1363.

Near the end of the measurement period, the culture seemed to stock up on intracellular macromolecules as GO categories associated with cellular polysaccharides biosynthesis (GO:0033692) were overrepresented among the up-regulated genes. This suggests that the cells are preparing for long-term survival under these conditions. The expression of genes in these pathways was highly transient, suggesting that the cells rather store energy in these macromolecules than spend it on the synthesis of other enzymes and proteins that aid in the uptake of nutrients from the environment. In addition to these pathways, GO
categories representing with lipo polysaccharides (GO:0008653), lipid A (GO:0008610), glycolipid (GO:0009247) biosynthetic processes were differentially expressed in the late stationary phase. We currently have no hypothesis explaining these fluctuations in gene expression.

Ribosomal protein genes were highly differentially expressed throughout growth. Through clustering analysis at least 2 groups of ribosomal proteins could be identified on the basis of the expression patterns of their genes. The gene expression patterns might indicate that the protein composition of the ribosomes of L. lactis changes during growth. This supposition is complementary to other studies that suggest that, although most ribosomal proteins are essential, some might only offer a growth advantage under certain conditions. The pronounced difference in the expression of ribosomal proteins is unlikely to occur without a functional role, but its elucidation is beyond the scope of this study.

The chrono-transcriptome presented here was used to determine trends in the physical properties of the expressed genes throughout growth on M17 medium. The observed patterns for the expression weighted iso-electric point, codon adaption index and molecular weight were highly distinctive and in case of the pI and molecular weight clearly not the mean of their properties (Fig. 13). In order to determine the CAI per protein, the relative codon frequencies in the genome of L. lactis MG1363 were used to determine the codon weights. This procedure was necessary as there is no independent set of highly expressed available.

Previously we have performed a chrono-transcriptomics analysis of L. lactis MG1363 growing as a batch culture in milk. The main differences between this and the present study are the choice of medium and the number of time-points tested. GM17 is the most-used medium for growth of L. lactis in the laboratory and as such the data presented here are of eminent importance to the scientific community; the use of milk by de Jong et al. allowed describing many processes relevant for the dairy industry. Comparing the findings from both studies it is evident that many of the processes that are different between the growth-phases are not medium specific. The shorter time intervals between samples in the present study allowed describing gene expression in greater detail. For example, a strong increase in gene expression at tp 8 for the genes responsible for pyrimidine metabolism was missed in the milk chrono-transcriptome (Suppl. Fig. 3). Preliminary analyses show that the chrono-transcriptomes of GM17- or milk-grown L. lactis give comparable results for the expression of genes responsible for several pathways (Suppl. Fig. 2). By further
comparing gene expression in *L. lactis* grown in milk or GM17 may elucidate the genes that are uniquely expressed in either of the media.

In conclusion, this chrono-transcriptome dataset represents a rich repository for researchers working in the fields of both fundamental and applied research in molecular and systems biology of lactic acid bacteria. We believe that the data will provide ample leads for the future study of these prokaryotes as well as provide researchers with the expression patterns of their favorite genes, which will allow them to more precisely judge the behavior of these genes.

**Materials and methods**

**Growth conditions**

*Lactococcus lactis* subspecies *cremoris* MG1363 was cultivated from a -80°C aliquot of the sequenced strain 5,8. These cultivated bacteria were used for the inoculation of the fermentor culture was performed with a total of 0.0025 OD units a of starter culture (1/100 final optical density). The starter culture was growing exponentially at the time of inoculation and on media from the same preparation as the sampled culture. In order to verify that the growth-curve was indeed reproducible the fermentation procedure was repeated 3 times. The samples of which the gene expressions were measured were all obtained from a single fermentation. The inocula and the end-cultures were examined by plating, visual microscopical inspection and by continued growth in microtiter plates. No contaminations were observed in any of these control experiments.

The culture was grown in 12 l. M17 medium (Difco laboratories) supplemented with 0,5% Glucose (Acros Organics) at 30°C in a temperature-controlled fermentor with a total volume of 16 l. To ensure homogeneity of the culture, a mild stirring rate of 30 RPM was maintained and the acidity of the medium was monitored with a pH electrode in the fermentor. The optical density of the samples was determined using at 600nm. Glucose measurements of specific time-points were performed using a glucose measuring kit according to the manufacturer’s instructions.

**RNA isolation**

The equivalent of 10 OD₆₀₀ units or more of culture was taken in duplicate every 15 min from 1 h 45 min after inoculation up to 12 h after inoculation. These samples correspond to time-points (tp) 1 to tp
Three further samples were taken at 24, 36 and 48 h after inoculation. Three duplicates were obtained for each of these samples using 15-min sampling intervals. These samples were labeled tp 43 to tp 51. Cells in the samples were spun down in Greiner tubes using a table top centrifuge for 1 min at 10,000 RPM and 30°C. The cells were subsequently resuspended in 0.5 ml of diethylpyrocarbonate-treated T_{10}E_{1} buffer (pH 8.0) and transferred to 2 ml tubes. These were immediately frozen in liquid nitrogen and kept at -80°C prior to RNA isolation. The subsequent RNA isolation was performed as described previously 25.

**DNA microarray analysis procedure**

The cDNA labeling and subsequent hybridizations were performed as described before 25. DNA microarray slides were scanned using a GenePix Autoloader 4200AL confocal laser scanner (Molecular Devices, USA). Labeled cDNAs were hybridized according to a loop/hop hybridization design in which each sample was hybridized to samples of the previous and of the next time-point in growth (loop) as well as to a sample 3 time-points later (hop) (Suppl. Fig. 1). In this design balanced dye-swaps were performed and up to 6 technical replicates were obtained per time-point. A total of 76 separate hybridizations were performed.

**Normalization and data analysis**

Mean signal intensities were quantified using the ArrayPro Analyzer software (www.mediacy.com; version 4.5.1.). Background intensities were determined per spot with the 'local corners' method. The resulting net signals were normalized and scaled using the MicroPrep software 26,121. The resulting tables were loaded into R and were subsequently analyzed using existing and newly developed scripts 122.

**Data sources**

Gene names and annotations were obtained from NCBI (http://ncbi.nih.gov/) under accession number NC_009004. KEGG mappings were obtained from the KEGG SOAP web-service by following the locus tags for *L. lactis* MG1363 with KEGG organism code llm. GO annotations for *L. lactis* MG1363 were obtained by submitting the uniprot protein identifiers to the EMBL QuickGO webservice (www.ebi.ac.uk/QuickGO/GAAnnotation). Service queries were performed in R using the RCurl package 122.
**Overrepresentation analysis**

To uncover overrepresentation of specific COGs among groups of genes, a contingency matrix was calculated with the number of affected and not-affected genes in the group, as well as the number of (not-)affected genes that were not in the COG group. This matrix served as the input for the Fisher's exact test, which is available through the R base library 122.

For the GO and KEGG overrepresentation analysis, the GOstats package from the Bioconductor project was used 33,123. This package employs a hypergeometric test to determine overrepresentation of GO terms and KEGG maps. Both of these annotation sources are organized in directed graphs that cannot be correctly analyzed using the Fisher's exact test.

**Expression-weighted properties**

In order to determine possible trends in the properties of the expressed proteins, the expression-weighted properties were calculated (Eq. 1). This measure is analogous to calculating the average of protein properties of all the gene products specified by the genome, with the exception that these properties are first weighed according to the expression levels of the corresponding genes, assuming that an increased expression of a transcript directly correlates with an increase in the amount of the encoded protein. The resulting expression-weighted property has the same dimensions and units as the original property and is in that respect equivalent. Standard property calculators from the BioPerl project were used (http://www.bioperl.org). The scripts in which these property calculators are implemented are available from the supplementary website.

\[
P = \frac{\sum_{i=1}^{n} (p_i \times 2^{E_i})}{\sum_{i=1}^{n} 2^{E_i}}
\]

Eq. 1 Expression weighted properties.

An expression-weighted property \(P\) is determined by dividing the sum of the product of the expression \(2^{E_i}\) and property \(p_i\) of a gene by the total gene expression in that dataset. The expression data is set as a power of 2 as the original data was transformed using a log2 transformation. This calculation yields a value with the same range and
dimensions as the property, but it is weighted using the relative expression of a particular gene.
Supplementary materials

Suppl. Fig. 1 DNA microarray hybridization scheme. DNA microarray hybridization scheme used in this study. It consisted of a combined loop (bottom) and hop (top) design. The time-points are indicated in the white ovals. The individual DNA microarray slides are represented by the black rectangles. Cy3- or Cy5-labeled cDNA are represented by the green and red arrows, respectively. For example, the cDNA for tp 6 was hybridized on 3 DNA microarrays. On the first microarray, it was labeled with Cy5 and hybridized with cDNA from tp 5. On the second DNA microarray cDNA from tp 6 was labeled with Cy3 and hybridized with cDNA from tp 7. On the third DNA microarray, cDNA from tp 6 was labeled with Cy5 again and hybridized with Cy3 labeled cDNA from tp 3. The first 2 hybridizations are called loops and the last hybridization is a hop as it skips 2 tps. The cDNA from tp 6 is not hybridized to cDNA from tp 9 as the hops were not dye-balanced.
Suppl. Fig. 2 The expression of secondary amino acid transporters during growth on M17 medium. The genes identified as llmg_1452, lysQ, llmg_2011, and llmg_2477, llmg_0375, llmg_0376 encode the following transporters, AcaP, HisP, FywP, LysP, SerP1, SerP2.
Suppl. Fig. 3 Expression of genes responsible for the purine, pyrimidine and guanine metabolism in *L. lactis* MG1363

The expression signals were transformed with a power 2 transformed to remove analysis differences between this study and the time-course performed by de Jong *et al.* 112. The end of transition phase (tps 18 and 19) is indicated with the vertical dotted line. The *bmpA* gene encodes a nucleoside transporter 124.