ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Cell Cycle Progression in *Escherichia coli* B/r Affects Transcription of Certain Genes: Implications for Synthetic Genome Design

Patricia L. Echtenkamp,[1] David B. Wilson,[2] Michael L. Shuler[1,3]

[1]School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853-5201

[2]Molecular Biology and Genetics, Cornell University, Ithaca, New York

[3]Department of Biomedical Engineering, Cornell University, Ithaca, New York 14853-5201; telephone: 607-255-7577; fax: 607-255-1136; e-mail: mls50@cornell.edu

Received 30 March 2008; revision received 31 July 2008; accepted 12 August 2008

Published online 8 September 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22098

ABSTRACT: We propose that transcript levels for some genes are affected by the bacterial cell division cycle and this may be an important factor to consider when designing synthetic bacterial genomes. To test this hypothesis, transcript levels of 58 genes in *Escherichia coli* B/r A were determined at five times during the cell division cycle. A two-step ANOVA technique was used to analyze data from custom oligonucleotide microarrays containing genes involved in important cellular processes including central metabolism, macromolecular synthesis, and transport and secretion. Consistent with results previously found in *Caulobacter*, approximately 17% of the transcript levels were cell cycle dependent. Cell cycle regulation can be divided into two classes: genes displaying increased transcript concentrations following gene replication and genes displaying an increased transcript concentration prior to replication initiation. Transcripts levels for *hns*, *uspA*, and *zwf* were affected by the cell division cycle, but did not fit well into either class. These results indicate that transcription of a significant fraction of the genome is affected by replication cycle progression. The results also show that both physical gene position and the physiological function of a gene affect when it is transcribed. In addition to the simple association with replication fork progression, other phenomena must be occurring to account for some of our observations. In conclusion, gene position, with regard to the C period, and gene function are important factors to incorporate into design criteria for synthetic bacterial genomes.

Biotechnol. Bioeng. 2009;102: 902–909.

Correspondence to: M.L. Shuler

Contract grant sponsor: National Science Foundation Graduate Research Fellowship Program

Contract grant sponsor: New York State Office of Science, Technology and Academic Research

Additional Supporting Information may be found in the online version of this article.

KEYWORDS: bacterial division cycle; periodic transcription; synchronous cultures; synthetic genome; bacterial chromosome structure

Introduction

Although synthetic biologists have made significant progress toward designing and synthesizing a minimal cell (Filipovska and Rackham, 2008; Forster and Church, 2006; Gibson et al., 2008; Kobayashi et al., 2004), the effects and implications of bacterial genome organization on transcription are still poorly understood. The dynamic nature of bacterial gene transcription is crucial to the survival of bacterial populations. With average mRNA half-lives in *Escherichia coli* ranging from 5.2 to 6.8 min (Bernstein et al., 2002; Selinger et al., 2003), transcription and RNA degradation play important roles in cellular dynamics and adaptability. Bernstein et al. showed that RNA degradation rates for individual transcripts are relatively constant, indicating that dynamic transcript levels reflect changes in transcription rates.

Several factors can affect transcription in bacteria including environmental and intracellular changes. In addition, crucial cellular processes may affect transcription. The cell cycle for all organisms involves mass doubling, chromosome replication and segregation, and cell division. Replication and division cause significant perturbations of cellular physiology and must be carefully coordinated to ensure viable daughter cells and prevent aberrant cell division. In bacteria, DNA replication forks change local DNA structure by increasing positive supercoiling in front of the replication fork and dislodging DNA-associated

proteins, including the transcriptional machinery (French, 1992). We hypothesize that transcript levels for some genes are affected by the bacterial cell cycle and for this reason the cell cycle is an important factor to consider when designing synthetic bacterial genomes. It may be that not only the genes selected for inclusion in a minimal gene set, but also their physical arrangement within the chromosome are important for the optimum physiological response.

Transcript levels throughout the cell cycle are dynamic in all three kingdoms. Cell cycle-related fluctuations were detected for 2%, 5%, and 3–13% of the transcripts in human primary fibroblasts (Cho et al., 2001), HeLa cells (Whitfield et al., 2002) and yeast cells (Cho et al., 1998; Oliva et al., 2005; Peng et al., 2005; Rustici et al., 2004; Spellman et al., 1998), respectively. Despite the range of organisms studied, one characteristic common to these gene sets was that functionally related genes were co-transcribed. More recently, two groups reported that 3% and 10% of transcripts show cell cycle related changes in the archaea *Halobacterium salinarum* and *Sulfolobus acidocaldarius*, respectively (Baumann et al., 2007; Lundgren and Bernander, 2007). Again, functional relationships could be identified for groups of genes that showed similar transcript level profiles.

Less is known about cell cycle related changes to bacterial transcription. Global transcriptional profiles during the cell cycle are reported for only one bacterial species, *Caulobacter crescentus*. Using populations synchronized by centrifugation, Laub et al. (2000) showed that at least 19% of *Caulobacter* genes have cell cycle-related expression profiles, and that at least 25% of those genes are regulated by CtrA, a DNA-binding response regulator. However, the regulator for the remaining 75% of the cell cycle-regulated genes remains a mystery (Ryan and Shapiro, 2003). Stalk formation and compartmentalization accompany cell division in *Caulobacter*, making it difficult to determine whether cell cycle related transcriptional changes in *Caulobacter* are relevant for the greatly simplified genome of a hypothetical synthetic bacterium. The goals of this work were (a) to determine whether transcript levels for genes not related to replication or division processes are affected by the cell cycle for a simple model bacterium, *E. coli*, and (b) to identify design criteria important for developing transcriptionally efficient synthetic genomes.

Here we present evidence that both the location of a gene and its function can affect transcript levels during the cell cycle. Relative RNA abundances were measured for 58 genes in synchronous populations of *E. coli* B/r A at five points during the cell division cycle. The genes are located at different physical positions on the *E. coli* chromosome, and are involved in important processes within the bacterium. We detected changes in transcript levels related to the cell cycle for 10 genes (17%). Two patterns were seen for genes with dynamic transcript levels: increased transcript concentrations following gene replication and increased transcript concentration prior to replication initiation. These results show that both physical gene position and physiological function of a gene affect when a gene is transcribed, indicating that the cell cycle is an important factor to consider when designing synthetic bacterial chromosomes.

Materials and Methods

Bacterial Strain, Growth Conditions and Synchronization

E. coli B/r A (ATCC 12407) were grown in C medium (17.2 mM K_2HPO_4, 11.0 mM KH_2PO_4, 9.5 mM $(NH_4)_2SO_4$, 0.41 mM $MgSO_4$, 0.17 mM NaCl, 3.6 μM Fe_2SO_4, 1.0 μM EDTA) containing 0.1% glucose (Roberts et al., 1955). Populations were synchronized using the membrane elution technique developed by Helmstetter (1969) and Helmstetter et al. (2003). Briefly, cells were attached to a 0.22 μm pore diameter nitrocellulose membrane. C-medium containing 0.1% glucose was continuously pumped through the membrane at 4.5 mL/min. As the cells divided, newborn daughter cells were collected for 2 min (<5% of the doubling time) and incubated at 37°C and 400 rpm. To analyze the synchrony of the *E. coli* populations, the cell concentration was measured using a Coulter Counter Model ZM (Beckman-Coulter, Inc., Fullerton, CA) with a 30 μm aperture diameter. The settings utilized were $I = 500$ μA, polarity = auto, $T_L = 7.0$, $T_U =$ out, attenuation = 1, preset gain = 4 and manometer volume = 50 μL. Synchronous populations were approximately 9 mL with a concentration of approximately 6×10^6 cells/mL.

Microarray Sample Preparation and Hybridization

At 5, 12, 20, 27, 35, and 49 min after cell division and collection, synchronous *E. coli* populations were chilled in a swirling ice bath for 60 s, centrifuged and treated with RNAprotect bacterial reagent (Qiagen, Inc., Valencia, CA). Suspensions were incubated at room temperature for a minimum of 10 min and the cells were pelleted by centrifugation. Total RNA was isolated using RNeasy mini-kits including DNAse I digestion according to the manufacturer's recommendations (Qiagen, Inc.). Samples from asynchronous *E. coli* populations were collected analogously. RNA from several asynchronous populations was pooled to obtain a consistent control.

Aminoallyl cDNA was synthesized from 1 to 3 μg of isolated RNA and cDNA was synthesized and indirectly labeled with Cy3- or Cy5-monoreactive esters as previously described (Hegde et al., 2000). Two-color hybridizations were performed as described earlier (Rhee et al., 2004), with minor modifications. Briefly, slides were prehybridized for 45 min at 42°C in prehybridization solution (50% formamide, 3× SSC, 0.31% SDS, 1 μg/mL salmon sperm DNA). Labeled cDNA samples (20 pmol dye each) were combined with hybridization solution to yield the following final concentrations: 50% formamide, 3× SSC, 0.31% SDS,

0.01 μg/mL salmon sperm DNA. Samples were boiled for 2 min, cooled by pulse centrifugation and hybridized to the microarrays for 20 h at 42°C. The microarrays were scanned using a GenePix 4000B scanner (Axon Instruments, Inc., Union City, CA) with GenePix Pro 4.0 software.

Oligonuceotide Microarray Design and Analysis

Microarrays contained 760 elements (8 replicates of 95 gene probes), that were covalently attached to aminosilane coated, 3 in. × 1 in. microscope slides. The genes included in the arrays were chosen to represent important cellular processes (central metabolism, macromolecular synthesis, nucleotide biosynthesis, transport/secretion and cell structure). In addition, several genes previously shown to have cell cycle-related transcription in either *E. coli* or *C. crescentus* were included in the array. Finally, genes were located throughout the genome (see Fig. 1).

The oligonucleotide probes were 50–55 bases in length. Sequences were designed using Oligoarray 2.1 software (Rouillard et al., 2003) and probes were synthesized by Illumina, Inc. (San Diego, CA). Complete probe sequences and microarray locations are included in Table SII.

Of the genes included in the microarray, cDNA hybridized to 58 probes with intensities at least three times greater than the background intensity. Gene probes that did not hybridize to the cDNA were located throughout the genome (Table SI, Fig. 1). Genes were distributed among classes, excluding nucleotide biosynthesis, but transport/secretion classified genes were over-represented in this subset, indicating these genes displayed lower average transcript levels.

To analyze the microarray data, a two-step ANOVA procedure was used (Lee, 2004). The 532 nm and 635 nm intensities of each gene spot were transformed as follows:

$$y = \sum_{i=8} \log_2 \left(\frac{\text{Intensity}_{\lambda_S} - \text{Intensity}_{\lambda,\text{background}_S}}{\text{Intensity}_{\lambda_A} - \text{Intensity}_{\lambda,\text{background}_A}} \right)$$

where $\text{Intensity}_{\lambda_S}$ refers to the Intensity of the wavelength measuring the synchronous sample and $\text{Intensity}_{\lambda_A}$ refers to the intensity of the wavelength measuring the asynchronous sample. Each biological replicate was averaged over the eight technical replicates. The following model was used to normalize the data

$$y_{\text{gene, array, time, dye, replicate}} = \mu + \alpha_{\text{array}} + \delta_{\text{dye}} + \tau_{\text{time}} + u_{\text{gene, array, time, dye, replicate}}$$

where μ represents the overall average transformed ratio of the hybridization intensity of the synchronous population to the hybridization intensity of the asynchronous population, α represents effects introduced by each array, δ represents effects introduced by the Cy3 or Cy5 dye, τ represents overall time effects, and u represents the residuals. The residuals are the normalized microarray data, and were analyzed for cell cycle related changes specific to each gene as follows:

$$u_{\text{gene, array, dye, time, replicate}} = \gamma_{\text{gene}} + (\gamma\tau)_{\text{gene}\times\text{time}} + \varepsilon_{\text{gene, array, dye, time, replicate}}$$

where γ_{gene} represents the gene's average residual, $(\gamma\tau)_{\text{gene}\times\text{time}}$ represents changes due to time, and ε represents normally distributed random error. Statistical analysis of the $(\gamma\tau)_{\text{gene}\times\text{time}}$ term identified differentially expressed genes during the cell cycle. A false discovery rate of 0.1 was utilized (Benjamini and Hochberg, 1995).

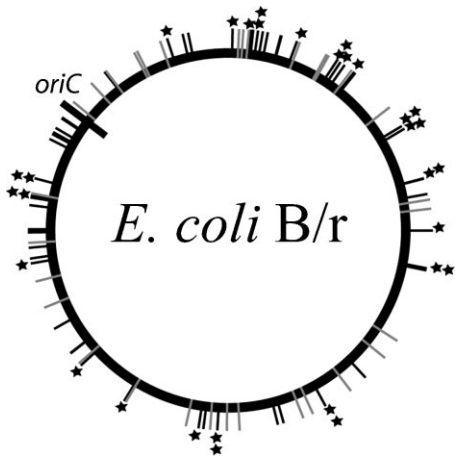

Figure 1. *E. coli* circular genome diagram showing location of genes included in custom microarrays. Black lines represent genes with successful transcript hybridizations; gray lines represent genes with unsuccessful transcript hybridizations. Transcripts for genes marked by stars showed statistically significant fluctuations during the division cycle. Double stars indicate the transcript level was verified at the 49 min time point.

Results

E. coli Synchronous Cultures and Transcript Level Changes

E. coli B/r A cultures grown in a minimal glucose medium were synchronized according to the membrane-elution technique developed by Helmstetter et al. (2003). We adopted this technique to isolate cell cycle-related changes in transcript levels and to avoid changes in the gene expression profiles caused by temperature shifts or nutrient starvation techniques. This procedure relies on separation and collection of newly divided cells from a population of *E. coli* immobilized on a nitrocellulose membrane. Figure 2 shows a representative population profile of a synchronous *E. coli* culture. During exponential phase, the growth rate was 0.95 h^{-1}, which corresponded to a doubling time of 44 min.

Transcript levels for 58 model genes were measured using two-color custom oligonucleotide microarrays. The genes

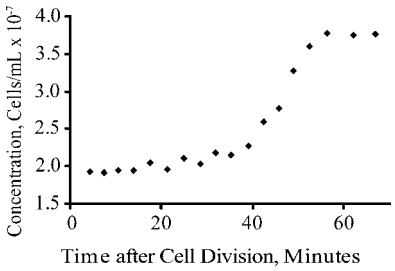

Figure 2. *E. coli* B/r cell concentration profile during a typical synchronous experiment performed in C-medium +0.1% glucose. Average cell division occurs 44 min after collection (0 min).

included in these microarrays represent the following cellular processes: (i) central metabolism, (ii) macromolecular synthesis, (iii) nucleotide biosynthesis, (iv) transport and secretion, and (v) cell structure (Table SI). In addition, the genes were located throughout the genome. Total RNA was isolated from populations of synchronous *E. coli* cells incubated at 37°C and 400 rpm for 5, 12, 20, 27, 35, and 49 min after cell division. To identify changes in transcript level, we used a two-step analysis of variance (ANOVA) technique (Lee, 2004) to analyze data collected from microarray hybridization experiments. Data were normalized to remove slide-specific and dye-specific effects, and subsequently each gene was analyzed for changes in transcript level during one cell cycle.

We detected statistically significant fluctuations in transcript level during the cell cycle for 21 of the 58 genes analyzed (36%) at a false discovery rate of 0.1 (Benjamini and Hochberg, 1995). To further distinguish cell cycle-induced changes, we compared transcript levels for populations incubated for 5 and 49 min ($\tau_D + 5$ min) after collection. Ten of the 21 genes (17% of the total) displayed similar expression levels, indicating the transcript dynamics were related to the cell cycle. Seven genes did not show the same expression levels at 5 and 49 min, and four genes were not verified at 49 min due to low hybridization intensity and high background in the 49 min measurements. The genes displaying changes in transcript level are listed in Table I. Transcript level profiles for the 10 genes showing cell cycle related changes are shown in Figures 3–5.

Changes in Transcript Levels Are Related to the C Period

To verify that our methodology is sufficiently robust to detect cell cycle related changes, we included three genes in our analyses that were shown previously to be cell cycle related. In addition, we included *rpoA*, a gene that has been shown to not fluctuate during the cell cycle (Theisen et al., 1993). As expected, in our studies *rpoA* transcript levels did not fluctuate during the cell cycle (data not shown). Transcript levels for both *ftsZ* and *nrdA* were detected as cell cycle related (Fig. 3). Transcript levels for *dnaA* showed a profile similar to previous studies, but the fluctuations were not statistically significant (P-value = 0.154). The lack of statistical significance may be due to the relatively low level of hybridization, which increases errors between measurements, compared to the other genes in this study.

To visualize the relationship of cell cycle related changes in *E. coli* transcription with the cell replication and division cycles, we aligned our data for *ftsZ*, *dnaA*, and *nrdA* with data collected from previous studies (Fig. 3; Theisen et al., 1993; Zhou and Helmstetter, 1994; Zhou et al., 1997). Although all studies used *E. coli* B/r populations synchronized by the membrane elution technique, the average doubling time and C period for the cells in previous studies were 22 and 45 min, respectively. For our studies the average doubling time was 44 min while the C period was 45 min, reflecting differences in the culture media used. As shown in Figure 3, the transcript level changes align well when considered with respect to the C period. This approach contrasts with attempts to correlate the data based on the fraction of the division cycle, where the correlation was much poorer. These results show that fluctuations in transcript levels are correlated with DNA replication and not cell division. The data corroborate previous results showing that changes in transcript levels are correlated with the C period (Zhou et al., 1997). Our results are also consistent with the finding that transcription of *ftsZ*, a gene involved in cell division, is correlated with the replication cycle, and in particular replication initiation (Garrido et al., 1993; Zhou and Helmstetter, 1994).

Transcript Level Increases Following Gene Replication for Some Genes

The C and D periods for *E. coli* B/r A grown in minimal glucose medium are 45 and 25 min, respectively (Bremer and Chuang, 1981). Using these values and assuming a constant rate of DNA synthesis, we plotted the transcript level throughout the cell cycle with the average times of replication initiation, gene replication and cell division marked (Figs. 4 and 5). Of the 10 transcripts that showed cell cycle related changes, five displayed an increase in transcript level at the first time point following replication of the gene (*secD*, *cmk*, *asd*, *sdhA*, and *gltA*; Fig. 3). *SdhA* and *gltA* are adjacent genes and show very similar transcript level fluctuations during the cell cycle. These genes, and *asd*, encode products involved in central metabolism and therefore demand for their gene products is not expected to change during the cell cycle. *Cmk*, a gene involved in nucleotide biosynthesis, may be expected to fluctuate based on DNA synthesis and will be discussed further below. The remaining gene, *secD*, is involved in transport and therefore we also would not expect it to be correlated with the cell cycle. The increase in transcript level detected shortly after replication of each of these genes indicates that replication fork progression may affect gene transcription.

Table I. Genes displaying statistically significant changes in transcript level during the cell division cycle.

Gene	Protein function	Map position (centisomes)[a]	*P*-value
Cell cycle related			
secD	Sec secretion pathway component	9.2	0.0006
gltA	Citrate synthase (central metabolism)	16.2	0.0011
nrdA	Ribonucleotide reductase	50.5	0.0017
zwf	Glucose-6-phosphate-1-dehydrogenase	41.7	0.0017
sdhA	Succinate dehydrogenase	16.3	0.0025
ftsZ	Essential cell division protein FtsZ	2.3	0.0063
uspA	Universal stress protein	78.4	0.0108
asd	Aspartate semialdehyde dehydrogenase	77.0	0.0132
hns	H-NS, DNA binding protein	27.8	0.0147
cmk	Cytidylate kinase	20.7	0.0313
Dynamic			
ftsL	FtsL (essential cell division protein)	2.0	0.0023
ackA	Acetate kinase A	52.0	0.0088
hflB	Integral membrane peptidase (degrades σ^{32})	71.6	0.0089
dnaK	DnaK chaperone (hsp70 family)	0.3	0.0128
secF	SecF (component of Sec Secretion Complex)	9.2	0.0151
eno	Enolase	62.6	0.0161
groS	GrpE co-chaperone protein	94.2	0.0305
Not verified at 49 min			
tmk	Thymidylate kinase	24.9	0.0208
dnaX	DNA polymerase III	10.6	0.0224
proA	Glutamylphosphate reductase	5.6	0.0318
era	Essential GTP-binding protein	58.2	0.0354

[a]Values calculated by Ecocyc database (Keseler et al., 2005).

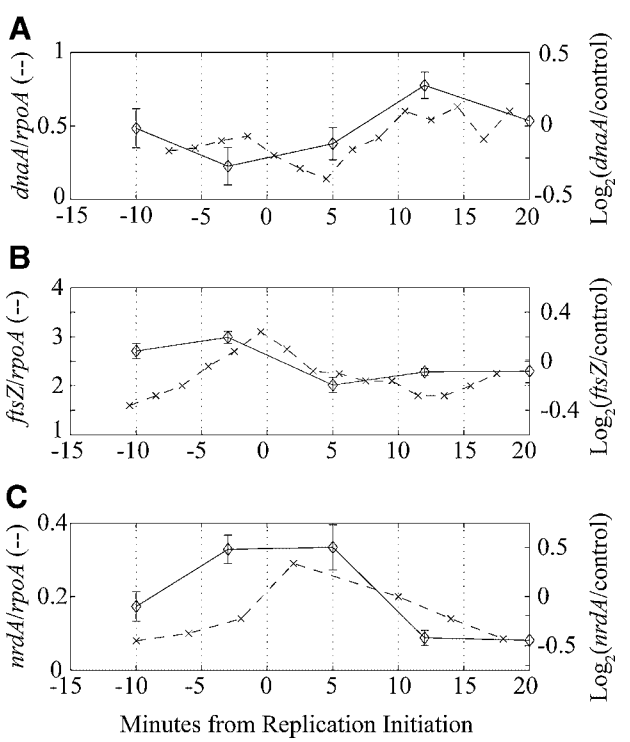

Figure 3. Comparison of transcript levels with previous measurements. Solid lines represent data measured in these studies. Broken lines represent data from Theisen et al. (1993), Zhou and Helmstetter (1994) and Zhou et al. (1997) for *dnaA*, *ftsZ* and *nrdA*, respectively. Data correlate well when plotted by C period and aligned for replication initiation.

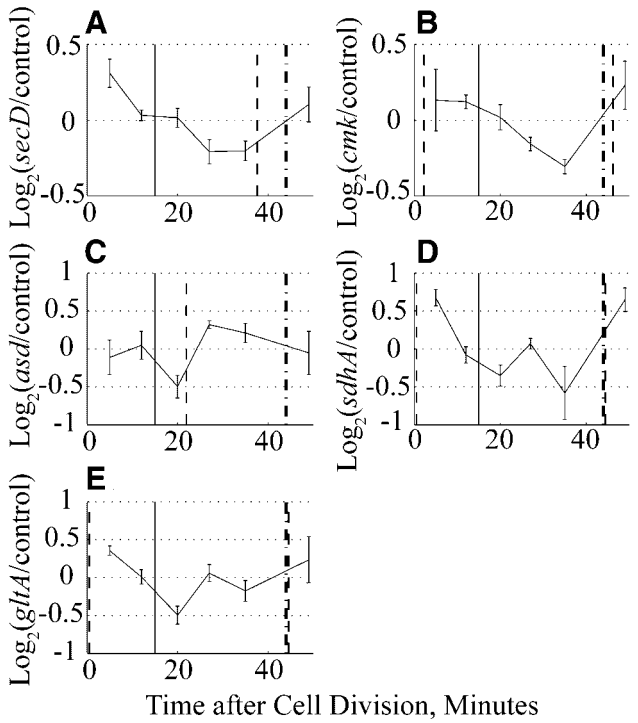

Figure 4. Transcript levels during the cell division cycle for statistically significant genes showing an increase in transcript level following gene replication. The vertical solid line represents replication initiation; the vertical broken line (- - -) represents gene replication and the vertical double broken line (- · -) represents average time of cell division. Plot shows relative transcript levels for: (**A**) *secD*; (**B**) *cmk*; (**C**) *asd*; (**D**) *sdhA*; (**E**) *gltA*.

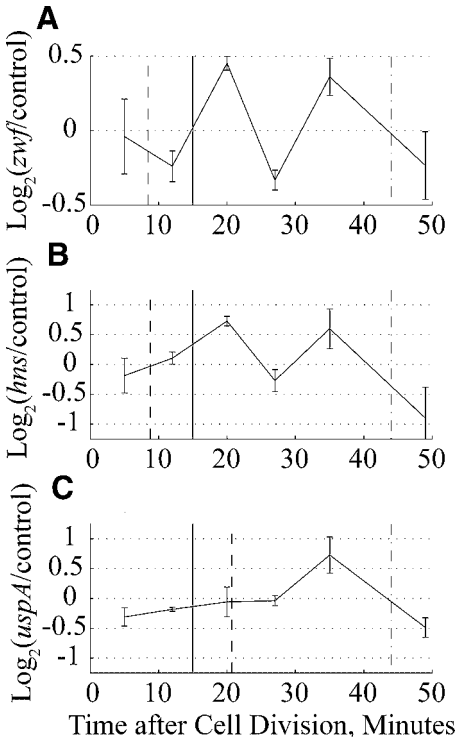

Figure 5. Differentially expressed transcript levels during the cell division cycle. The vertical solid line represents replication initiation; the vertical broken line (- - -) represents gene replication and the vertical double broken line (- - -) represents average time of cell division. Plot shows relative transcript levels for: (**A**) *zwf*; (**B**) *hns*; (**C**) *uspA*.

Additional Transcriptional Mechanisms Exist for *zwf*, *hns*, and *uspA*

Three genes that displayed differential expression during the cell cycle did not show fluctuations correlated to gene replication: *zwf*, *hns*, and *uspA* (Fig. 5). Our results show the transcript level peaks twice during the division cycle for both *zwf* and *hns*, indicating a transcriptional control mechanism beyond replication fork progression is somehow coupled to the division cycle for these genes. Although these genes are not functionally related to replication or division (the *zwf* gene product is involved in central metabolism and the *hns* gene product is a DNA-binding protein and environmental transcription regulator), they both show clear fluctuations in transcript level during the division cycle.

We also detected an increased *uspA* transcript level prior to cell division. Nachin et al. (2005) recently showed UspA is involved in oxidative stress resistance in *E. coli*, and therefore our results may indicate division induces expression of genes designed to protect cells from DNA-damaging agents.

Transcript Level for Nucleotide Biosynthesis Genes Increase Near Replication Initiation

Three of 10 genes measured that were involved in nucleotide biosynthesis displayed statistically significant changes in transcript levels during the cell cycle (*nrdA*, *cmk*, *tmk*). In addition to analyzing each gene separately for transcript level fluctuations, we performed an ANOVA test on each class of genes. As a group, nucleotide biosynthesis genes (*dut*, *upp*, *apt*, *tmk*, *gmk*, *nrdA*, *hpt*, *cmk*, *adk*, *thyA*) displayed changes in transcript levels during the cell cycle (P-value <0.0001). The maximum average transcript level was detected 3 min prior to replication initiation, while the minimum transcript level occurred 20 min after replication initiation, or approximately mid-C period (Fig. 6). To ensure the gene group was not dominated by the three genes showing cell cycle related transcript fluctuations (*nrdA*, *cmk*, *tmk*), the ANOVA test was also performed for only nucleotide biosynthesis genes not showing statistically significant transcript fluctuations on an individual level (*dut*, *upp*, *apt*, *gmk*, *hpt*, *adk*, *thyA*). This analysis yielded equivalent results (P-value $= 0.0096$). Although our results were not sensitive enough to detect cell cycle related changes for all genes involved in nucleotide biosynthesis individually, the gene class shows strong correlation with the C period, indicating that transcription of this set of genes is regulated during the cell cycle.

Discussion

In recent years, there has been a growing effort to construct a synthetic bacterium. A major step in this process is defining and assembling a minimal genome de novo. To this end, researchers have attempted to identify a minimal bacterial gene set necessary for life (reviewed by Koonin, 2003; Smalley et al., 2003). However, our understanding of transcriptional organization in even the simplest organisms remains incomplete. Because DNA replication and division are major processes that must be undertaken by any synthetic organism, it is important to understand their

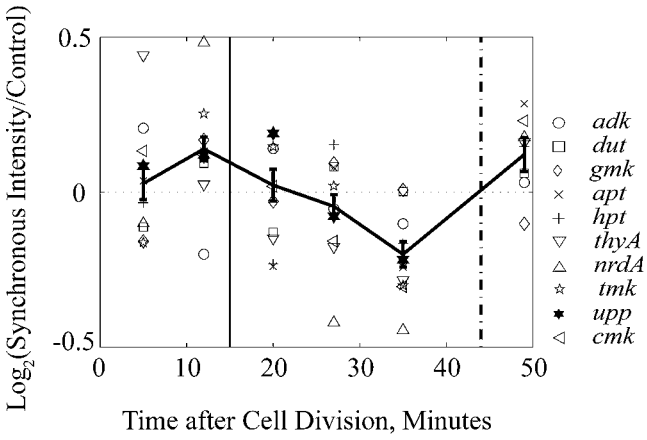

Figure 6. Compiled transcript levels for nucleotide biosynthesis genes included in microarray analysis. The vertical solid line represents replication initiation and the vertical double broken line (- - -) represents average time of cell division. Symbols represent genes show in legend on right.

effects on transcript profiles. In this study, we identify two transcriptional phenomena related to cell cycle progression: transcription of some genes increases transiently after the gene is replicated; and transcription of genes involved in nucleotide biosynthesis are activated prior to replication initiation.

We employed an ANOVA procedure to analyze transcript measurements and did not limit our analysis to detecting only transcripts with cyclic profiles. To eliminate changes due to factors outside of cell cycle progression, we required that data collected for populations of cells at 49 min was statistically indistinguishable from the 5 min time point. Ten of the 21 genes displaying dynamic transcript levels fit this criterion. Of the seven genes that did not fit this criterion, our results may show real cell cycle effects for some while others may be dynamic due to other cellular processes. For example, fluctuations for *ftsL* and *hflB* may be related to the cell cycle based on the division cycle related function of the gene products (e.g., FtsL and FtsH, respectively). In addition, we would expect to see changes in *secF* transcript levels similar to the *secD* transcript levels because these genes are expressed in the same operon. Indeed the *secD* and *secF* profiles are similar (Table SI). Both *dnaK* and *groS* also display dynamic transcript levels but do not recover to the 5 min level at the 49 min point. These genes are both stress-induced and could be up-regulated based on the minor changes in the cellular environment (i.e., shaking during incubation).

Several of the cell cycle related transcripts displayed a peak shortly after the replication fork progressed through the coding gene. Previously *E. coli* transcript levels for several genes (most not studied here) have been shown to fluctuate with the cell replication cycle (Garrido et al., 1993; Theisen et al., 1993; Zhou and Helmstetter, 1994; Zhou et al., 1997). *E. coli* cultures that were synchronous with respect to either cell division or replication initiation showed that transcript levels for *mnmG* (*gidA*), *mioC*, *dnaA*, *dam*, *ftsZ*, *seqA*, *nrdA*, and *mukB* all decreased transiently near gene replication. Due to the longer sampling intervals we employed, we would not expect to see this inhibition in our studies. However, several of the transcripts in previous studies also displayed a peak in transcript level following the decrease, which is consistent with our findings. In addition, previous studies found transcription of *rpoA*, *minE*, and *tus* were unaffected by cell cycle progression, and transcription of *argP* (*iciA*) increased near the time of its replication. These results indicate that not all genes are affected by the cell cycle, which is corroborated by the data presented here. In previous studies, genes were analyzed based on their functional or structural relationship with the cell division or replication cycles. Our studies provide evidence that transcript levels for a significant fraction of genes in *E. coli* are affected by the cell cycle, whether or not the function or position of the gene product is related to cell cycle progression.

In addition to showing induction following replication of some genes, we show that fluctuations in transcript level depend on the length of the C period as opposed to the division cycle. The replication cycle in *E. coli* can be divided into I, C, and D periods. The I period is simply defined as the time required between replication initiation events. The C period is defined as the time it takes to complete one round of replication and the D period is the time between replication termination and cell division. At high growth rates, *E. coli* replication cycles overlap as cell division is not a prerequisite for initiating the next round of replication, therefore the doubling time (τ_D) is often shorter than the sum of the C and D periods. Our results show that cell cycle related transcription is correlated to the C period, not the division cycle. Based on this relationship, we would expect changes in transcript level to be somewhat lower in *E. coli* populations synchronized by the division cycle rather than the replication cycle. Indeed, the fluctuations in transcript level that we measured, although statistically significant, were in general less than twofold. Based on the observed correlation with the C period, transcript level changes may be more pronounced in populations synchronized with respect to the replication cycle.

Finally, although we were able to detect two distinct phenomena affecting transcription, these do not fully account for all of our results. Specifically, the *zwf*, *hns*, and *uspA* transcript profiles during the cell cycle are not predicted by either gene location or gene function. In addition, *hns* and *zwf* transcripts do not display a purely cyclic profile, but peak multiple times during the cell cycle. Therefore, other mechanisms must occur during the cell cycle to control transcript levels for important genes. Further study of these genes may reveal important information about transcriptional regulation.

The overall goal of this work was to identify important design criteria for construction of a bacterial genome de novo. Our results show the time at which a gene is replicated is an important factor to consider when organizing a synthetic genome. Because the transcriptional variations introduced by the cell cycle are related to replication, key design criteria that must be considered when designing synthetic genomes include replication initiation timing, the rate of DNA polymerization and the distance of the gene from the origin of replication. Our results show that transcription of 15–20% of the genome fluctuates with the cell cycle, and about half of the fluctuating transcripts reach a maximum following gene replication. Therefore genes should be arranged within the genome to allow for transcription at their optimum time within the cell cycle. The rate of DNA polymerization must also be considered, and this may be especially important when designing a minimal genome. Because DNA replication may serve as a key process to make regions of the genome accessible for transcription, it is possible that replication must persist for a specified amount of time to optimize the ratio between the fraction of the genome in a transcriptionally accessible state and the concentration of the transcriptional machinery, and to enable efficient transcription of accessible genes. Consistent with this hypothesis, the chief DNA polymerase in *Mycoplasma capricolum*, a bacterium with a greatly

reduced genome size, is approximately 10 times slower than *E. coli* DNA polymerase III (Seto and Miyata, 1998).

Another interesting observation from this study is that nucleotide biosynthesis genes are up-regulated concurrently with DNA replication initiation. Although each of these genes is replication related, they are located throughout the genome and are replicated in different segments of the cell cycle. Therefore a hierarchy of gene regulation mechanisms may exist in bacterial systems. Further understanding of all transcription regulation mechanisms will aid in designing efficient and robust synthetic genomes.

The authors thank Kelvin H. Lee and Jordan Atlas for helpful discussions. This work was supported in part by the National Science Foundation Graduate Research Fellowship Program, and funds from the New York State Office of Science, Technology and Academic Research to Michael L. Shuler.

References

Baumann A, Lange C, Soppa J. 2007. Transcriptome changes and cAMP oscillations in an archaeal cell cycle. BMC Cell Biol 8:21.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Roy Stat Soc Ser B (Stat Method) 57(1):289–300.

Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci USA 99(15):9697–9702.

Bremer H, Chuang L. 1981. The cell cycle in *Escherichia col*i B/r. J Theor Biol 88:47–80.

Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Wolfsberg TG, Gabriellian AE, Landsman D, Lockhart DJ, Davis RW. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2(1):65–73.

Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, Sapinoso L, Hampton G, Elledge SJ, Davis RW, Lockhart DJ. 2001. Transcriptional regulation and function during the human cell cycle. Nat Genet 27(1): 48–54.

Filipovska A, Rackham O. 2008. Building a parallel metabolism within the cell. ACS Chem Biol 3(1):51–63.

Forster AC, Church GM. 2006. Towards synthesis of a minimal cell. Mol Syst Biol:Article 45.

French S. 1992. Consequences of replication fork movement through transcription units in vivo. Science 258:1362–1365.

Garrido T, Sanchez M, Palacios P, Aldea M, Vicente M. 1993. Transcription of *ftsZ* oscillates during the cell cycle of *Escherichia coli*. EMBO J 12(10):3957–3965.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Naskov VN, Glass JI, Venter JC, Hutchison III CA, Smith HO. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 319:1215–1220.

Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Earle-Hughes J, Snesrud E, Lee N, Quackenbush J. 2000. A concise guide to cDNA microarray analysis—II. BioTechniques 29(3):548–562.

Helmstetter CE. 1969. Methods for studying the microbial division cycle. In: Norris JR, Ribbons DW, editors. Methods in microbiology. London, New York: Academic Press. p 327–363.

Helmstetter CE, Thornton M, Romero A, Eward KL. 2003. Synchrony in human, mouse and bacterial cell cycles. Cell Cycle 2:42–45.

Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, Peralta-Gil M, Karp PD. 2005. EcoCyc: A comprehensive database resource for Escherichia coli. Nucleic Acids Res 33:D334–D337.

Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ. 2004. Programmable cells: Interfacing natural and engineered gene networks. Proc Natl Acad Sci USA 101(22):8414–8419.

Koonin EV. 2003. Comparative genomics, minimal gene sets and the last universal common ancestor. Nat Rev Microbiol 1(2):127–136.

Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L. 2000. Global analysis of the genetic network controlling a bacterial cell cycle. Science 290:2144–2148.

Lee M-LT. 2004. Analysis of microarray gene expression data. Boston: Kluwer Academic Publishers.

Lundgren M, Bernander R. 2007. Genome-wide transcription map of an archaeal cell cycle. Proc Natl Acad Sci USA 104(8):2939–2944.

Nachin L, Nannmark U, Nyström T. 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J Bacteriol 187(18):6265–6272.

Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J. 2005. The cell cycle-regulated genes of Schizosaccharomyces pombe. PLoS Biol 3(7):e225.

Peng X, Karuturi RKM, Miller LD, Lin K, Jia YH, Kondu P, Wang L, Wong LS, Liu ET, Balasubramanian MK, et al. 2005. Identification of cell cycle-regulated genes in fission yeast. Mol Biol Cell 16(3):1026–1042.

Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J. 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl Environ Microbiol 70(7):4303–4317.

Roberts RB, Abelson PH, Cowie DB, Boulton ET, Britten RJ. 1955. Studies of biosynthesis in *Escherichia coli*. Washington, DC: Carnegie Institute of Washington.

Rouillard J-M, Zuker M, Gulari E. 2003. OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31(12):3057–3062.

Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bahler J. 2004. Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36(8):809–817.

Ryan KR, Shapiro L. 2003. Temporal and spatial regulation in prokaryotic cell cycle progression and development. Annu Rev Biochem 72:367–394.

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res 13:216–223.

Seto S, Miyata M. 1998. Cell reproduction and morphological changes in *Mycoplasma capricolum*. J Bacteriol 180(2):256–264.

Smalley DJ, Whiteley M, Conway T. 2003. In search of the minimal *Escherichia coli* genome. Trends Microbiol 11:6–8.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. 1998. Comprehensive identification of cell cycle-related genes of the yeast *Saccharomyces cervisiae* by microarray hybridization. Mol Biol Cell 9(12):3273–3297.

Theisen PW, Grimwade JE, Leonard AC, Bogan JA, Helmstetter CE. 1993. Correlation of gene transcription with the time of initiation of chromosome replication in *Escherichia coli*. Mol Microbiol 10(3):575–584.

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, et al. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13(6):1977–2000.

Zhou P, Helmstetter CE. 1994. Relationship between *ftsZ* gene expression and chromosome replication in *Escherichia coli*. J Bacteriol 176(19): 6100–6106.

Zhou P, Bogan JA, Welch K, Pickett SP, Wang H-J, Zaritsky A, Helmstetter CE. 1997. Gene transcription and chromosome replication in *Escherichia coli*. J Bacteriol 179(1):163–169.