Global Transcriptional Programs Reveal a Carbon Source Foraging Strategy by Escherichia coli

Received for publication, December 14, 2004, and in revised form, January 26, 2005
Published, JBC Papers in Press, February 10, 2005, DOI 10.1074/jbc.M414050200

Mingzhu Liu†, Tim Dufee‡, Julio E. Cabrera§, Kai Zhao¶, Ding J. Jin§, and Frederick R. Blattner¶

From the ‡Department of Genetics and §McArthur Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706 and the ¶Transcription Control Section, Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, NCI at Frederick, National Institutes of Health, Frederick, Maryland 21702

By exploring global gene expression of Escherichia coli growing on six different carbon sources, we discovered a striking genome transcription pattern: as carbon substrate quality declines, cells systematically increase the number of genes expressed. Gene induction occurs in a hierarchical manner and includes many factors for uptake and metabolism of better but currently unavailable carbon sources. Concomitantly, cells also increase their motility. This, as the growth potential of the environment decreases, cells appear to devote progressively more energy on the mere possibility of improving conditions. This adaptation is not what would be predicated by classic regulatory models alone. We also observe an inverse correlation between gene activation and rRNA synthesis suggesting that reappportioning RNA polymerase (RNAP) contributes to the expanded genome activation. Significant differences in RNAP distribution in vivo, monitored using an RNAP-green fluorescent protein fusion, from energy-rich and energy-poor carbon source cultures support this hypothesis. Together, these findings represent the integration of both substrate-specific and global regulatory systems, and may be a bacterial approximation to metazoan risk-prone foraging behavior.

Jacob and Monod originally studied Escherichia coli gene regulation using carbon catabolism as the experimental system (specifically lactate). They elucidated that genes needed to metabolize lactose are specifically induced by that substrate (1). Over the ensuing 40 years this model has been refined and extended to many substrates and it is now generally accepted that virtually all carbohydrate catabolic genes can be regulated by substrate-specific induction (2). This mechanism is attractive because it promotes efficient use of cellular resources, and energy need not be wasted producing enzymes and transporters for substrates unless they are available.

In addition to these highly specific mechanisms, cells also have several levels of global regulation. A prime example is carbon catabolite repression, a multifactorial system that blocks expression of alternative carbon utilization pathways when glucose is present (3). Relief from carbon catabolite repression, among other effects, activates the cyclic AMP receptor protein (CRP),† a global transcription factor that positively regulates most carbon catabolic pathways including those for glucose (4). Such global mechanisms have the advantage of responding to many different conditions and can potentiate a spectrum of transcriptional outcomes.

While much is known about how individual substrate-specific and global systems regulate a limited set of operons under a defined condition, much less is known regarding how the two coordinate genome-wide transcription under a range of conditions. In this paper, we report the findings from a global gene expression study of E. coli growing on a series of six carbon sources. A transcription pattern emerges from these profiles that expands hierarchically as the growth rate declines. Inspection of the up-regulated genes suggests that, instead of maximizing energy conservation through strict adherence to substrate-specific induction, cells growing on poor substrates devote progressively more of their limited reserves to expand their genome transcription needed to broaden a search for alternative energy sources.

EXPERIMENTAL PROCEDURES

Bacterial Growth—MG1655 cells were grown overnight in MOPS minimal medium (5) supplemented with 0.1% glucose. Overnight cultures were then diluted 1:50 (or 1:10² for the deeper adapted experiments) into the same base medium with either glucose, glycerol, succinate, L-alanine, acetate, or L-proline as the carbon source and grown aerobically at 37 °C in shake flasks. The carbon compound concentration in each case was adjusted so that the number of carbon atoms was equivalent to that of 0.1% glucose.

Sample Preparation for Microarray Analysis—Detailed protocols for RNA preparation and labeling can be found at www.genome.wisc.edu/ functional/protocols.htm. Briefly, at an A₂₆₀ of 0.2 (~6 generations except for deeper adapted cultures, ~30), 15-ml samples were mixed with 30 ml of RNase A for 10 min at 37 °C. Total RNA was then isolated using MasterPure kits (Epicenter Technologies).

cDNA was synthesized and labeled using a protocol similar to that described by Rosenow et al. (6). Briefly, 10 μg of total RNA was reverse transcribed with 1,200 units of Superscript II (Invitrogen) using 500 ng of random hexamers at 42 °C for 90 min. The reaction buffer was used according to the manufacturer’s recommendations. Remaining RNA was removed with 2 units of RNase H and 1 μg of RNase A for 10 min at 37 °C. Synthesized cDNA was then purified with QiaQuick (Qiagen) and fragmented to 50–100 bp with 0.2 unit of DNase I (Epicenter) for 10 min at 37 °C. The fragmented cDNA was 3’-end-labeled with 25 μM

† The abbreviations used are: CRP, cyclic AMP receptor protein; RT, reverse transcription; MOPS, 4-morpholinepropanesulfonic acid; DAPI, 4’,6-diamidino-2-phenylindole; AI-2, autoinducer-2; RNAP, RNA polymerase; GFP, green fluorescent protein.
biotin-N^4-ddATP (Applied Biosystems, Foster City, CA) using 50 units of terminal transferase (New England Biolabs, Beverly, MA) at 37 °C for 2 h. The labeled cDNA was hybridized to Affymetrix GeneChip® E. coli antisense genome arrays as recommended by the manufacturer (www.affymetrix.com). Following a 16-h hybridization at 45 °C, the array was washed and stained with streptavidin-phycocerythrin (Molecular Probes) using an antibody intermediate to enhance the signal. The arrays were read at 570 nm with a resolution of 3 μm using a Gene-Array® confocal laser scanner (Affymetrix). Washing and scanning were automated by a GeneChip® Fluidics Station controlled by Affymetrix® Microarray Suite 5.0 software (Affymetrix). For each carbon source, two biological replicates were done except glucose which was replicated five times.

Image Processing and Data Analysis—Image analysis was carried out by Affymetrix® Microarray Suite 5.0 software. This calculates a detection call, detection p value and signal (background-subtracted and adjusted for noise) for each gene. Those values are then imported into a relational database, converted to log2 values, and averaged for each gene across replicates. The correlation coefficient of log2 values between any two replicates was greater than 0.95. Genes were considered up-regulated relative to glucose if they increased at least 3-fold in signal intensity and the signal intensity in experiment had a log2 value of at least 8.5. Conversely, genes with at least a 3-fold reduction in signal intensity were down-regulated if at least a 3-fold decrease was deemed down-regulated. The higher log2 intensity values were used to limit the analysis to those genes for which we have a high degree of confidence in their level of expression. Note, the use of “up-regulation” and “down-regulation” refers only to the change in measured RNA abundance and has no regulatory implication.

Real-time Quantitative RT-PCR—RNA leading sequence RNAs were quantified using real-time PCR in an ABI Prism 7700 DNA Analyzer (Applied Biosystems), and the QuantiTect SYBR green RT-PCR kit (Qiagen) was used according to the manufacturer’s protocol. A 115-base pair segment was amplified using forward primer, 5′-TGCAGCGGAA-CAACGGCAAACACG-3′, and reverse primer, 5′-TCAGTAATAACCGCT-TCCGCTACA-3′. This fragment assesses transcripts originating from the P1 promoter of three rrn operons (A–C). PCR cycling conditions were 10 cycles of denaturation at 94 °C for 15 s, annealing at 94 °C for 1 min, and extension at 60 °C for 1 min. Relative gene expression data analysis was carried out with the standard curve method (7). The frr gene was chosen as the internal control because of its consistent expression levels across conditions (see www.genome.wisc.edu).

Measurement of α and β Protein Levels—Cells were harvested at the appropriate times and lysed in 1× SDS lysis buffer (0% (w/v) SDS, 0.06 M Tris-HCl (pH 6.8), 5% (v/v) β-mercaptoethanol, 5% (v/v) glyc- erol). Extracts were separated by SDS-polyacrylamide gel electrophore- sion and then transferred electrophoretically to a nitrocellulose mem- brane. Blots were probed with specific mouse monoclonal antibodies NT73, 1B51, and 1RE53 (8), which detect the RNA polymerase (RNAP) β subunit, α2, and β, respectively. The blot was rinsed three times with TBST (0.1% phosphoric acid, 0.1% Tris-HCl (pH 7.4), 150 mM NaCl, and 0.1% Tween 20) and then incubated with goat anti-mouse IgG horseradish peroxi- dase-labeled secondary antibody for 1 h at room temperature. The blots were again washed three times with TBST buffer and proteins visual- ized with SuperSignal® West Femto Maximum Sensitivity Substrate (Pierce).

CRP-binding Site Analysis—Seventy known CRP-binding sequences (9, 10) were used as input for scanACE (11). The input sequence is 22 base pairs in length and two segments (position 5–9 and 16–20) were indicated as active columns representing the strong consensus TGTGA-N^4-pTCACA. 500 base pair upstream region of each operon in the three sets (see “Results”) were scanned using the above input motif.

Microscopy—D25399 cells (12) were grown in either glucose or pro- line supplemented MOPS minimal media. At an A590 of 0.2, 5-ml ali- quots of culture were removed from the culture flasks and formaldehyde added to a final concentration of 3.7%. Cells were fixed for 1 h at room temperature, centrifuged, and resuspended in 1 ml of 1× PBS. Before mounting the cell mixture on the slides, 15 μl of a 10 μg/ml solution of DAPI was added. The final mixture of fixed cells was mounted on slides using 1% low melting point agarose. Microscopy was performed in a Zeiss Axiophot II microscope equipped with a Plan-Apo 100× objective, epifluorescence filters, and a 2.5 optivar. Images were captured with a CCD camera (Micromax) working at 2×2 binning. The images were processed with Adobe Photoshop.

Contrast Analysis of Nucleoids—Normalized contrast measurements were performed as described previously (12) to 100 cells per glucose or proline culture. Briefly, in each region of interest at each nucleoid, we measured the intensity of each pixel and its 8 neighbor pixels. The differences in intensities were used to feed a gray level co-occurrence matrix as described by Haralick (13). This gray level co-occurrence matrix is a representation of the gray level transitions within the region of interest, and the contrast textural feature can be calculated from it. As described previously, the normalized contrast was obtained by di- viding the contrast textural feature by the area and by the mean’s square of the region of interest. This ensures a normalization of the contrast parameter by area and gray level intensity.

Results

Growth on Alternative Carbon Sources Elicits a Scalable Transcriptional Program—To investigate the effect of different carbon sources on E. coli global gene expression, we grew MG1655 cells aerobically in MOPS minimal medium (5) with either glucose, glycerol, succinate, t-alanine, acetate, or t-proline as the carbon supply. These carbon sources are of differing quality as defined by the resulting log phase growth rates which range from 0.97 generation h−1 in glucose to 0.13 generation h−1 in proline (Fig. 1A). Samples were taken from each culture at mid-log phase (six generations), and total RNA was purified, labeled, and hybridized to Affymetrix E. coli Anti- sense GeneChips. Expression profiles of the five alternative carbon sources were compared with the fastest growing culture, glucose, to identify genes whose steady state abundance depends on the quality of the carbon source. To ensure we were monitoring an adapted state rather than a transitional response, acetate and glycerol cultures were also diluted 1:10^7 into the test medium such that ~30 generations were needed to reach the same density. Expression profiles from these cultures were identical to the initial samples indicating steady state patterns are established during the course of the experiments.

The comparative analysis revealed two unexpected results. First, the number of up-regulated genes in each alternative carbon source far exceeds the number of down-regulated trans- scripts, and this effect is inversely correlated with growth rate (Fig. 1A). Second, the up-regulated genes are activated in a hierarchical manner with respect to growth rate (Fig. 1B, i.e. the genes expressed at higher growth rates are largely a subset of those expressed in slower growing cells. This sug- gests a common response system is employed across carbon sources.

Specifically, glycerol and succinate cultures have similar growth rates and share 40 up-regulated genes. The set of up-regulated genes in alanine is larger (188 genes) and includes all 40 genes in the glycerol/succinate common set. The two con- pounds with the slowest growth rates, acetate and proline, have similar profiles despite their different generation times, and both profiles contain 144 of the 188 genes in the alanine data set including the 40 glycerol/succinate core set. To simplify the following discussion, we consider the common 201 up-regulated genes as three nested sets: set 1 (glycerol ~ succinate) ⊆ set 2 (alanine) ⊆ set 3 (acetate ~ proline), where set 1 constitutes the core group that is up-regulated in all five alternative carbon experiments (40 genes); set 2 contains those up-regulated in alanine, acetate, and proline only (85 genes); and set 3 includes genes that are only up-regulated in the two slowest carbon sources, acetate and proline (76 genes). These 201 genes are contained in 154 operons. Transcript abundances across the sets generally increase linearly with declining growth rate so the difference between the three sets is due primarily to genes having different induc- tion slopes, not because they are turned on in discrete steps (Fig. 1C).

Genes for Metabolizing Unavailable Compounds Are Systematically Induced—Although genes necessary for specifically utilizing the available carbon compound are up-regulated, they constitute only a small percentage of the total genes activated.
in each profile. Instead, a significant number of transport and catabolic genes for carbon sources not present in the medium are progressively induced in each set (Fig. 2). There is a general, but not strict, trend toward activation of transporters for compounds yielding growth rates comparable with or better than the source present. While activation of pathways for unavailable substrates has been seen previously in single carbon source expression studies (14–16), the extent and systematic nature of the phenomenon is striking. It appears that as the substrate quality decreases, cells increase their ability to rapidly switch to preferable compounds if they become available.

**Cell Motility Increases as Carbon Source Quality Decreases**—Induction of the ego-lsrCDBFG-tam operon was also observed in the three slowest growing cultures. In *Salmonella typhi-murium*, this operon is induced by the extracellular autoinducer-2 (AI-2) and is required for AI-2 uptake and processing (17). AI-2 has been proposed to be a universal signal for interspecies communication in bacteria (18) and has different species-specific signaling roles including enhancing motility in *E. coli* (19). This observation led us to examine cell motility in each culture using video microscopy. A gradient of activity correlated with carbon quality was observed ranging from basically sessile cells in glucose to highly motile cells in acetate, although motility declined in proline (movie files at wwwgenome.wisc.edu/functional.htm). This observation, together with the systematic up-regulation of transport genes, further suggests that cells actively search for better conditions as the quality of the available substrate(s) decreases. The possible involvement of AI-2 further suggests this search may be an aspect of group behavior as well.

**Accumulation of the Stress Response Sigma Factor, $\sigma^B$, and Its Regulon**—Other genes in set 2 include carbon utilization regulators and a subset of the $\sigma^B$-controlled regulon. The $\sigma^B$-encoding gene, rpoS, is slightly up-regulated as is a second alternative sigma factor gene, rpoE (Fig. 3A). More impor-
tantly, production of both proteins substantially increases in alanine, acetate, and proline (Fig. 3B). Accumulation of these two factors is critical for priming a variety of stress defense systems for cellular adaptation to changes in external environments (20–22). Set 3 enlarges the number and abundance of activated σ²-dependent genes. Approximately 30% of the remaining up-regulated genes in the three sets is of unknown function.

Examination of CRP Involvement by in Silico Promoter Analysis—CRP is known to positively control most carbon catabolic pathways when the carbon source changes from glucose. To assess how pervasive CRP control may be over the shared 154 up-regulated operons, we built a binding consensus from 70 experimentally verified CRP-binding sites (9, 10) and scanned each promoter region using scanACE (11). We found that 24 operons contain consensus CRP-binding sites, 22 of which are known members of the regulon and two of which are new (paaXY and gatYZABCDR_2). Eleven of these 24 are in set 1 and 13 in set 2. That is, 15.6% of the operons are under known or putative CRP control.

Global Gene Activation Is Inversely Correlated with rRNAs Leader Sequence Abundance—The apparent coupling of growth rate and gene activation in these experiments prompted us to measure the RNA synthesis rate in each culture, since this parameter is known to be correlated with growth rate (23, 24). rRNAs leader sequence abundance has been shown to accurately reflect the rRNAs synthesis rate (25) due to the fact that they are processed and degraded very rapidly in the cell. We used real-time RT-PCR to measure the rRNAs leader sequence abundance in cells grown in alternative carbon sources relative to that in glucose-grown cells. Results confirm that rRNAs synthesis decreases proportionately with reduced growth rate in these cultures, although a limit is reached in acetate and proline cultures (Fig. 4). This result indicates that the increased expression of the nested set is more correlated with decreased rRNA gene activity than growth rate.

RNAP Distribution Is Significantly More Homogenous in Slow Growing Cells—We have previously shown, using a functional rpoC-gfp allele (rpoC encodes the β’ subunit of RNAP), that RNAP distribution is sensitive to environmental factors that affect growth rate and the synthesis of stable RNA (12). In rapidly dividing cells, RNAP is largely concentrated in transcription foci that likely represent sites of stable RNA transcription and rapidly dissipate under conditions that suppress growth and thereby rRNA synthesis, e.g. amino acid starvation, leading to more diffuse fluorescence throughout the nucleoid (12). To determine whether RNAP distribution is affected by different carbon sources, DJ2599 (MG1655 ΔrpoC::gfp) cells were grown in both glucose and proline supplemented media to assess changes at the two growth rate extremes studied here. Cell morphology, RNAP-GFP distribution, and nucleoid compaction were assessed as described previously (12).

Glucose grown cells were significantly larger and more rod-
Carbon Source Affects on Global Transcription

The distribution of RNAP in glucose grown cells was relatively heterogeneous compared with that in proline as evidenced by the increased number of distinct foci in the former (compare Fig. 5, E and C). In addition, visual inspection of the relative cellular space occupied by the nucleoid suggests the chromosome may be more decondensed in proline-grown cells (compare Fig. 5, C and D and G and H).

To measure the degree of homogeneity of the RNAP distribution in nucleoids from different cells, we quantified the intensity variations in the RNAP-GFP fluorescence signal within individual nucleoids using texture analysis (13) as described by Cabrera and Jin (12). This analysis has been used extensively to analyze nuclear structure in eukaryotic cells (26–28). Larger values of normalized contrast, a parameter that reflects variations in gray intensities within an image, indicate a more heterogeneous distribution of RNAP within the nucleoids. The normalized contrast parameter was measured from 200 nucleoids as described under “Experimental Procedures.” Statistics test shows that glucose-grown cells have significant higher values than proline-grown cells, indicating the distribution of RNAP from glucose-grown cells is more heterogeneous in nucleoids because of the presence of transcription foci, whereas the distribution of RNAP proline cells is relatively homogeneous (Fig. 5I). These results demonstrate a clear difference in RNAP distribution between the two cultures and suggest that the expanded transcription pattern observed in proline-grown cells may result from RNAP being diverted from the stable RNA promoters to those of the up-regulated genes.

DISCUSSION

E. coli has multiple substrate-specific regulatory mechanisms that allow cells to fully activate pathways necessary for utilizing a given carbon compound once it becomes available. However, the global transcription profiles of E. coli grown on a range of carbon sources presented here show that, as the quality of the available carbon source decreases, instead of predominantly adapting to metabolize the available substrate, cells also initiate a much broader transcription program including up-regulating many genes for utilizing unavailable compounds. This program expands in a manner inversely correlated with the rRNA synthesis rate and primes the ability of cells to find and utilize better carbon sources as well as respond quickly to environmental stresses.

The majority of up-regulated genes are involved in different aspects of metabolism. Among these are multiple pathways for uptake and utilization of preferable carbon compounds. Transporter activation in the absence of substrate has been seen in other experiments where the growth rate was reduced either by glucose limitation in a chemostat (29) or by growth in a single alternative carbon source (14–16). The scope and systematic nature of induction observed here, however, is unanticipated and emphasizes that the profiles are not an ad hoc response to individual substrates but instead a common gradual response to a series of compounds. That expression levels of individual genes increase with declining growth rate is also consistent with earlier studies that focused on a limited set of operons (30–33) and demonstrates that the regulatory mechanisms impart both qualitative and quantitative effects.

Coordinating transcription factor activity with growth rate is one means of establishing a gradated response. CRP, for example, requires binding of cAMP for dimerization and activation (34), and cAMP levels vary inversely with the growth rate in...
the absence of glucose (3, 35, 36). This results in the progressive up-regulation of the CRP regulon in sets 1 and 2. Initial promoter analysis indicates that other global transcription factors, such as ArcA and Cra, whose activity can also be modulated across growth conditions, are involved in establishing the pattern (data not shown).

Increased expression of $\sigma^S$ and the resulting partial activation of its regulon in the three slowest growing alternative carbon sources also contribute to the pattern. $\sigma^S$ levels are normally kept low in log phase cells grown in rich media largely through rapid degradation by the ClpXP protease (37, 38) and the orphan response regulator, RssB (39, 40). However, consistent with our data, a significant elevation of $\sigma^S$ protein during exponential growth on acetate as the carbon supply in minimal medium was found in *Salmonella typhimurium* (41). Furthermore, $\sigma^S$ was recently shown to accumulate during log phase in mutants lacking the high-affinity inorganic phosphate (P) transporter, Pst (42). Similar to glucose, P3 is the preferred phosphate source for *E. coli*, and switching to other sources causes a severe reduction in growth rate (43). This suggests an analogous or overlapping mechanism for $\sigma^S$ stabilization which is growth rate-dependent may be used in the presence of either alternative carbon or phosphate substrates.

While these transcription factors are important for specifying aspects of the pattern, they explain only a subset of the genes. The hierarchical nature of the gene induction across growth conditions, the presence of sets, and the transcriptional responses appropriate to the specific circumstances (growth factors, intrinsic promoter strength and condition-associated transcriptional mechanisms) may underline the transcription pattern. The distribution mechanism may balance stable RNA synthesis with new gene transcription. The forward model of the rate of gene activation and decreased rRNA synthesis we observed is consistent with our data, a significant elevation of *S* levels are during exponential growth on acetate as the carbon supply in minimal medium was found in *S. typhimurium*. This is in line with early studies where the ratio of the level of the CRP regulon in *E. coli* and switching to the presence of either alternative carbon or phosphate substrates.

**REFERENCES**
45. Zhou, Y. N., and Jin, D. J. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 2908–2913
46. Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001) J. Mol. Biol. 305, 673–688
47. Gentry, D. R., and Cashel, M. (1996) Mol. Microbiol. 19, 1373–1384
48. Norris, T. E., and Koch, A. L. (1972) J. Mol. Biol. 64, 633–649
49. Bremer, H., and Dennis, P. P. (1996) in Escherichia coli and Salmonella (Neidhardt, F. C., ed) Vol. 2, pp. 1553–1569, ASM Press, Washington, D. C.
50. Hastjarjo, T., Silberberg, A., and Hursh, S. R. (1990) J. Exp. Anal. Behav. 53, 155–161
51. Bockhorst, J., Craven, M., Page, D., Shavlik, J., and Glasner, J. (2003) Bioinformatics 19, 1227–1235