

## Effect of an Artificial RNA Marker on Gene Expression in *Escherichia coli*

Don L. Tucker,<sup>1</sup> Fathi Karouia,<sup>1</sup> Jim Wang,<sup>2</sup> Yi Luo,<sup>2</sup> Tong-Bin Li,<sup>2</sup>  
Richard C. Willson,<sup>1,3</sup> Yuriy Fofanov,<sup>2</sup> and George E. Fox<sup>1,3\*</sup>

Department of Biology and Biochemistry,<sup>1</sup> Department of Computer Science,<sup>2</sup> and Department of  
Chemical Engineering,<sup>3</sup> University of Houston, Houston, Texas 77204-5001

Received 25 August 2004/Accepted 4 January 2005

**Transcriptional analysis was used to examine the effect of a genomically encoded artificial RNA on *Escherichia coli* in rich and minimal media. Only the expression of a single gene, *deoC*, was unequivocally affected under both conditions. *E. coli* marker strains of this type may be useful in monitoring the fate and transport of bacteria in various applications.**

A previously constructed *Escherichia coli* tracking strain known as PCPHR expresses a genomic gene encoding an artificial RNA (aRNA) which contains a unique 17-nucleotide identifier sequence (2). The aRNA gene was originally derived from *Vibrio proteolyticus* 5S rRNA by deleting a key functional region and replacing it with the identifier sequence (21). The aRNA product is expressed at high levels from an rRNA promoter and accumulates in the cell without an obvious effect on growth rate (2). Although it resembles 5S rRNA, the aRNA is inactive and not integrated into the *E. coli* ribosomes (2, 11, 14, 21). The high accumulation level of the aRNA allows its presence to be readily detected by many routine molecular methods, including RNA size, hybridization, fluorescence, and immunohybridization (15, 16, 22). Large numbers of alternative but equally unique identifier sequences could be introduced into the aRNA (12), thereby making it possible to simultaneously differentiate between multiple sources of *E. coli*. These aRNA strains are, however, genetically modified organisms whose use in field applications, such as source tracking or bioremediation, would require regulatory approval.

The issue, which was addressed here, is whether or not the presence of the aRNA tracking gene will significantly modify the behavior of the cells carrying it. *E. coli* has eight genes for 5S rRNA that are found in seven operons. The synthesis of the rRNAs is strongly linked with cell growth and subject to sophisticated regulation (9). For example, it has been shown that if the levels of 16S and 23S rRNA are artificially reduced by deletions in one or more rRNA operons, the cell will up-regulate its remaining rRNA operons to compensate for the loss (9). In contrast, the successive deletion of 5S rRNA genes revealed that there is apparently no ability to directly compensate for underexpression of 5S rRNA (3). These results suggest that the addition of an artificial RNA to the genome would not have a significant impact on other cellular processes. Nevertheless, the safety and utility of tracking strains expressing aRNAs that resemble 5S rRNA will depend on whether their

presence affects other components of the cellular machinery in a significant or unexpected way.

In order to address this issue, arrays of PCR products targeting each known open reading frame in the *E. coli* genome were used to determine what transcriptional changes are induced in the PCPHR strain compared to the wild-type (WT) *E. coli* EMG2 strain. Such arrays have been widely employed in the study of *E. coli* gene expression with considerable success (5–7, 17, 23–25). These experiments have generated extensive lists of genes that respond to various experimental variables (10).

Cultures of the WT EMG2 and PCPHR strains were grown aerobically in either rich YT medium (8 g Bacto tryptone, 5 g yeast extract, 5 g NaCl per liter) or defined minimal MOPS (morpholinepropanesulfonic acid) medium supplemented with glucose (19). Total RNA was harvested from the mid-log phase of growth. RNA samples from both strains were isolated twice from YT medium and three times from M9 minimal medium. Hybridizations were conducted using matched pairs of Panorama *E. coli* gene arrays (Sigma-Genosys Biotechnologies Inc., Houston, TX). These arrays were spotted with PCR products in duplicate by the manufacturer. They contain 4,290 PCR products coinciding with all known *E. coli* MG1655 open reading frames. The arrays were rinsed, prehybridized (65°C for 1 to 4 h), hybridized (65°C for 16 to 18 h), washed, and wrapped in Saran Wrap as described previously (24) with minor modifications (25). Exposed phosphor screen images (Fuji BAS-IP MS phosphor screen from FUJIFILM Medical Systems USA Inc., Stafford, TX) were visualized at a pixel density of 50  $\mu\text{m}$  (40,000 dots/cm<sup>2</sup>) with a Storm 860 PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The array membranes used in these experiments were subsequently stripped of hybridized probe by boiling inverted arrays in a microwave (100°C) in 200 ml of stripping solution (10 mM Tris, pH 7.5, 1 mM EDTA, 1% sodium dodecyl sulfate) and reused as many as eight times. Two hybridization probes were created from each RNA sample for hybridization to each array in an array pair resulting in four hybridizations for each strain grown in rich YT medium and six hybridizations each in M9 minimal medium.

A software package, ImageAnalyzer, was developed to quantify the hybridization signals. ImageAnalyzer utilizes spe-

\* Corresponding author. Mailing address: Dept. Biology and Biochemistry, University of Houston, Houston, TX 77204-5001. Phone: (713) 743-8363. Fax: (713) 743-8351. E-mail: fox@uh.edu.

TABLE 1. Genes expressed in *E. coli* PCPHR and WT EMG2 in M9 (minimal) and YT (rich) media, sorted by fold change in gene expression

Gene	Gene no.	<i>P</i> value <sup>a</sup>	Change in gene expression (fold)	Cotranslation <sup>b</sup>	Gene product description	Functional group
M9 minimal medium-expressed genes						
PCPHR expressed						
<i>deoC</i>	b4381	$7.7 \times 10^{-6}$	2.34	+	2-Deoxyribose-5-phosphate aldolase	Salvage of nucleosides and nucleotides
<i>deoB</i>	b4383	0.000565	1.48	+	Phosphopentomutase	Salvage of nucleosides and nucleotides
<i>serA</i>	b2913	0.007776	1.447		D-3-Phosphoglycerate dehydrogenase	Amino acid biosynthesis; serine
<i>rplL</i>	b3986	0.011943	1.387		50S ribosomal subunit protein L7/L12	Ribosomal proteins; synthesis, modification
<i>ybiX</i>	b0804	0.023121	1.299		Putative enzyme	Unknown
<i>metJ</i>	b3938	0.029677	1.363		Methionine sulfoximine	Amino acid biosynthesis; methionine
<i>deoD</i>	b4384	0.031111	1.303	+	Purine-nucleoside phosphorylase	Salvage of nucleosides and nucleotides
<i>moaE</i>	b0875	0.032103	1.32		Molybdopterin converting factor	Biosynthesis of cofactors, carriers; molybdopterin
<i>aceB</i>	b4014	0.042603	1.326		Malate synthase A	Central intermediate metabolism; glyoxylate bypass
WT EMG2 expressed						
<i>(vijC)</i>	b3963	0.039152	1.358		<i>fabR</i> transcriptional repressor	Unknown
YT-expressed genes						
PCPHR expressed						
<i>fimI</i>	b4315	0.003396	2.958	+	Fimbrial protein	Surface structures
<i>fimA</i>	b4314	0.008934	2.499	+	Major type 1 subunit fimbriae (pilin)	Regulatory function
<i>deoC</i>	b4381	0.018262	2.025	+	2-Deoxyribose-5-phosphate aldolase	Salvage of nucleosides and nucleotides
<i>ykgM</i>	b0296	0.037316	1.625		Putative ribosome protein	Putative structure
<i>tmk</i>	b1098	0.049627	1.625		Thymidylate kinase	2'-Deoxyribonucleotide metabolism
<i>deoB</i>	b4383	0.048425	1.608	+	Phosphopentomutase	Salvage of nucleosides and nucleotides
<i>deoD</i>	b4384	0.004603	1.607	+	Purine-nucleoside phosphorylase	Salvage of nucleosides and nucleotides
<i>ribH</i>	b0415	0.030334	1.582		Riboflavin synthase, beta chain	Biosynthesis of cofactors, carriers: riboflavin
<i>chpR</i>	b2783	0.018250	1.518		Suppressor of inhibitor ChpA, PcmI-like	Cell processes
WT EMG2 expressed						
<i>glpA</i>	b2241	0.018796	1.898		<i>sn</i> -Glycerol-3-phosphate dehydrogenase	Energy metabolism, carbon: anaerobic respiration
<i>glpF</i>	b3927	0.009194	1.699		Facilitated diffusion of glycerol	Small molecule transport (carbohydrates, organic acid, alcohol)
<i>ydaC</i>	b1347	0.008747	1.637		orf, hypothetical protein	Unknown
<i>napA</i>	b2206	0.001326	1.549	+	Probable nitrate reductase 3	Energy metabolism, carbon: anaerobic respiration
<i>napH</i>	b2204	0.000005	1.548	+	Ferredoxin-type protein: electron transfer	Energy metabolism, carbon: electron transport
<i>yedE</i>	b1929	0.002599	1.512		Putative transport system permease	Putative transport protein
<i>yhhQ</i>	b3471	0.000817	1.503		orf, hypothetical protein	Unknown
<i>icc</i>	b3032	0.000005	1.496		Regulator of <i>lacZ</i>	Degradate small molecules: carbon compounds

<sup>a</sup> *P* value (the Student *t* test) of gene expression between cultures.

<sup>b</sup> +, potential cotranslation of a gene with other significantly expressed genes in this list (determined by genomic proximity and direction of transcription).

cific metrics that take into consideration the nonuniformity of pixels inside a spot. These metrics exclude the highest-intensity pixels, which may reflect spike noises, and lowest-intensity pixels resulting from the surface tension effect of the probe printing process. The results obtained with ImageAnalyzer were compared to results obtained with the demonstration version of the commercially available ArrayVision Software (Imaging Research Inc., St. Catharines, Ontario, Canada), and ImageAnalyzer was found to produce spot pixel density data of similar or better quality (data not shown). Analysis of variance and the Student *t* test were performed between the groups of arrays obtained for each probe/gene, and the genes were sorted according to their *P* values, which represent the level of significance in differential expression, and according to the changes (*n*-fold) in gene expression between samples. The spot finding and statistical analysis software can be obtained upon request from Yuriy Fofanov (yfofanov@bioinfo.uh.edu). A descriptive user's manual is available at <http://www.bioinfo.uh.edu/ImageAnalyzer/>.

Genes were determined to have significant changes in gene expression based on three previously identified criteria: (i) an overall *P* value of <0.05, which implies a 95% probability that a change in expression between strains or media was significant (5); (ii) an overall ratio of gene expression of >2.0 between strains or media (7); and (iii) a gene expression ratio of >2.0 in each of the individual comparisons. Combining statistical methods in this manner increases the probability that genes remaining after statistical analysis are in fact changing significantly under the experimental conditions (23, 25). The raw image data and results of the analysis of the expression studies are available at [http://www.bioinfo.uh.edu/publications/aRNA\\_marker/](http://www.bioinfo.uh.edu/publications/aRNA_marker/).

Repeated growth studies revealed at most minor differences in lag phase, growth rate, and final cell density between WT EMG2 and PCPHR. The average doubling times from four high-quality growth curves in minimal media were 75 min in EMG2 and 73 min in PCPHR with standard deviations of 13 and 12 min, respectively. In accordance with previously published data (2), the average doubling times based on six high-quality growth curves in rich media for EMG2 and PCPHR were 25.1 and 27.7 min with respective standard deviations of 1.7 and 2.6 min. As expected, rich-medium cultures grew faster and reached a higher final cell density than cultures grown in M9 minimal medium. These data suggest that the stable aRNA present in PCPHR does not appreciably affect growth rate or energy consumption in either rich or minimal medium.

The array studies revealed that the expression level of only one gene, *deoC*, which encodes a 2-deoxyribose-5-phosphate aldolase, in the PCPHR strain was incontrovertibly elevated in both minimal and rich media. To ensure that all putative aRNA-impacted genes in minimal media had been identified, the 10 genes with the greatest overall changes in gene expression ratios between EMG2 and PCPHR were also examined (Table 1). The role that many of these genes could play in response to the stable aRNA is unclear, but it is noteworthy that the genes *deoB* (which encodes phosphopentomutase) and *deoD* (which encodes purine-nucleoside phosphorylase) were present in this less-stringent analysis. Along with *deoA*, whose expression was not elevated, and *deoC*, these genes belong to the *deo* operon, which is involved in deoxyribonucleoside and

ribonucleoside catabolism (1, 4, 7, 8, 13, 18, 27). It therefore seems likely that the primary response to the presence of the aRNA gene in the PCPHR strain is an increase in ribonucleotide catabolism as the accumulating aRNA is gradually degraded. However, whereas *deoB* and *deoD* operate on ribonucleotides and deoxyribonucleotides, prior work suggests that the *deoC* gene product is involved only in deoxynucleoside catabolism (7, 13), and hence, its up-regulation was not actually expected. The PCPHR strain may have significant unexpected value in future work aimed at fully understanding the regulation and function of the *deo* operon.

When rich medium is considered alone (Table 1), the *fim* genes (*fimI* and *fimA*) also met all three criteria for increased gene expression in PCPHR compared to the wild-type strain EMG2 described in this report. When gene transcription in rich medium is analyzed using only the Student *t* test (*P* value, <0.05), all genes of the *fimA* operon (*fimAICDFGH*) were more highly transcribed in PCPHR (data not shown). FimI is involved in type 1 pili biosynthesis, while a repeating monomer of FimA, arrayed helically to form a hollow core fiber, serves as a major component of type 1 pili (20, 26). Type I pili are filamentous bacterial surface attachment organelles that have been associated with virulence (20). The reason for the increased *fimA* operon gene transcription is unclear.

None of the genes whose expression was elevated in EMG2 (i.e., expression decreased in PCPHR) passed the original >2.0-fold-change test for significance. Decreasing the stringency of the change (*n*-fold) in gene expression of YT medium-grown cultures from >2.0 to >1.5 revealed six additional genes with increased expression in PCPHR and eight genes with increased expression in EMG2. This slight increase in expression of some YT medium-grown PCPHR genes could be caused by the medium composition or the presence of additional false positives in the smaller YT medium data set. Thus, in conclusion, gene expression in the PCPHR strain is extremely similar to that in the WT EMG2 strain lacking the aRNA identifier in both minimal and rich media.

Quantitative methods of tracking specific microbial populations are essential to basic research on complex microbial ecosystems. Such capability is also needed in practical applications, such as source tracking and microbial monitoring in bioremediation. For example, the release of naturally occurring or genetically engineered organisms as part of a bioremediation effort raises considerable public concern regarding the possibility of organisms escaping to nearby ecosystems. Groups participating in such activities will require incontrovertible evidence as to where their organisms did or did not end up. Appropriately designed tracking strains possibly based on aRNA identifiers can meet these requirements.

This work was supported in part by grants from the Gulf Coast Hazardous Waste Research Center (081UHH2793) and Environmental Institute of Houston to G.E.F., the Institute of Space Systems Operations to D.L.T., G.E.F., and R.C.W., and the Texas Learning and Computational Center to Y.F., R.C.W., and B. Montgomery Pettit.

#### REFERENCES

- Ahmad, S. I., and R. H. Pritchard. 1969. A map of four genes specifying enzymes involved in catabolism of nucleosides and deoxynucleosides in *Escherichia coli*. *Mol. Gen. Genet.* **104**:351–359.
- Ammons, D., J. Rampersad, and G. E. Fox. 1998. A genomically modified marker strain of *Escherichia coli*. *Curr. Microbiol.* **37**:341–346.

3. Ammons, D., J. Rampersad, and G. E. Fox. 1999. 5S rRNA gene deletions cause an unexpectedly high fitness loss in *Escherichia coli*. *Nucleic Acids Res.* **27**:637–642.
4. Amouyal, M., L. Mortensen, H. Buc, and K. Hammer. 1989. Single and double loop formation when *deoR* repressor binds to its natural operator sites. *Cell* **58**:545–551.
5. Arfin, S. M., A. D. Long, E. T. Ito, L. Toller, M. M. Riehle, E. S. Paegle, and G. W. Hatfield. 2000. Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J. Biol. Chem.* **275**:29672–29684.
6. Arnold, C. N., J. McElhanon, A. Lee, R. Leonhart, and D. A. Siegle. 2001. Global analysis of *Escherichia coli* gene expression during the acetate-induced acid tolerance response. *J. Bacteriol.* **183**:2178–2186.
7. Barbosa, T. M., and S. B. Levy. 2000. Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *J. Bacteriol.* **182**:3467–3474.
8. Buxton, R. S., H. Albrechtsen, and K. Hammer-Jespersen. 1977. Overlapping transcriptional units in the *deo* operon of *Escherichia coli* K-12. *J. Mol. Biol.* **114**:287–300.
9. Condon, C., C. Squires, and C. L. Squires. 1995. Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59**:623–645.
10. Conway, T., and G. K. Schoolnik. 2003. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol. Microbiol.* **47**:879–889.
11. D'Souza, L. M., R. C. Willson, and G. E. Fox. 2000. Expression of marker RNAs in *Pseudomonas putida*. *Curr. Microbiol.* **40**:91–95.
12. D'Souza, L. M., M. Larios-Sanz, R. A. Setterquist, R. C. Willson, and G. E. Fox. 2003. Small RNA sequences are readily stabilized by inclusion in a carrier rRNA. *Biotech. Prog.* **19**:734–738.
13. Hammer-Jespersen, K., A. Munch-Petersen, P. Nygaard, and M. Schwartz. 1971. Induction of enzymes involved in the catabolism of deoxyribonucleosides and ribonucleosides in *Escherichia coli* K 12. *Eur. J. Biochem.* **19**:533–538.
14. Hedenstierna, K. O. F., H. Y. Lee, Y. Yang, and G. E. Fox. 1993. A prototype stable RNA identification cassette for monitoring plasmids of genetically engineered microorganisms. *Syst. Appl. Microbiol.* **16**:280–286.
15. Kourentzi, K. D., G. E. Fox, and R. C. Willson. 2001. Microbial detection with low molecular weight RNA. *Curr. Microbiol.* **43**:444–447.
16. Kourentzi, K. D., G. E. Fox, and R. C. Willson. 2002. Microbial identification by immunohybridization assay of artificial RNA labels. *J. Microbiol. Methods* **49**:301–306.
17. Mori, H. 2004. From the sequence to cell modeling: comprehensive functional genomics in *Escherichia coli*. *J. Biochem. Mol. Biol.* **37**:83–92.
18. Mortensen, L., G. Dandanell, and K. Hammer. 1989. Purification and characterization of the *deoR* repressor of *Escherichia coli*. *EMBO J.* **8**:325–331.
19. Neidhardt, F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. *J. Bacteriol.* **119**:736–747.
20. Orndorff, P. E., A. Devapali, S. Palestrant, A. Wyse, M. L. Everett, R. R. Bollinger, and W. Parker. 2004. Immunoglobulin-mediated agglutination of and biofilm formation by *Escherichia coli* K-12 require the type 1 pilus fiber. *Infect. Immun.* **72**:1929–1938.
21. Pitulle, C., K. O. F. Hedenstierna, and G. E. Fox. 1995. A novel approach for monitoring genetically engineered microorganisms by using artificial, stable RNAs. *Appl. Environ. Microbiol.* **61**:3661–3666.
22. Pitulle, C., L. Dsouza, and G. E. Fox. 1997. A low molecular weight artificial RNA of unique size with multiple probe target regions. *Syst. Appl. Microbiol.* **20**:133–136.
23. Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner. 1999. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**:3821–3835.
24. Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.* **181**:6425–6440.
25. Tucker, D. L., N. Tucker, and T. Conway. 2002. Gene expression profiling of the pH response in *Escherichia coli*. *J. Bacteriol.* **184**:6551–6558.
26. Valenski, M. L., S. L. Harris, P. A. Spears, J. R. Horton, and P. E. Orndorff. 2003. The product of the *fimI* gene is necessary for *Escherichia coli* type 1 pilus biosynthesis. *J. Bacteriol.* **185**:5007–5011.
27. Valentin-Hansen, P., K. Hammer, J. E. Love Larsen, and I. Svendsen. 1984. The internal regulated promoter of the *deo* operon of *Escherichia coli* K-12. *Nucleic Acids Res.* **12**:5211–5224.