JB Journals.ASM.org

COMMENTARY

Where To Begin? Mapping Transcription Start Sites Genome-Wide in *Escherichia coli*

Joseph T. Wade[a,b]

Wadsworth Center, New York State Department of Health, Albany, New York, USA[a]; Department of Biomedical Sciences, University at Albany, Albany, New York, USA[b]

Recent genome-wide studies of bacterial transcription have revealed large numbers of promoters located inside genes. In this issue of the *Journal of Bacteriology*, Thomason and colleagues (J. Bacteriol. 197:18–28, 2015, doi:10.1128/JB.02096-14) map transcription start sites in *Escherichia coli* on an unprecedented scale. This work provides important insights into the regulation of transcripts that initiate inside genes and sources of variability between studies aimed at identifying these RNAs.

More than 50 years of work have led to a detailed mechanistic understanding of bacterial RNA synthesis. A single RNA polymerase is responsible for all transcription and is directed to specific promoter sequences by a σ factor. The level of transcription can be modulated by transcription activators and repressors that typically bind close to the promoter sequences. Until recently, the large majority of transcripts were believed to be mRNAs, with smaller numbers of noncoding RNAs. The noncoding RNAs included tRNAs, rRNAs, and so-called "small RNAs," which often have regulatory functions (1). However, the advent of genomic technologies has dramatically altered our view of bacterial transcriptomes. An early microarray study suggested the existence of thousands of antisense RNAs (2). This result was largely ignored until next-generation sequencing technologies facilitated interrogation of bacterial transcriptomes with unprecedented sensitivity. In particular, several methods based on transcriptome sequencing (RNA-seq) were developed to map transcription start sites (TSS). When applied to a variety of bacterial species, these methods revealed thousands of TSS inside genes, in the antisense orientation relative to the overlapping gene (3, 4). These TSS are referred to as "asTSS," and the RNAs they generate are known as "asRNAs." Genome-wide TSS mapping studies also suggested the existence of a similarly high number of TSS inside genes, in the sense orientation relative to the overlapping gene (4). These TSS are referred to as "iTSS" (internal TSS), and the RNAs they generate are known as "intraRNAs" (5). Such "pervasive transcription" has also been described in eukaryotes, and the function of the newly identified RNAs in all kingdoms of life is a source of great debate (6–8). In bacterial systems, a significant challenge in this emerging field has been the inconsistency between different studies in the number and composition of the RNAs identified. In this issue of the *Journal of Bacteriology*, Thomason and colleagues have directly addressed this challenge by generating the most comprehensive collection of TSS data sets to date for the model bacterium *Escherichia coli* (9). These easily accessible primary transcriptome data sets promise to be a valuable resource for future studies of pervasive transcription.

PERVASIVE TRANSCRIPTION OF THE *E. COLI* GENOME

Thomason and colleagues used the well-established differential RNA-seq (dRNA-seq) method (4) to map TSS in *E. coli* cells grown under a variety of conditions (9). dRNA-seq relies on the fact that primary transcripts are triphosphorylated at the 5′ ends, whereas processed RNAs are not. Thus, the authors identified 14,868 putative TSS, many more than identified to date. The validity of the putative TSS was supported by bioinformatic analysis showing enrichment of the expected promoter elements upstream of the TSS and experimental validation of selected TSS, using Northern blots to detect the associated RNAs. The genomic context of TSS (Fig. 1) was consistent with previous studies: TSS were enriched in intergenic regions relative to the entire genome (only ~11% of the *E. coli* genome is intergenic), but the majority of TSS were located inside genes, with similar numbers of iTSS and asTSS. Some iTSS and asTSS are <300 bp upstream of an annotated gene, suggesting they may be TSS for mRNAs; however, 83% of iTSS and 88% of asTSS are >300 bp from an annotated gene start, indicating that most intragenic TSS likely correspond to novel, noncoding RNAs. Thus, the *E. coli* genome is pervasively transcribed.

SOURCES OF VARIABILITY IN TSS MAPPING DATA SETS

Thomason and colleagues directly compared the lists of putative asRNAs identified in 7 studies, including their own (9). The largest overlap between any of these studies represented only 33% of the asRNAs identified. One possible explanation for the variability in TSS identified by different studies is that all studies suffer from a high false-positive rate. While this is likely to be the case for a few studies that have almost no overlap with others, three of the lists of TSS were supported by experimental validation of selected examples (3, 9, 10), and three were supported by bioinformatic analysis showing enrichment of the expected promoter elements upstream of the TSS (3, 9, 11). A more likely explanation for the variability between studies is that there is an extremely high number of low-abundance TSS, with each study sampling this pool differently due to methodological differences. Thomason and colleagues' data strongly support this explanation (9). They sequenced dRNA-seq

Accepted manuscript posted online 20 October 2014

Citation Wade JT. 2015. Where to begin? Mapping transcription start sites genome-wide in *Escherichia coli*. J Bacteriol 197:4–6. doi:10.1128/JB.02410-14.

Editor: R. L. Gourse

Address correspondence to joseph.wade@health.ny.gov.

doi:10.1128/JB.02410-14

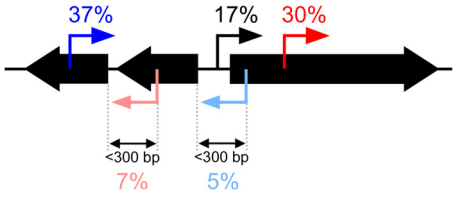

FIG 1 Summary of transcription start sites identified by Thomason and colleagues (9). The proportion of transcription start sites (TSS) in different classes is indicated. Genes are represented by thick arrows. TSS are represented by bent arrows. TSS in black indicate those upstream of annotated genes, <300 bp from the gene start. iTSS are shown in red, with those <300 bp upstream of an annotated gene shown in pale red. asTSS are shown in blue, with those <300 bp upstream of an annotated gene shown in pale blue. The percentages shown indicate the proportion of each class of TSS. Note that "orphan TSS" (>300 bp upstream of an annotated gene start and not overlapping a gene) and TSS that are both iTSS and asTSS are not shown.

libraries on two different instruments, with each instrument requiring a subtly different method for library preparation. Strikingly, the most dissimilar data sets were not those generated from cells grown under different conditions but rather were those from libraries sequenced on different instruments. Based on this result, large differences between studies would be expected, with different groups often using very different methods for library preparation, and sequencing libraries on unrelated instruments. Differences in growth conditions are also a likely factor, although the majority of TSS identified by Thomason and colleagues were detected under at least two of the three conditions tested (9).

REGULATION OF asRNAS AND intraRNAS

Previous studies have shown that transcription of asRNAs and intraRNAs is suppressed by H-NS and Rho (12, 13) and that they are subject to degradation by RNase III (14, 15). These phenomena are believed to be mechanisms to silence asRNA and intraRNA expression (7), suggesting that these RNAs are wasteful rather than functional. Thomason and colleagues analyzed expression of several asRNAs in RNase III and RNase E mutants (9). As expected, several asRNAs were more abundant in the mutant strains than in the wild-type strain. However, some asRNAs showed decreased expression in the RNase mutants. Thus, the effects of RNases on asRNA expression are not straightforward and may be dependent on base-pairing interactions with other RNAs, a phenomenon recently described in *E. coli* (14).

Thomason and colleagues' data indicate that many asRNAs and intraRNAs are differentially regulated according to the growth conditions; significant changes in RNA abundance were observed between rich and minimal media (9). Regulation of asRNA and intraRNA expression is suggestive of functional roles for these RNAs. A lower proportion of asRNAs and intraRNAs were differentially expressed than mRNAs, suggesting that transcription of many asRNAs and intraRNAs is driven solely by promoter contacts with RNA polymerase bound to σ^{70}, the primary sigma factor in *E. coli*. Thus, it seems likely that many asRNAs and intraRNAs are nonfunctional—transcribed from spurious promoters—but that others have functions. This is consistent with conservation of a subset of asRNAs and intraRNAs in *Shewanella* (16).

FUTURE PROSPECTS

The TSS identified by Thomason and colleagues will be a valuable resource for mRNA TSS and will likely be the benchmark for future studies of asRNAs and intraRNAs in *E. coli*. Many questions remain in this emerging field. First and foremost, is that the function, if any, of these RNAs remains a mystery. Although a handful of asRNAs and one intraRNA have been shown to regulate expression of mRNAs (17–19), the vast majority remain uncharacterized. Other key questions include how these RNAs are regulated, whether they impact expression of the overlapping gene, whether they are translated, what is their impact on cell fitness, and what is their role in genome evolution?

ACKNOWLEDGMENT

I thank David Grainger for comments on the manuscript.

REFERENCES

1. **Waters LS, Storz G.** 2009. Regulatory RNAs in bacteria. Cell **136:**615–628. http://dx.doi.org/10.1016/j.cell.2009.01.043.
2. **Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM.** 2000. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. Nat Biotechnol **18:**1262–1268. http://dx.doi.org/10.1038/82367.
3. **Dornenburg JE, DeVita AM, Palumbo MJ, Wade JT.** 2010. Widespread antisense transcription in *Escherichia coli*. mBio **1**(1):e00024–10. http://dx.doi.org/10.1128/mBio.00024-10.
4. **Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J.** 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464:**250–255. http://dx.doi.org/10.1038/nature08756.
5. **Bilusic I, Popitsch N, Rescheneder P, Schroeder R, Lybecker M.** 2014. Revisiting the coding potential of the *E. coli* genome through Hfq coimmunoprecipitation. RNA Biol **11:**641–654. http://dx.doi.org/10.4161/rna.29299.
6. **Lybecker M, Bilusic I, Raghavan R.** 2014. Pervasive transcription: detecting functional RNAs in bacteria. Transcription **5:**e944039. http://dx.doi.org/10.4161/21541272.2014.944039.
7. **Wade JT, Grainger DC.** 2014. Pervasive transcription: illuminating the dark matter of bacterial transcription. Nat Rev Microbiol 12:647–653. http://dx.doi.org/10.1038/nrmicro3316.
8. **Kapranov P, St Laurent G.** 2012. Dark matter RNA: existence, function, and controversy. Front Genet **3:**60. http://dx.doi.org/10.3389/fgene.2012.00060.
9. **Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G.** 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. J Bacteriol **197:**18–28. http://dx.doi.org/10.1128/JB.02096-14.
10. **Shinhara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, Tomita M, Mori H, Kanai A.** 2011. Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. BMC Genomics **12:**428. http://dx.doi.org/10.1186/1471-2164-12-428.
11. **Raghavan R, Sloan DB, Ochman H.** 2012. Antisense transcription is pervasive but rarely conserved in enteric bacteria. mBio **3**(4):e00156–12. http://dx.doi.org/10.1128/mBio.00156-12.
12. **Singh S, Singh N, Bonocora RP, Fitzgerald DM, Wade JT, Grainger DC.** 2014. Widespread suppression of intragenic transcription initiation by H-NS. Genes Dev **28:**214–219. http://dx.doi.org/10.1101/gad.234336.113.
13. **Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R.** 2012. Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. Genes Dev **26:**2621–2633. http://dx.doi.org/10.1101/gad.196741.112.
14. **Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R.** 2014. The double-stranded transcriptome of *Escherichia coli*. Proc Natl Acad Sci U S A **111:**3134–3139. http://dx.doi.org/10.1073/pnas.1315974111.
15. **Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de los Mozos IR, Vergara-Irigaray M, Segura V, Fagegaltier D, Penadés JR, Valle J, Solano C, Gingeras TR.** 2011. Genome-wide antisense transcription drives mRNA processing in bacteria. Proc Natl Acad Sci U S A **108:**20172–20177. http://dx.doi.org/10.1073/pnas.1113521108.
16. **Shao W, Price MN, Deutschbauer AM, Romine MF, Arkin AP.** 2014.

Conservation of transcription start sites within genes across a bacterial genus. mBio **5**(4):e01398-14. http://dx.doi.org/10.1128/mBio.01398-14.

17. **Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G.** 2014. MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev **28:**1620–1634. http://dx.doi.org/10.1101/gad.243485.114.
18. **Thomason MK, Storz G.** 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet **44:**167-188. http://dx.doi.org/10.1146/annurev-genet-102209-163523.
19. **Georg J, Hess WR.** 2011. *cis*-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev **75:**286–300. http://dx.doi.org/10.1128/MMBR.00032-10.