

Many expressed genes in bacteria and yeast are transcribed only once per cell cycle

Michaël Bon, Simon J. McGowan, and Peter R. Cook¹

Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

ABSTRACT The steady-state levels of all mature transcripts expressed in bacteria and yeast have been cataloged, but we do not yet know the numbers of nascent transcripts and so RNA polymerases engaged on all genes. Such catalogs are presented here. As mRNA levels depend on the balance between synthesis and degradation, we use published data to calculate the numbers of engaged polymerases required to maintain these levels in the face of the known rate of degradation. Most genes, including essential ones, prove not to be transcribed most of the time, and many produce only one message per cell cycle. Some cells even fail to produce an essential message during a cycle, and so must depend on their mother's messages and/or proteins for survival. We speculate that evolution sets the rate of message production so low to conserve energy, minimize transcription-induced mutation, and permit regulation over the widest range.—Bon, M., McGowan, S. J., Cook, P. R. Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *FASEB J.* 20, E000–E000 (2006)

Key Words: database • nascent RNA • transcription rate • transcriptomics

THE RELATIVE NUMBERS of mature mRNAs and their half lives in *Escherichia coli* growing in Luria broth (LB) or M9 plus glucose (Glc) (which support division every 30 and 90 min, respectively) have been cataloged using microarrays (1). As mRNA levels depend on the balance between synthesis and degradation, we calculate the numbers of engaged polymerases required to maintain these levels in the face of the known rate of degradation. We first convert relative levels of each mRNA to absolute numbers knowing the total number of all messages in the cell (2), then calculate how many engaged polymerases are required to maintain these levels using the known rates of transcript elongation (2) and degradation. We also apply this approach to equivalent data from yeast. The results in the form of searchable catalogs and the sources, equations used, underlying assumptions, and validations are presented at <http://users.path.ox.ac.uk/~pcook/data/catalogs.html>. We find that most genes, including essential ones, are not transcribed most of the time; many produce only one message per cell cycle, some less than one.

As microarray data for structural RNA genes (e.g.,

encoding rRNA, tRNA) are not available, polymerase densities on these bacterial genes are calculated using the known rates of initiation (2). Values for all bacterial genes are also corrected for gene length and copy number, which depend on genome position and growth rate (2). We also consider only actively elongating polymerases, but not inactive ones that might be initiating, terminating, or paused on their templates; we also assume that each active polymerase eventually produces a mature message and that none abort to release an unstable transcript that is degraded too rapidly to be detected using a microarray. Even if the amounts of pausing and abortive transcription prove to be significant, they are unlikely to change our values by more than a few fold.

MATERIALS AND METHODS

Sources, equations used, underlying assumptions, and validations are presented at <http://users.path.ox.ac.uk/~pcook/data/catalogs.html>.

RESULTS

E. coli

Results show that in LB, only three open reading frames (ORFs) are associated with >0.5 polymerases/gene, with the most active (*sucA*) being associated with 0.6. All others are associated with <0.5 polymerases/gene (Fig. 1A), and so fewer than half the copies in the population are being transcribed at any moment. Moreover, many are transcribed rarely (i.e., about once per cell cycle). Thus, an engaged polymerase traverses a typical gene in ~20 s (roughly one-hundredth of the cycle), and the mean and median values are in this range (i.e., 0.017 and 0.01 polymerases/gene respectively; Fig. 1A). In contrast to ORFs, many polymerases are engaged on the 7 *nm* cistrons encoding ribosomal RNAs (Fig. 1B). In nutrient-poor M9 + Glc, biosynthetic capacity switches away from ribosome genesis (2)

¹ Correspondence: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK. E-mail: peter.cook@path.ox.ac.uk
doi: 10.1096/fj.06-6087fje

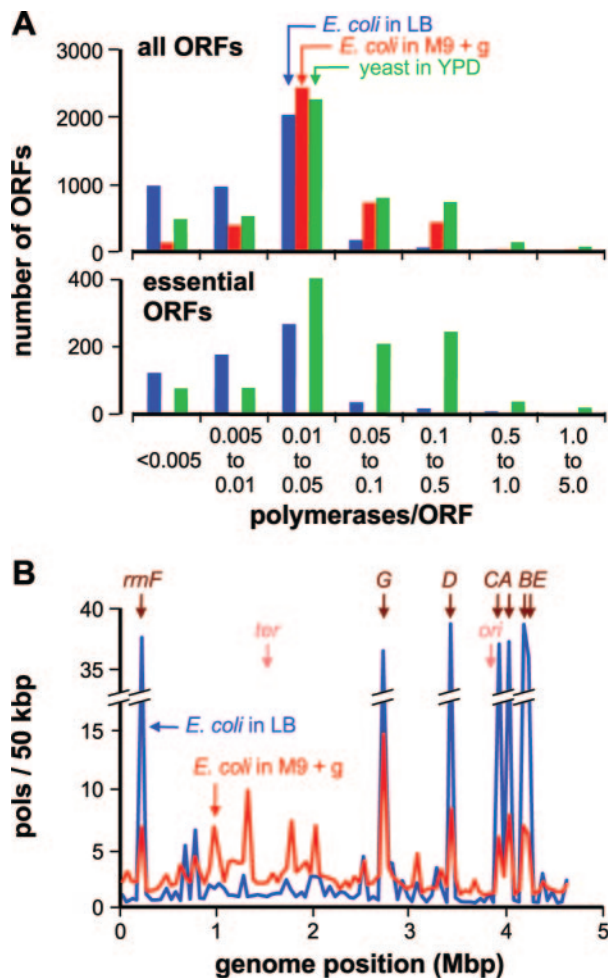


Figure 1. Densities of engaged polymerases. *A*) The range of transcriptional activity of all ORFs and essential ORFs in bacteria and yeast. The number of ORFs with different polymerase densities is indicated for bacteria grown in LB or M9 + Glc and for yeast in YPD. *B*) The number of polymerases engaged on successive 50 kbp segments (each containing ~50 genes) of the *E. coli* genome. Positions of the (active) *rm* operons (*A–G*) encoding rRNA and the origin (*ori*) plus terminus (*ter*) are indicated. Polymerases switch from the *rm* operons to ORFs on transfer from LB to M9 + Glc.

and the polymerase density on ORFs increases (Fig. 1*B*); polymerases are typically engaged every ~24 kbp along the genome in LB and every ~8.5 kbp in M9 + Glc (Fig. 1*B*). But despite this switch, most ORFs remain associated with few polymerases (Fig. 1). This increased transcription of ORFs during slow growth is accompanied by a reduced rate of translation (2), while mean mRNA half-lives in LB and M9 + Glc remain roughly the same (i.e., at 5.2 and 5.7 min, respectively; ref. 1).

About 600 genes are essential in LB (3) and so must be transcribed. Only one (*sucA*) is transcribed more often than not (i.e., associated with >0.5 polymerase); the rest are again mostly silent, with 295 essential genes being associated with <0.01 polymerase (Fig. 1*A*). In this medium, each cell contains ~3 genome equivalents, yet 46 essential genes produce on average less

than one message per cycle. If initiation occurs stochastically and follows Poisson statistics, it follows that the probability that all essential genes are transcribed at least once before a cell divides is essentially zero; therefore, most cells must depend for survival on mRNAs and/or proteins made in previous division cycles. [Unfortunately, data on the levels of all proteins and their half-lives are not yet available. However, in LB there are ~1000 copies/cell of each (abundant and essential) ribosomal protein (2), 1–15 copies/cell of each of the relevant messages (not shown), and 0.01–0.16 polymerases engaged on each gene (not shown).]

Results agree with others obtained using very different methods. For example, we calculate that each *rm* operon in LB is associated with 35–40 polymerases (Fig. 1*B*); this compares with the ~70 closely packed polymerases seen in the “Christmas trees” obtained in micrographs of spread DNA fibers (4). As 2 of the 7 operons can be deleted without affecting growth (5) and as inactive operons are missed in spreads, our value might be expected to be lower as it is an average of densities on both active and inactive operons. Values for ORFs also agree: spreads reveal one engaged polymerase every 10–20 kbp (4), equivalent to 0.05–0.1 polymerases/gene (genes are typically ~1 kbp long and closely packed), and again we would expect our values to be lower. Moreover, relative densities in genes around the *rm* operons (not shown) are similar to those seen in spreads (4). In addition, there are simply too few molecules of RNA polymerase in the cell to support transcription at levels more than a few fold higher than we find (see Assumptions and validation at the supporting web site). Finally, the relative numbers of polymerases on the structural RNA genes compared to all the ORFs are known accurately (2), and our values compare with these.

S. cerevisiae

Holstege *et al.* (6) similarly cataloged relative numbers of mature mRNAs and their half-lives in *S. cerevisiae* grown in yeast extract/peptone/dextrose (YPD; doubling time 90 min), and calculated the rate of mRNA production required to maintain those numbers. [Half-lives were determined using a temperature-sensitive mutant of the polymerase.] After correcting for copy number and using the known rate of elongation (7, 8), we calculate polymerase densities that are strikingly similar to those obtained with bacteria (Fig. 1*A*). Only 203 of 4942 ORFs (and 51 of 1004 essential ORFs; ref. 9) for which data are available are associated with >0.5 polymerases, while average and median values are 0.094 and 0.031 (and 0.13 and 0.05 for essential ORFs), respectively (Fig. 1*A*). Although each cell contains ~1.5 copies of a gene, 21 essential genes produce on average less than one message per cycle, and one-quarter of all ORFs are transcribed at a rate less than twice the minimum where a gene is copied once per division cycle (the minimum in this case is ~0.007 polymerases/gene; Fig. 1*A*); as a result, essentially all

cells must again depend on their mother's mRNAs and/or proteins for survival (not shown).

DISCUSSION

Using published data obtained from microarrays, we have generated a catalog of the numbers of RNA polymerases engaged on all the genes in bacteria and yeast. We first convert relative levels of each mRNA to absolute numbers knowing the total number of all messages in the cell, then calculate how many engaged polymerases are required to maintain these levels using the known rates of transcript elongation and degradation. We find that most genes, including essential ones, are not transcribed most of the time; many produce only one message per cell cycle, some less than one. As the transcription rate is so low, it follows that cells must depend on their mother's mRNAs and/or proteins for survival.

We might expect genes encoding abundant proteins to be relatively highly transcribed; they are, but the absolute rate is low. For example, each ribosomal protein is present in ~ 1000 copies per bacterial cell growing in LB (2); they are encoded by 1–15 copies/cell of each of the relevant messages, and these message levels are sustained by only 0.01–0.16 polymerases engaged on each copy of the relevant genes.

Unfortunately, no equivalent data for higher eukaryotes is available, but most so-called “active” polymerase II units are likely to be transcribed as rarely. If one excludes hyperactive units (e.g., chorion and heat shock in flies, globin, and actin in mammals; refs. 10–13), measured polymerase densities are <1 (14–17; reviewed in ref. 18). For example, fly genes encoding tubulin B1 and glyceraldehyde phosphate dehydrogenase associate with <1 polymerase at a time (19), and even the adenoviral unit—one of the most active in mammals—has only one every 7.5 kbp (20, 21). Studies of GFP-tagged polymerase II also show that most active human units spend most of their time waiting to be transcribed (22).

Why is the transcription rate set so low? Why did evolution not select a higher rate coupled to a less frequent translation of a higher number of messages? We can suggest three reasons. First, energy is saved, as fewer mRNAs are made (23). [Compare, for example, 10 ribosomes making 10 peptides on 10 mRNAs with the same peptide production by 10 ribosomes active on only one message.] Second, mutation is minimized. This follows because cytosine residues in single-stranded transcription bubbles deaminate $140 \times$ more readily than those in double-stranded DNA, and this ensures that the process of transcription is itself mutagenic (24). Third, polymerase densities can vary from the lowest to the highest possible (i.e., from about one polymerase engaging per cycle to close packing) enabling control over the widest possible range (i.e., from 0.01–10 polymerases per ORF, or ~ 1000 -fold; refs. 23,

25). But if only one transcript is made per division cycle, how might a cell respond rapidly to changes in the environment (e.g., in response to stress)? Polymerases frequently contact promoters without going on to produce a mature mRNA (e.g., during abortive initiation), so a rapid response can be achieved by increasing the ratio of productive to nonproductive contacts (26).

In many current studies it is assumed that expressed genes are active, and so are being transcribed. However, our results show that almost all such genes actually spend much more time not being transcribed. Then, such studies tell us more about the properties of silent genes that have the potential to be transcribed than those of active ones. We also hope these catalogs will prove useful to those studying how gene position affects activity. For example, the copy number of ORFs in the 1 Mbp around the bacterial origin of replication is double that in an equivalent region around the terminus, and the polymerase density is halved to compensate for this (Fig. 1B). The production of so few messages per cycle is also consistent with findings that gene expression is “noisy” (27), as some bacteria in the population will contain 0, 1, 2, 3 . . . messages and so express very different numbers of proteins. EJ

We thank Davide Marenduzzo for his help. M.B. and P.R.C. were supported by a grant from the Zilkha Fund of Lincoln College, Oxford.

REFERENCES

1. Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., and Cohen, S. N. (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. U. S. A.* **99**, 9697–9702
2. Bremer, H., and Dennis, P. P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. E. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds) pp 1553–1569, ASM Press, Washington, D.C.
3. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., *et al.* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684
4. French, S. L., and Miller, O. L. (1989) Transcription mapping of the *Escherichia coli* chromosome by electron microscopy. *J. Bacteriol.* **171**, 4207–4216
5. Condon, C., Liveris, D., Squires, C., Schwartz, I., and Squires, C. L. (1995) rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. *J. Bacteriol.* **177**, 4152–4156
6. Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728
7. Edwards, A. M., Kane, C. M., Young, R. A., and Kornberg, R. D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* **266**, 71–75

8. Shermoen, A. W., and O'Farrell, P. H. (1991) Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303–310
9. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre B., *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391
10. Osheim, Y. N., Miller, O. L., and Beyer, A. L. (1985) RNP particles at splice junction sequences on *Drosophila* chorion transcripts. *Cell* **43**, 143–151
11. O'Brien, T., and Lis, J. T. (1991) RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila melanogaster hsp70* gene. *Mol. Cell. Biol.* **11**, 5285–5290
12. Wijgerde, M., Grosveld, F., and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**, 209–213
13. Femino, A. M., Fay, F. S., Fogarty, K., and Singer, R. H. (1998) Visualization of single RNA transcripts in situ. *Science* **280**, 585–590
14. Laird, C. D., and Chooi, W. Y. (1976) Morphology of transcription units in *Drosophila melanogaster*. *Chromosoma* **58**, 193–218
15. McKnight, S. L., and Miller, O. L. (1979) Post-replicative nonribosomal transcription units in *D. melanogaster* embryos. *Cell* **17**, 551–563
16. Fakan, S., Leser, G., and Martin, T. E. (1986) Immunoelectron microscope visualization of nuclear ribonucleoprotein antigens spread within transcription complexes. *J. Cell Biol.* **103**, 1153–1157
17. Jackson, D.A., Iborra, F.J., Manders, E. M. M., and Cook, P. R. (1998) Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. *Mol. Biol. Cell.* **9**, 1523–1536
18. Jackson, D.A., Pombo, A., and Iborra, F. J. (2000) The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **14**, 242–254
19. Rougvie, A. E., and Lis, J. T. (1990) Postinitiation transcriptional control in *Drosophila melanogaster*. *Mol. Cell. Biol.* **10**, 6041–6045
20. Wolgemuth, D. J., and Hsu, M.-T. (1981) Visualization of nascent RNA transcripts and simultaneous transcription and replication in viral nucleoprotein complexes from adenovirus 2-infected HeLa cells. *J. Mol. Biol.* **147**, 247–268
21. Beyer, A.L., Bouton, A. H., Hodge, L. D., and Miller, O. L. (1981) Visualization of the major late R strand transcription unit of adenovirus serotype 2. *J. Mol. Biol.* **147**, 269–295
22. Kimura, H., Sugaya, K., and Cook, P. R. (2002) The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* **159**, 777–782
23. McAdams, H. H., and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 814–819
24. Beletskii, A., and Bhagwat, A. S. (1996) Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13919–13924
25. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2005) Gene regulation at the single-cell level. *Science* **307**, 1962–1965
26. Hsu, L. M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta* **1577**, 191–207
27. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science* **297**, 1183–1186

*Received for publication March 6, 2006.
Accepted for publication March 31, 2006.*