

What is faster, transcription or translation?

Transcription, the synthesis of mRNA from DNA, and translation, the synthesis of protein from mRNA, are the main pillars of the central dogma of molecular biology. How do the speeds of these two processes compare? This question is made all the more interesting as a result of observations like those shown in figure 1, namely, the existence of the beautiful “Christmas tree” structures observed in *E. coli* using electron microscopy. These stereotyped structures reflect the simultaneous transcription and translation of the same gene and raise the question of how the relative rates of the two processes compare making such synchronization possible.

Transcription of RNA in *E. coli* is carried out by ~1000-10,000 RNA polymerases (BNID 101440) proceeding at a maximal speed of about 40-80 bp/sec (BNID 104900, 104902). Translation of proteins in *E. coli* is similarly carried out by ~10,000-100,000 ribosomes (BNID 101441) and proceeds at a maximal speed of about 20 aa/sec (BNID 100059, 105067). Interestingly, since every 3 base pairs code for one amino acid, the rates of the two processes are nearly matched and as noted above, this is probably not accidental. Translation that is faster than transcription would cause the ribosome to “collide” with the RNA polymerase. Is there some biological reason why transcription is not significantly faster than

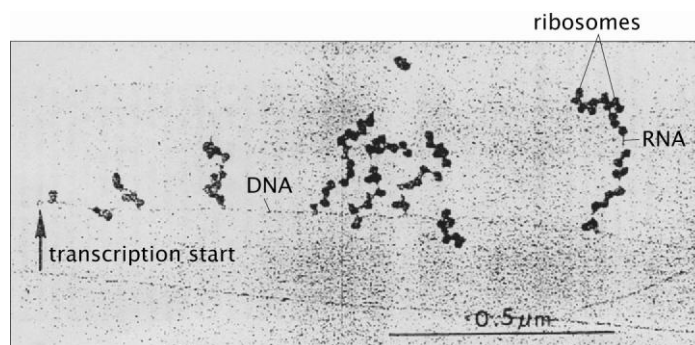


Figure 1: Electron microscopy image of simultaneous transcription and translation. The image shows bacterial DNA and its associated mRNA transcripts, each of which is occupied by ribosomes. (Fig. 3.9 of PBOC2, Adapted from O. L. Miller et al., Science 169:392, 1970.)

translation? There are probably multiple explanations, and one of them centers on the ability of the cell to carry out viable feedback. Arguments from control theory demand that to have stable regulation there needs to be time to respond to the feedback signal itself. If transcription were much faster than translation the response to a signal for a change in gene expression could overshoot the required response; the cell will produce a lot of mRNA and the signal that enough proteins are now present will come too late.

How are the rates of these key processes of the central dogma measured? This is an interesting challenge even with today's advanced techniques. Let's consider how we might attack this problem. One idea might be: "let's express a GFP and measure the time until it appears". To see the flaws in such an approach, check out the section on maturation time (minutes to an hour) of a fluorescent protein, which demonstrates a mismatch of time scales between the processes of interest and that of the putative readout. This question was even more challenging given the more limited experimental arsenal available in the 1970's when the answers were first convincingly obtained. Yet, in a series of clever experiments, Miller *et al.* using electron microscopy and Bremer and colleagues were able to come up with ingenious techniques based on using radioactive labeling to precisely determine these rates. As will be shown below, they relied on a subtle quantitative analysis in order to tease out the rates.

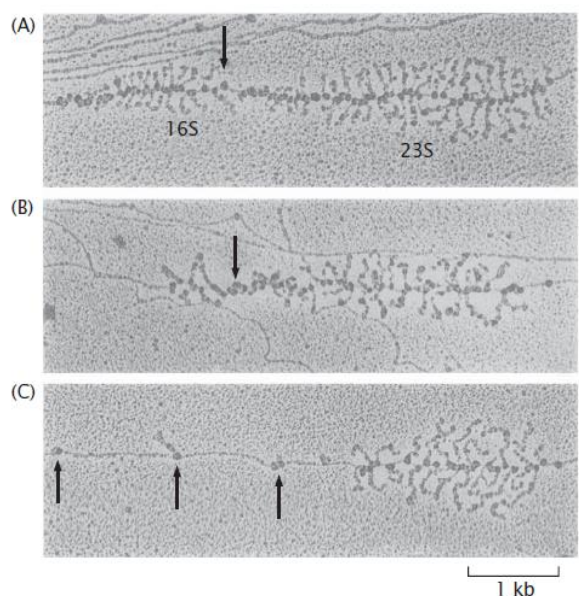
Measurements on transcription rates were based upon a trick in which transcription initiation was shut down by using the drug rifampin. Though no new transcription events can begin, those that are already proceeding continue unabated, i.e. rifampin inhibits the initiation of transcription, but not the elongation of RNA transcripts. As a result, this drug treatment effectively begins the running of a stopwatch which times how long since the last transcription process began. By fixing the cells and stopping the transcription process at different times after the drug treatment and then performing electron microscopy, resulting in images like that shown in Figure 2, it was possible to measure the length of RNA polymerase-free DNA. By taking into account the elapsed time since drug treatment the rate at which these polymerases are moving is inferred.

The measurement of translation rates similarly depended upon finding an appropriate stopwatch, but this time for the protein synthesis process. The crux of the method is the following: add a pulse of labeled amino acid at time zero and follow ("chase" as it is often called) the

fraction of labeled products of mass m with respect to the overall fraction that is labeled of any protein mass (λ). What would you expect to see? Immediately after the pulse of labeled amino acids one starts to see labeled proteins of mass m ; these are the nascent chains that were almost complete at time zero and which then had radioactive labeling incorporated only onto their unfinished ends. These newly synthesized proteins contain only a small fraction of radioactive labeling since only in the final stages of their translation were there labeled amino acids available. With time the fraction will increase as the newly synthesized chains have a larger proportion of their length labeled. After a time τ_m , depending on the transcript length, the whole chain will be labeled, as these are proteins that began their translation at time zero when the label was added. At this time one observes a breakpoint, a change in the slope of the accumulation of labeled proteins of mass m . Given the time that elapsed, τ_m , and by knowing how many amino acids are in a polypeptide chain of mass m it is possible to derive an estimate for the translation rate. For rigorous understanding of the method, one will benefit from the original study by Young & Bremer, *Biochem. J.* 1976. It remains as a reliable value for bacterial translation rate to this day.

What are the corresponding rates in eukaryotes? In yeast the rate of translation is about 2 fold slower, but one should note that the “physiological” temperature at which it is measured is 30°C whereas for *E. coli* measurements are at 37°C. As discussed in the vignette on the dependence of enzyme action on temperature, the slower rate is what one would expect based on the general dependence of a factor of 2-3 per 10 deg (Q10 value, BNID 100919). Transcription in mammalian cells was

Figure 2: (PBOC2, Figure 3.29) Effect of rifampin on transcription initiation. Electron micrographs of *E. coli* rRNA operons: (A) before adding rifampin, (B) 40 s after addition of rifampin, and (C) 70 s after exposure. No new transcripts have been initiated, but those already initiated are carrying on elongation. In parts (A) and (B) the arrow signifies the site where RNaseIII cleaves the nascent RNA molecule producing 16S and 23S ribosomal subunits. RNA polymerase molecules that have not been affected by the antibiotic are marked by the arrows in part (C). (Adapted from L. S. Gotta et al., *J. Bacteriol.* 20:6647, 1991.)



shown to consist of elongation at rates that are similar to those measured in *E.coli* (50-100 nts/sec, BNID 105113, 100662) interspersed with pauses leading to an average rate that is about an order of magnitude slower (~6 nts/sec, 100661). Remember that in eukaryotes, transcription and translation are not as intimately coupled as in bacteria due to the fact that they are actually geometrically segregated with transcription taking place in the nucleus and translation in the cytoplasm. Further, as introns are excised from transcripts prior to translation the correspondence between the rates is trickier. We have focused here on transcript elongation but in some cases the rate limiting process seems to be the initiation of transcription. This is the process in which the RNA polymerase complex is being assembled, the two DNA strands are separated to form a bubble and extra steps are performed. After several decades of intense investigation and ever more elaborate techniques at our disposal we seem to have arrived at the point where the quantitative description of the different steps of the central dogma can be integrated to reveal its intricate temporal dependencies.

How fast do RNAs and proteins degrade?

The central dogma focuses on the production of the great nucleic acid and protein polymers of biology. However, the control and maintenance of the functions of the cell depends upon more than just synthesis of new molecules. Degradation is a second key process in the lives of the macromolecules of the cell and is itself tightly controlled. Indeed, in the simplest model of mRNA production, the dynamics of the average level of mRNA is given by

$$\frac{d\bar{m}}{dt} = r - \gamma\bar{m},$$

where r is the rate of mRNA production and γ is the rate constant dictating mRNA decay. The steady-state value of the mRNA is given by

$$\bar{m} = \frac{r}{\gamma},$$

showing that to first approximation, it is the balance of these processes that controls the steady-state levels of these molecules.

The lifetime of mRNA molecules is usually short in comparison with one of the fundamental time scales of cell biology, namely, the time between cell divisions. As shown in Figure 1(A), for *E. coli*, the majority of mRNA molecules have lifetimes between 3 and 8 minutes. The experiments leading to these results were done by inhibiting transcription through the use of the drug rifampicin that interacts with the RNA polymerase and then querying the cells for their mRNA levels in two minute intervals after drug treatment. In particular, the RNA levels were quantified by hybridizing with complementary DNAs on a microarray and measuring the relative levels of fluorescence at different time points.

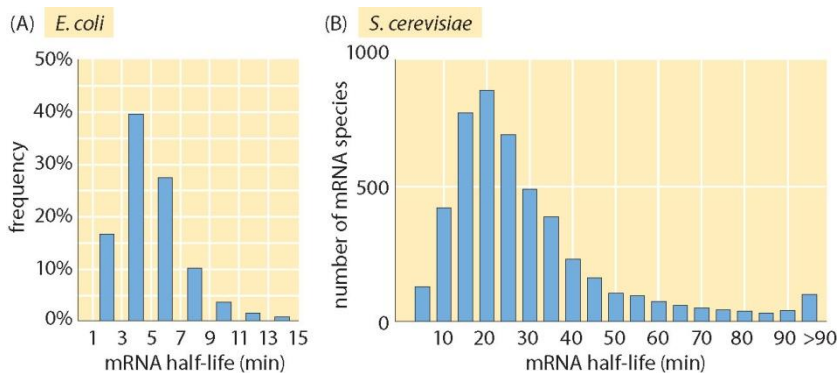


Fig. 1 – Measured half lives of mRNAs in *E. coli* and yeast.

Figure source for Yeast is Quantification of protein half-lives in the budding yeast proteome Archana Belle et al. PNAS

Given such genome-wide data, various hypotheses could be explored for the mechanistic underpinnings of the observed lifetimes. For example, is there a correlation between the abundance of certain messages and their decay rate? Are there secondary structure motifs or sequence motifs that confer differences in the decay rates? One of the big surprises of the measurements leading to Figure 1(A) is that none of the conventional wisdom on the origins of mRNA lifetime was found to be consistent with the data which revealed no clear correlation with secondary structure, message abundance or growth rate.

How far does Monod's statement that "what is true for *E. coli* is true for the elephant" (depicted by Monod in Figure 2) take us in our assessment of mRNA lifetimes in other organisms? The short answer is not very. Whereas the median mRNA degradation lifetime is roughly 5 minutes in *E. coli*, the mean lifetime is 21 minutes in the case of yeast (see Figure 1(B)) and a whopping 600 minutes (BNID 106869) in human cells. Interestingly, a clear scaling is observed with the cell cycle times for these three cell types, of roughly 20 minutes (*E. coli*), 90 minutes (yeast) and 3000 minutes (human), under the fast exponential growth rates that the cells of interest were cultivated in for these experiments (Yang, genome research 2003). As a rule of thumb, these results suggest a mRNA degradation time scale in these cases is thus about a quarter to a fifth of the fastest cell cycle.



Fig. 2 – L'éléphant et l'Escherichia coli, décembre 1972 "Tout ce qui est vrai pour le Colibacille est vrai pour l'éléphant". From: http://www.pasteur.fr/infosci/archives/mon/im_ele.html

Messenger RNA is not the only target of degradation. Protein molecules are themselves also the target of specific destruction, though generally, their lifetimes tend to be longer than the mRNAs that lead to their synthesis. Because of these long lifetimes, under fast growth rates what happens is that the number of copies of a particular protein per cell is reduced not because of an active degradation process, but simply because the cell doubles all its other constituents and divides into two daughters leaving each of the daughters with half as many copies of the protein of interest as were harbored in the mother cell. To understand

the dilution effect, imagine that all protein synthesis for a given protein has been turned off while the cell keeps on doubling its volume and shortly thereafter divides. In terms of absolute values, if the number of copies of our protein of interest before division is N , afterwards it is $N/2$. In terms of concentrations if it started with a concentration c , during the cell cycle it got diluted to $c/2$ by the doubling of the volume. This mechanism is especially relevant in the context of bacteria where the protein lifetimes are often dominated by the cell division time. As a result, the total protein loss rate α (the term carrying the same meaning as γ for mRNA) is the sum of a part due to active degradation and a part due to the dilution that occurs when cells divide and we can write the total removal rate in the form $\alpha = \alpha_{\text{active}} + \alpha_{\text{dilution}}$.

Like with the genome-wide studies of the mRNA lifetimes described above, protein lifetimes have been subjected to similar scrutiny by using drugs to inhibit further macromolecular synthesis, in this case cycloheximide, to inhibit the synthesis of further proteins (i.e. a translation inhibitor) and then by performing Western blots on the cell contents at various time points after translation inhibition. The Western blot technique is a scheme in which the proteins of interest are fished out by specific binding (for example by antibodies) to some part of the protein and the amount of protein is read off of the intensity of some reporter which has been calibrated against a standard. For the yeast proteome, the measured lifetimes are shown in Figure 3(A) and reveal the longer lifetimes of proteins in comparison with their mRNA counterparts, with the mean lifetime of roughly 40 minutes (BNID 104151), though the precision of these results still calls for the development of new methods for constructing such surveys.

Using modern fluorescence techniques it has become possible to measure degradation rates of human proteins *in vivo* without the need to lyse the cells. The long removal times observed in human cells are shown in Figure 3(B). The measurements were done by fusing the protein of interest to a fluorescent protein. Then, by splitting the population into

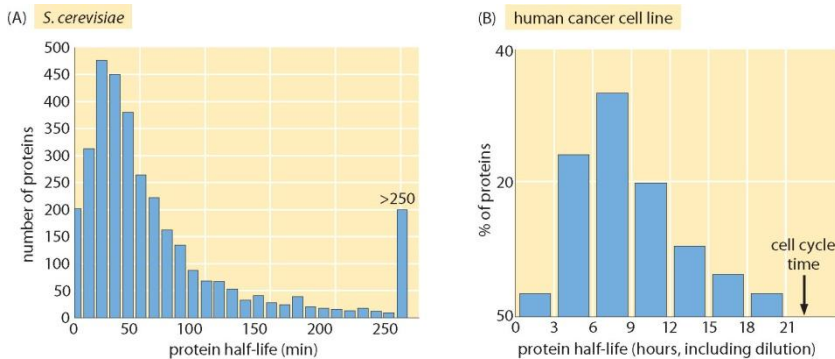
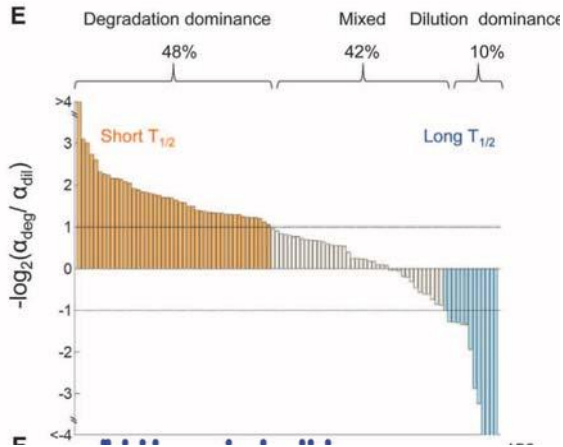


Fig. 3 – Measured half lives of proteins in yeast and a human cancer cell line.

two groups, one of which is photobleached and the other of which is not, and watching the reemergence of fluorescence in the photobleached population, it is possible to directly measure the degradation time. Ultimately, as noted above, the disappearance of proteins from cells results both from active degradation and from the fact that upon cell division, the protein complement of each cell is reduced by half. As shown in figure 4, for human cells there is an interesting interplay between active degradation and protein removal by dilution.



What is the error rate in transcription and translation?

The central dogma recognizes the flow of genomic information from the DNA into functional proteins via the acts of transcription which results in synthesis of messenger RNA and the subsequent translation of that RNA into the string of amino acids that make up a protein. This chain of events is presented in textbooks as a steady and deterministic process, but is in fact full of glitches in the form of mistakes or errors in the incorporation of nucleotides in the case of RNA or amino acids in the case of proteins. How common are these mistakes?

One approach to measure the error rate in transcription (Rosenberger and Foskett, 1981) is to use an *E. coli* mutant carrying a nonsense mutation in *lacZ* (i.e. one that puts a premature stop codon conferring a loss of function) and then assay for activity of this protein which enables utilization of the sugar lactose. The idea of the experiment is that the only way in which *functional* LacZ could be produced is through erroneous transcription resulting from a misincorporation event that bypassed the mutation. The sensitivity of the assay enabled measuring this residual activity due to "incorrect" transcripts giving an indication of an error rate in transcription of $\approx 10^{-4}$ (BNID 103453, Table 1), which in this well orchestrated experiment changed the spurious stop codon to some other amino acid, thus resurrecting the functional protein. Later measurements reviewed by Ninio suggested a value an order of magnitude better of 10^{-5} (BNID 105212). His analysis of these error rates led to the hypothesis of an error correction mechanism termed kinetic proofreading, paralleling a similar analysis performed by John Hopfield.

Recently, the error rate of RNA polymerase III, in charge of tRNA rather than mRNA transcription was measured in yeast. The authors (Alic et al 2007) were able to tease apart the contribution to fidelity arising from the first stages of selectivity from the later stages of proofreading. The total error rate was estimated to be 10^{-7} which should be viewed as a product of two error rates, $\approx 10^{-4}$ arising from initial selectivity and an extra factor of $\approx 10^{-3}$ arising from proofreading (105213, 105214).

Perhaps the best way to develop intuition for these error rates is again through analogy. An error rate of 10^{-4} corresponds to the authors of this book making one typo every several vignettes. An error rate of 10^{-7} corresponds more impressively (and impossibly for book authors) to one error in a thousand page textbook.

Recently, incorporation into the genome of GFP that is in the wrong reading frame facilitated the study of error rates in terms of frameshifts in the bacterium *B. Subtilis*. A high error rate at about 1% was observed indicating that the prevalence and implications of errors are still far from completely understood.

Error rates are generally thought to be about an order of magnitude higher than transcription error rates. An error in transcription would lead to many erroneous protein copies whereas an error in translation affects only one protein copy. Moreover, the correspondence of 3 nucleotides to one amino acid means that mRNA messages require higher fidelity per “letter” to achieve the same overall error rate. A standing challenge is to elucidate what limits decreasing the error rates in biological systems even further, say to values similar to those achieved by DNA polymerase. Is there a biophysical tradeoff at play or maybe the observed error rates have some selective advantages ?

Table 1: Error rates in transcription and translation

Organism	Errors per bp/codon	BNID, Measurement method, Comments
Transcription		
<i>E. coli</i>	10^{-4} per bp	103453, Activity in strains carrying lacZ mutations
<i>E. coli</i>	10^{-5} per bp	105212, In vitro- selection for rifampicin resistance and increased leakiness of an early, strongly polar nonsense mutation of lacZ.
<i>Saccharomyces cerevisiae</i> , RNA pol III	2×10^{-4} per bp	105213, determined based on selectivity
Translation		
<i>E. coli</i>	0.0013 to 0.004 per codon	105215, Identify cases that do not contain the amino acid cysteine responsible for the missense substitution.
<i>E. coli</i>	10^{-3} - 10^{-4} per codon	103454, Identify cases that do not contain the amino acid cysteine responsible for the missense substitution.
<i>E. coli</i> - Lys-tRNA	3.4×10^{-4} per codon	105069, reporter system for frequency of each type of misreading error
<i>Saccharomyces cerevisiae</i>	5×10^{-6} - 2×10^{-5} per codon	105216, Measurement of rescue rate of inactivating mutations of type III chloramphenicol acetyl transferase