The ups and downs of making a protein

Christine Vogel (New York University, USA)

Correctly adjusting the concentrations of the over 25,000 different proteins that reside in our cells is a complex task – and finely regulated via many processes. Recent technological advances allow us now to measure these concentrations and the rates of production and degradation for the different genes and provide first insights into how much protein is made per mRNA, and how the production and degradation rates change when the cell responds to a stimulus.

The two steps of making a protein

Proteins are the workhorses of our cells, fulfilling many different functions: they can be transporters, they can be enzymes that catalyse reactions, they can provide stability to the cell's structure, make more proteins, degrade unneeded and damaged molecules and regulate these processes. The human genome, the DNA blueprint for all of the different protein molecules in the cell, encodes ~25,000 major types of proteins. However, this estimation is on the low side, as variations and protein modifications increase this number several fold.

Depending on its function, each protein is required to be present at a well-defined concentration. Some signalling proteins might only be present at ten copies per human cell; others, for example ribosomes, the gigantic machineries whose role it is to make more proteins, are thought to exist in hundreds of thousands, if not millions, of copies. Over two-thirds of all cellular proteins are ribosomes, illustrating the enormous effort that a cell puts into producing enough of these molecules – the machinery that makes proteins is itself the most abundant one in the cell. The cell's effort to maintain the concentrations of these different proteins at the required levels through fine-regulation of production and degradation is called proteostasis, and it is the central goal of many regulatory processes. Each cell from any organism requires the correct number of ribosomes, enzymes, transporters and regulatory proteins at a given time.

Making a protein from a gene encompasses two steps that are summarized in the Central Dogma of Biology\(^1\) (Figure 1). The Central Dogma was first postulated in 1958 by Francis Crick, and it states that genetic information flows from DNA to mRNA and then from mRNA to protein, but not back: proteins cannot confer information about their make-up to the genome, the DNA. The first step, making the mRNA from the DNA blueprint, is called transcription; the second step, making proteins from the mRNA using the ribosomes, is called translation. Nowadays, much of the research is concerned not only with the flow of information from DNA to RNA to protein, but also with the rates of production and degradation of the respective molecules. We look at the cell from an engineering point of view: how fast does a process respond? What is the shape of the response signal? What is the relationship between transcription and translation across individual genes?

Unsurprisingly, transcription and translation are highly regulated, by proteins called 'transcription factors' and 'translation factors' respectively. These regulators help to determine which protein is produced and when – they are the key players during the response of a cell to an outside stimulus. For this important role, transcription factors have been studied for many years with respect to the number and types of target genes. Some transcription factors are quite famous as they are 'master regulators' that regulate the transcription of thousands of genes. One such master regulator is a transcription factor called p53. A mutation in p53's gene, i.e. a mistake in its DNA sequence design and subsequently the mRNA blueprint and the final protein, leads to dysregulation of its many cellular targets and cell division. A disturbance in cell division in turn leads to uncontrolled tissue growth – i.e. cancer. With over 50% of all cancers having a mutation in p53, it is one of the key causes of the disease\(^2\).
RNA Revisited

The human genome encodes approximately ~2000 different transcription factors, which all have different sets of target genes. Not all factors are as crucial as p53, but the entirety of these regulators and their combinatorial action determines which mRNAs (and therefore proteins) are produced at a given point in cellular life. Although the crucial role of transcription regulation has been known for a while, only recently have we learned that translation is regulated in a similarly complex manner: the human genome also encodes ~1000 potential translation factors and proteins that affect RNA degradation – a repertoire far bigger than has been assumed for a long time. The identity, precise function and set of target mRNAs is often unknown, and subject to ongoing investigation.

Finally, simply producing a protein is not enough for cellular proteostasis – the cell also needs to activate the protein, transport it to its destination and degrade it when damaged or unneeded. And, you guessed it, hundreds of regulators are encoded in our genome, ready to orchestrate these processes as needed.

**How much protein do you have – and how can you change this?**

Recent years have seen many technological advances that allow us to measure concentrations of the mRNA and protein molecules in cells, but also to estimate the rates of their synthesis and degradation. The data are noisy and sometimes sparse, but allow us to build first models of the inner workings of a cell and gain insights into its regulatory principles.

The first questions that were asked were very simple: how much protein is produced, how much per corresponding mRNA molecule, and is it the same relationship across all different genes? What is the importance of transcription relative to translation in making a protein? Indeed, if the cell produced the same number of protein molecules per mRNA, many researchers’ work would be in vain as there would be no translation regulation and just one simple rate of translation across the entire protein repertoire. This is not the case: although genes whose mRNA is more abundant than that of other genes also have higher protein concentrations, both transcription and translation rates vary across genes, demonstrating much fine-regulation at the per-gene level. This fine-regulation is also exemplified by several evolutionary insights that demonstrate Nature’s selection for optimal protein production. If a protein needs to be highly abundant, like ribosomes, it is usually shorter than less-abundant proteins, probably because it is easier to synthesize. Highly abundant proteins are also often more stable; the same applies to their mRNAs.

Furthermore, the nucleotide code for these highly abundant proteins has been optimized: they only use codons (i.e. amino acid codes) for which amino acids and adaptor RNAs are easily available in the cell, again ensuring fast production without encountering bottlenecks.

Next, after having examined how much mRNA and protein we obtain per gene, researchers are now asking how these concentrations change when the cell responds to a stimulus. Suppose a cell is responding to a chemical that has been placed into the growth medium, it now needs to increase the concentrations of so-called stress-response proteins that help the cell to reverse any damage and secure adaptation to the new environment (Figure 2). Sometimes, the concentrations changes are very big, tens- or hundred-fold, but sometimes they are very small. In nerve cells, for example, protein concentrations can change as little as 1.2-fold – meaning that only 20% of the concentration of a specific protein makes a difference in how the cell behaves. Detecting such small changes poses a challenge to experimental methods – but even for the cell realizing these changes can be far from trivial. For example, for a protein whose normal concentration is 20 molecules per cell, up-regulation by 20% simply requires production of four additional molecules. In contrast, for a different protein, for example from the cytoskeleton, the concentration might normally be

---

**Figure 1.** The Central Dogma of Biology. The Central Dogma of Biology describes the information flow from DNA to RNA to protein and builds the framework for the two steps to make proteins. Each step of synthesis and degradation is subject to multiple regulatory pathways. The numbers provided reflect approximations.
200 000 molecules per cell. Producing 20% more of this protein, i.e. 40 000 new molecules, is much more difficult and demands much cellular energy and time. We are now trying to understand different strategies to manage this task.

The first aspect to consider might be resources that are needed to produce a protein. On a per-molecule basis, making an mRNA is much more expensive than making a protein. The average combined synthesis and polymerization costs are more than 50% higher for nucleotides than for amino acids, the building blocks for RNA and protein respectively. Furthermore, because it takes three nucleotides to encode one amino acid, mRNAs are usually at least three times longer than their corresponding protein product. Finally, mRNAs have a much higher turnover rate than proteins, requiring more synthesis per time unit. However, all of these factors are outweighed by an order of magnitude when considering total molecule concentrations: because proteins are, in general, more abundant than their mRNAs by a factor of 1000 or more, it costs the cell much more energy to produce a gene's protein product than the corresponding mRNA despite the cost for individual steps showing the opposite.

Secondly, the cell might consider response time in its decision which process to regulate, and thirdly, the type of response and its sustainability: should a change in protein concentration be implemented very rapidly or is it sufficient to change slowly? Would one perhaps prefer a pulsed signal in which cells return to the original state once the threat is over or permanently switch to a new state (Figure 3)? If the environmental condition is life-threatening for the cell, the response might have to be very fast and very drastic – and crucial response proteins may have to be readily available at all times, since producing them from scratch takes time. Transcription is faster than translation, but the fastest, albeit highly expensive, route to quickly bring up concentrations is to produce the required protein all of the time, but then to shut off protein degradation when the molecules are needed. This route for regulation – producing a protein continuously, but degrading it when not needed – is certainly very costly, but might be worth it for the cell to ensure its survival. For a less threatening response, the cell can slowly shift to a new cellular state and adjust molecule concentrations more gradually by changing transcription and later translation. What do you choose between all of these scenarios, and how so for different types of genes?

Recent work has brought datasets that provide first insights into these questions. One study examined mammalian cells responding to an immune stimulus. The work required precise measurements of mRNA and protein concentrations in a 24-hour time-course experiment, but also complex labelling of the molecules to estimate the rates of synthesis and degradation. These measurements were then placed into a mathematical model that helped to deconvolute the different processes, to eliminate measurement noise and to calculate the final result: >80% of the cellular response could be attributed to transcription changes, only a small fraction to changes in translation. Indeed, in a switch-like fashion, mRNA concentrations reached new steady-state levels, defining a new cellular phenotype. Much more slowly and less drastically, protein concentrations followed suit.
In comparison, a different study examined mammalian cells responding to a more acute threat to a chemical which damaged much of the existing proteins and even caused a proportion of the cells to die. In this situation, the cell activated survival mechanisms and, bypassing transcription, immediately changed translation and protein degradation, therefore much more rapidly adjusting the protein repertoire. Later, this first-line response was followed by changes in mRNA levels, but, in a pulse-like fashion, these concentrations returned to normal.

Where do we go from here?

Many open questions remain. Under which circumstances and for which genes do we observe a pulse- versus switch-like response? How does that differ for transcription compared with translation? Is there feedback or coupling between these processes that leads to a more complex response? Anecdotal evidence indeed suggests such mechanisms.

Several transcription factors, for example, are regulated themselves via protein degradation. Protein modifications can affect translation. Some transcription factors 'moonlight' in their function and also affect the degradation of mRNA molecules. The list continues.

I sometimes liken the cell to a city, for example New York, and the proteins in our cells to the different cars that busily drive on Manhattan's streets (Figure 4). What we are trying to understand is what types of different cars are on the streets at a given time, what their purpose is, how many of each type exist, and what happens if there is an accident on Times Square. How do you manage a city? How do you adjust to changes in supply and demand? What are the principles of optimization? Where do we focus energy and effort? Given the enormous increase in molecular data that we gather and integrative models that arise, the coming years are bound to produce many exciting insights into the inner workings of this city called the cell.

I thank B. Klingele and M. Schroeter for valuable feedback. I also acknowledge funding by the National Science Foundation (NSF) [EAGER], the National Institutes of Health (NIH) [grant number Ro1 GM113237] and the Zegar Family Foundation Fund for Genomics Research at New York University.

References

1. Crick, F.H.C. (1958) On protein synthesis. Symp. Soc. Exp. Biol. 12, 138–163
2. Surget, S., Khoury, M.P . and Bourdon, J.C. (2013) Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther. 7, 57–68
3. Wagner, A. (2005) Energy constraints on the evolution of gene expression. Mol. Biol. Evol. 22, 1365–1374
4. Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227–232
5. Jovanovic, M., Rooney, M.S., Mertins, P. et al. (2015) Immunogenetics: dynamic profiling of the protein life cycle in response to pathogens. Science 347, 1259038
6. McManus, J., Cheng, Z., and Vogel, C. (2015) Next-generation analysis of gene expression regulation--comparing the roles of synthesis and degradation (2015). Mol. Biosyst. 11(10), 2680-9. doi: 10.1039/c5mb00310e–2689
7. Cheng, Z., Teo, G., Krueger, S. et al. (2016) Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. Mol. Syst. Biol. 12, 855