EVOLUTION

Protein Expression Under Pressure

Cellular protein concentrations are generally under stronger evolutionary pressure than mRNA concentrations.

Christine Vogel

Which is more conserved across species—the concentrations of proteins or the concentrations of the messenger RNAs (mRNAs) that encode them? When examining orthologous genes, it's protein concentrations that are more similar to each other. This observation was first made in worm and fly (*1*), and later for eight organisms ranging from bacteria to yeast, plant, and human (*2*). However, because the measurement platforms, data sets, and cell samples were heterogeneous in these studies, it has been difficult to separate possible biological trends from technical artifacts. On page 1100 of this issue, Khan *et al*. (*3*) show that the biological trend is very real. The authors show that protein concentrations from identical cell types across three primate species are under stronger evolutionary constraints than the respective mRNA expression levels.

New York University, Center for Genomics and Systems Biology, New York, NY 10003, USA. E-mail: cvogel@nyu.edu

Khan *et al*. subjected lymphoblastoid cell lines from humans, chimpanzees, and rhesus macaques (five of each) to RNA sequencing and mass spectrometry–based proteomics experiments. Protein and mRNA concentrations were quantified for ~3400 proteins across at least three individuals from each species, controlling strictly for variation across replicates, ambiguous quantification, and artifacts introduced by extremely high- or low-abundance genes. The authors found that the expression is more tightly controlled for orthologous proteins compared to corresponding mRNAs (see the figure).

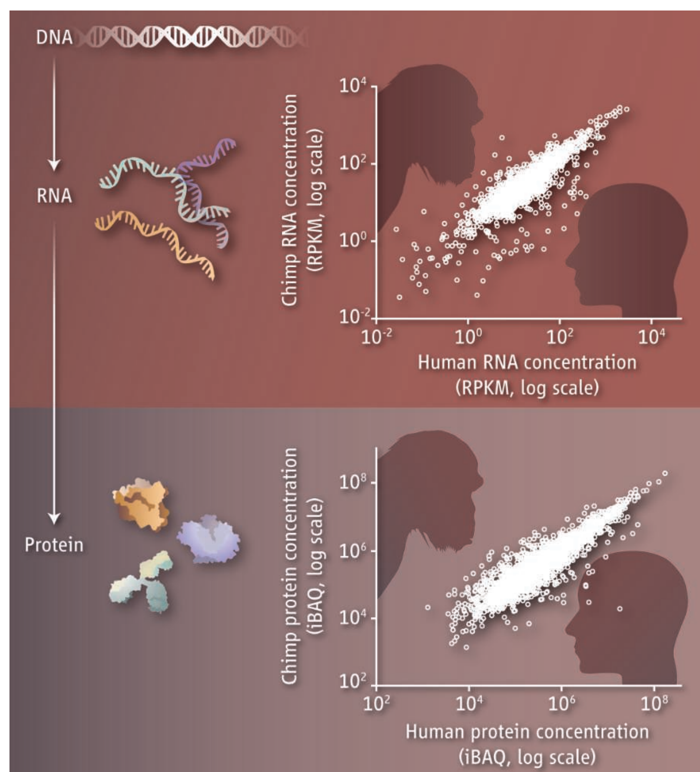
Evolutionarily, this observation seems obvious: Proteins are the cell's workhorses, and for proper cellular function, one would expect their concentrations to be firmly set at desired levels. Khan *et al*. demonstrate that protein concentrations diverge at a slower rate than mRNA concentrations, suggesting higher evolutionary constraints. These constraints may be larger for some protein functions than for others. Due to mass spectrometry's bias toward high-abundance proteins, it remains to be seen how the observation holds true for less-abundant proteins, such as transcription factors.

One hypothesis to explain the conservation of protein concentrations is inspired by work on the chaperone heat shock protein 90 (HSP90), which supports proper folding of protein substrates. HSP90 can act as an evolutionary capacitor and enable the accumulation of mutant proteins across a cell population to provide a repository of functional variants when needed (*4*). This explanation can now be extended to possible capacitor roles in gene expression regulation. On the basis of Khan *et al*.'s observation, protein concentrations may be assumed to vary less across a population of cells than the respective mRNA concentrations—i.e., protein concentrations are buffered—and the regulation of transcription and therefore mRNA abundances are allowed to evolve more freely. Indeed, this is what Khan *et al*. conclude from their data, and it is consistent with the previously

Published by AAAS

observed rapid divergence of transcription regulation (*5*). Similar to sequence variants accumulating under HSP90 function, variants of mRNA expression levels may provide a repository for regulatory adaptation if the cell population is under selection pressure. Single-cell studies that resolve the variation of mRNA and protein concentrations across individual cells will be needed to show if this interpretation holds true.

But how is the conservation of protein concentrations achieved? Several observations point to possible explanations, and the answer may lie in a combination of these. For example, mRNA concentrations appear to be often regulated in a switchlike manner, turning transcription on or off, without much attention to exact concentrations (*6*). Translation, in turn, may be much more finely regulated: MicroRNAs (miRNAs), for example, have very small individual effects on protein abundances (*7*, *8*). Similarly, other translation regulators, such as RNA-binding proteins, have also small effects on protein concentrations, as with the RNA-binding protein human antigen R (*9*). These data point to a highly combinatorial mode of action for these translation regulators that may counterbalance large variations in transcript concentrations (*10*).



Protein versus mRNA expression. Absolute concentrations of mRNA are more divergent between chimpanzee and human than absolute protein concentrations, suggesting tighter evolutionary constraints on protein abundances. Graphs are generated from table S8 in (*3*). Protein concentrations were determined by intensity-based absolute quantification (iBAQ) of mass spectrometry data; mRNA concentrations were measured by reads per kilobase of transcript per million mapped reads (RPKM) from RNA-sequencing data, as described in (*3*).

Furthermore, buffering of expression divergence at the mRNA level will likely require extensive feedback between the different steps of protein synthesis. Such coupling has been observed in several circumstances (*11*). To regulate protein concentrations and counteract variations introduced at the mRNA level, the cell would have to sense the levels of transcription (how much mRNA is present), and adjust translation and protein degradation accordingly. Conversely, once desired protein concentrations are achieved, the cell may reduce translation (and transcription) or increase protein degradation to maintain proteostasis. The existence of such feedback mechanisms could also explain the large number of possible translation regulators that have now been found: The human genome encodes at least 800 miRNA genes (*12*), and possibly ~1000 RNA-binding proteins (*13*). However, in most cases, the mechanisms of regulatory feedback between the different processes of protein synthesis are unknown, and we do not yet understand the extent to which coupling is positive or negative—i.e., whether processes are working in the same or opposite directions.

Part of the observation by Khan *et al.* can be explained by simple mathematics. In their data (but also in general), mRNA concentrations vary by three to four orders of magnitude. In comparison, protein concentrations cover four to six orders of magnitude, suggesting an amplification step at the level of translation: A single mRNA may be translated ten, hundreds, or thousands of times before being degraded. Because the resulting protein and mRNA concentrations are compared at a logarithmic scale, the amplification step alone can partly explain the observation, even when acting randomly. Indeed, there may be a very large variation in mammalian translation and degradation rates (*14*). While explaining some of the observed trends, this logic produces another intriguing question: How would the cell decide which rate to use for a particular mRNA species? Innovative techniques for measuring translation and protein degradation rates will have to provide answers (*14*, *15*).

Khan *et al.*'s observation points to several future analyses. We still lack a basic understanding of the fundamental principles of gene expression regulation, starting from simple descriptions of the absolute concentrations of mRNA and proteins in different cells, across tissues, and conditions—it is unclear how mRNA expression and protein abundance are coordinated in dynamic systems responding to a stimulus. In addition, these concentration measurements should now be extended to estimates of underlying rates of synthesis and degradation of mRNA and proteins. Despite the availability of large-scale methods for determining these rates (*14*, *15*), such experiments are still rare. One next step would involve measurements of translation in parallel to protein and mRNA concentrations. The regulation of protein synthesis is a complex process involving at least two levels (transcription and translation, plus the respective degradation), and deconvoluting these processes from observed protein and mRNA concentration measurements is all but trivial, requiring new models, molecular tools, and computational approaches—many of which are under way.

References and Notes

1. S. P. Schrimpf *et al.*, *PLOS Biol.* **7**, e48 (2009).
2. J. M. Laurent *et al.*, *Proteomics* **10**, 4209 (2010).
3. Z. Khan *et al.*, *Science* **342**, 1100 (2013); 10.1126/science.1242379.
4. C. Queitsch, T. A. Sangster, S. Lindquist, *Nature* **417**, 618 (2002).
5. P. Khaitovich *et al.*, *Science* **309**, 1850 (2005).
6. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLOS Biol.* **4**, e309 (2006).
7. D. Baek *et al.*, *Nature* **455**, 64 (2008).
8. M. Selbach *et al.*, *Nature* **455**, 58 (2008).
9. S. Lebedeva *et al.*, *Mol. Cell* **43**, 340 (2011).
10. C. Vogel, E. M. Marcotte, *Nat. Rev. Genet.* **13**, 227 (2012).
11. O. Dahan, H. Gingold, Y. Pilpel, *Trends Genet.* **27**, 316 (2011).
12. I. Bentwich *et al.*, *Nat. Genet.* **37**, 766 (2005).
13. A. Castello *et al.*, *Cell* **149**, 1393 (2012).
14. B. Schwanhäusser *et al.*, *Nature* **473**, 337 (2011).
15. A. J. Howden *et al.*, *Nat. Methods* **10**, 343 (2013).

Acknowledgments: C.V. acknowledges funding by the New York University (NYU) Whitehead Fellowship and the NYU University Research Challenge Grant.

10.1126/science.1247833

CREDIT: H. MCDONALD/*SCIENCE*

Published by AAAS