DTU Library

Principles of Systems Biology, No. 2

Klausen, Michael Schantz; Sommer, Morten Otto Alexander

[Link back to DTU Orbit](#)

Principles of Systems Biology, No. 2

It's a translation bonanza. This month's responses to our Cell Systems Call (Cell Systems *1*, 307) include four that illuminate principles governing the production of proteins from mRNA.

mRNA Sequence Features Controlling Protein Expression
G. Boël, CNRS Institut de Biologie Physico-Chimique; G.T. Montelione, Rutgers University; D.P. Aalberts, Williams College; J.F. Hunt, Columbia University
Principles
A vast number of synonymous genes can encode a protein, but uncertainty remains concerning the mechanisms by which sequence variations influence the transcription, translation, and mRNA decay processes that control protein synthesis. We developed and tested a multiparameter mathematical model describing the nucleotide sequence features that influence protein level in large-scale *E. coli* overexpression experiments (Boël et al., Nature *529*, 358). We demonstrated that base composition and mRNA folding effects are influential near the start of the coding region, whereas codon usage dominates thereafter. Our new codon-influence metric correlates with endogenous *E. coli* protein concentrations and mRNA concentrations and lifetimes, suggesting tight coupling between translation efficiency and mRNA stability.

> *"We demonstrated that base composition and mRNA folding effects are influential near the start of the coding region, whereas codon usage dominates thereafter."*

What's Next?
Studies should address whether the influence of sequence parameters changes under different physiological conditions in *E. coli* and other organisms. Our results suggest mathematical analyses of steady-state mRNA levels can be used for facile characterization of these effects. Rigorous characterization is needed of related molecular mechanisms, especially the relative influence of parameters on transcription elongation/termination versus translation versus mRNA-decay and the degree and mechanism of linkage between these processes.

Systematic Discovery of IRESs
Shira Weingarten-Gabbay and Eran Segal, Department of Computer Science and Applied Mathematics and Department of Molecular Cell Biology, Weizmann Institute of Science
Principles
Ribosome recruitment to internal ribosome entry sites (IRESs) is critical for the synthesis of numerous human and viral proteins. However, since no high-throughput methods for discovering and characterizing IRESs exist, relatively few IRESs are known and the mechanisms by which they recruit the ribosome remains elusive. We recently devised a high-throughput bi-cistronic assay and quantitatively measured the IRES activity of 55,000 designed sequences, uncovering thousands of novel IRESs within human and viral genomes and regulatory sequences that underlie IRES activity (Weingarten-Gabbay et al., Science *351*, 240).

> *"We…measured the IRES activity of 55,000 designed sequences, uncovering thousands of novel IRESs within human and viral genomes…."*

What's Next?
Our assay revealed a striking enrichment of IRESs in 3′ UTR regions of human transcripts, suggesting further studies on whether ribosomes are recruited to the 3′ UTR to enhance translation of an upstream sequence. In addition, our finding of IRESs along the coding region of many [+]ssRNA viruses led us to hypothesize that in addition to the known mechanism of polyprotein synthesis and cleavage, [+]ssRNA viruses can also translate only part of their genome. Finally, the 50-fold increase in the number of known IRESs provided by our study enables in-depth computational analyses aimed at decoding the relationship between RNA sequence and IRES activity.

Tuning Native Protein Expression
Michael Schantz Klausen and Morten O. A. Sommer, Novo Nordisk Foundation Center for Biosustainability and Technical University of Denmark
Principles
Effective tuning of protein expression levels is of fundamental importance to synthetic biology and metabolic engineering in general. While good methodologies exist for reliable tuning of protein expression, these are rarely applicable in native chromosomal loci, which often times have overlapping genes. We established an empirical approach to map the effect on protein expression of a majority of the possible Shine-Dalgarno sequences in *E.coli.* This dataset allowed us to devise a tool (EMOPEC) for efficient tuning of protein expression through subtle modifications to the Shine-Dalgarno sequence (Bonde et al, Nat. Meth. http://dx.doi.org/10.1038/nmeth.3727).

> *"This dataset allowed us to devise a tool (EMOPEC) for efficient tuning of protein expression through subtle modifications to the Shine-Dalgarno sequence."*

What's Next?
One finding from our study was that a thermodynamics based rRNA:mRNA interaction poorly explains protein expression level. However, we believe that further comprehensive empiric characterization in conjunction with bottom up models will greatly advance our ability to predict biological behavior. At a more applied level we wish to utilize EMOPEC to implement more effective engineering strategies for sampling potential cell states combinatorially for cell factory engineering. EMOPEC should enable effective and comprehensive explorations of protein expression space.

Cell Systems Call

Slowing Translation to Improve Protein Solubility

Kevin Vasquez and Lydia Contreras, University of Texas at Austin

Principles

Codon optimization has long been synonymous with maximization of protein synthesis. However, growing evidence suggests that governing nascent protein behavior can be achieved by optimizing translation. Our study highlights that strategic pausing may prevent interactions between protein domains that generate kinetically misfolded proteins. Given the demonstrated positive correlation between ribosomal occupancy and the affinity between the mRNA and the ribosomal anti-Shine-Dalgarno sequence, we have engineered selective local regions within mRNAs to undergo slower translation by introducing synonymous mutations that increase the affinity between these regions and the ribosome. We have demonstrated that these effects can increase protein solubility by up to ~7-fold.

"…strategic pausing may prevent interactions between protein domains that generate kinetically misfolded proteins…."

What's Next?

In any manufacturing process, there are trade-offs between quality and productivity. With growing interest in expression of increasingly complex proteins, better guidelines for systematically designing mRNAs are needed to achieve high yields of soluble proteins. Our study recommends adjustments to how we optimize codons in a way that considers translation kinetics of all local regions within an mRNA.

Ecological Intensification Improves Crop Yields in Small And Large Farms

Lucas A. Garibaldi, Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural, Sede Andina, Universidad Nacional de Río Negro and Consejo Nacional de Investigaciones Científicas y Técnicas

Principles

Ecological intensification, the improvement of crop yield through biodiversity management, has recently been proposed as a sustainable approach to ensure that the world's poorest people have enough food. However, the link between crop yield and biodiversity has remained unexplored in small holdings in developing countries. We created a global dataset by applying the same field protocol on 344 fields from 33 crop systems in 12 countries of Africa, Asia, and Latin America to quantify the relationship between agricultural yield (kg per ha) and pollinator abundance and diversity in landscapes dominated by small or large farms (Garibaldi et al., Science, *351*, 338–391). We found that pollinator abundance was the most important predictor of crop yield. Moreover, our data indicate that the effectiveness of ecological intensification through improved pollinator abundance was greater for small, rather than large, holdings and when pollinator biodiversity was higher.

"…pollinator abundance was the most important predictor of crop yield."

What's Next?

Recent studies show that bee declines are driven by combined stress from lack of floral resources, pesticides, and parasites. Our study demonstrates that these declines have already had an impact on global food security as pollinator abundance, pollinator diversity, and field size all interact to affect yield gaps. Ecological intensification, with smaller-holder famers as the primary focus, should be implemented to create win-win scenarios between biodiversity and crop yields.

Data-Driven Cell Conversion

Owen J.L. Rackham, Duke-NUS Medical School, Singapore; Jose Polo, Monash University; Julian Gough, Bristol University

Principles

Transdifferentiation, the process of converting from one cell type to another, has been achieved between a handful of human cell types. The limiting factor to increasing the number of successful conversions has been the trial-and-error approach taken to the identification of key transcription factors for reprogramming. By integrating gene expression and regulatory network information, we systematically predict the set of transcription factors for any conversion, and experimentally validate this with two novel human cell conversions (Rackham and Firas et al., Nat. Genet., published online January 18, 2016. http://dx.doi.org/10.1038/ng.3487). Our system "Mogrify" exemplifies how computational design with experimental data integration can break through barriers to the progress of a field.

"…we systematically predict the set of transcription factors for any [cell] conversion…."

What's Next?

Having provided Mogrify conversions between over 300 cell types as an online resource (http://mogrify.net), we anticipate that it will facilitate a rapid increase in the number of successful cell conversions. These new conversions will allow for the development of new regenerative medicine, and to this end we are continuing to add specialized medically relevant cell types to Mogrify. Also underway are further advances in the technology behind Mogrify that will not only extend the scope of its application but will also allow us to better understand the biological underpinnings of cellular identity.