CellPress

Principles of Systems Biology, No. 26

This month: metabolic networks (Picotti, Shoichet/Sali, Maranas), evolution (Solé, Sunyaev), chaperones shape protein production costs (Pal), the size-control strategy of archaea (Schmid/Garner/Amir), and monitoring mRNA degradation (Singer).

How Proteins and Metabolites Handshake

Ilaria Piazza and Paola Picotti, ETH Zurich

Principles

Metabolites are essential biomolecules that participate in a multitude of cellular processes by providing building blocks and chemical energy for metabolic reactions. The intracellular concentration of metabolites is influenced by environmental cues, and biological systems can sense their environment through the binding of metabolites to proteins. Elucidating metabolite-protein interactions is essential to understanding mechanisms of cellular adaptation and ecosystems' dynamics, but our knowledge of these interactions lags far behind that of protein-DNA or protein-protein interactions. Metabolite-protein interactions are typically discovered via hypothesis-driven experiments that rely on *in vitro* activity assays. Higher-throughput assays are limited to specific metabolite classes.

We developed a systematic approach to discover metabolite-protein interactions and metabolite binding sites. It relies on the proteome-wide mass spectrometric analysis of altered proteolytic patterns upon metabolite binding (Piazza et al., Cell *172*, 358–372). Application of our approach to *E. coli* and to metabolites covering a broad range of chemical properties resulted in a map of the protein-metabolite interactome. The majority of interactions and binding sites we identified were previously unknown. Our data enabled the discovery of novel allosteric mechanisms, alternative enzymatic reactions, and cases of metabolite-induced remodeling of protein complexes. Further, our results suggest that enzyme promiscuity is widespread.

…proteome-wide mass spectrometric analysis of altered proteolytic patterns upon metabolite binding… resulted in a map of the protein-metabolite interactome.

What's Next?

Our approach is suitable to discover various types of protein-small-molecule interactions. The next obvious application will be the unbiased identification of drug targets directly from whole-cell or -tissue extracts.

Integrative Approach for Enzyme Function Prediction

Sara Calhoun, Magdalena Korczynska, Brian K. Shoichet, and Andrej Sali, University of California, San Francisco

Principles

The functional assignments for millions of genes from genome-sequencing projects and even for proteins targeted by structural genomics are still unknown or uncertain. There is an unmet need for automated strategies for reliable functional assignment of uncharacterized, orphan enzymes.

We introduced a general integrative framework in which multiple experimental and computational observations are combined to simultaneously infer enzymes' molecular functions and their context in a metabolic pathway (Calhoun et al., eLife *7*, e31097). All pathways consistent with input information are found by satisfying "network" restraints implied by various experiments and computational analyses, such as virtual screening, cheminformatics, genomic context analysis, and ligand binding experiments. The utility of the method is demonstrated by predicting the L-gulonate degradation pathway in *Haemophilus influenza*. The predicted pathway was confirmed by X-ray crystallography, *in vitro* assays, genetic analyses, and metabolomics. More broadly, the integrative pathway-mapping approach may formally bridge the gap between structural biology and systems biology.

…[we] infer enzymes' molecular functions and their context in a metabolic pathway… by satisfying "network" restraints implied by various experiments and computational analyses, such as virtual screening, cheminformatics, genomic context analysis, and ligand binding experiments.

What's Next?

The method can be expanded by including additional types of input information, better pathway sampling methods, and consideration of branched pathways in addition to linear pathways. In addition, other kinds of networks, not only enzymatic networks, could be predicted. Finally, simultaneous combination of integrative structural biology and integrative pathway mapping could result in better spatial models of multi-subunit enzymatic machines that are organized in space.

Pathway Design Using *De Novo* Steps through Uncharted Chemical Space

Lin Wang and Costas D. Maranas, Pennsylvania State University

Principles

Newly available, versatile genetic engineering tools such as CRISPR enable the assembly and expression of complex pathways in microbial production hosts. Protein engineering, either *de novo* or through directed evolution, provides the means of re-purposing existing enzymes for novel conversions, thus substantially expanding the biochemistry parts list available for constructing pathways.

The key challenge of minimizing the number of novel steps and combining them with existing reactions within a carbon- and energy-balanced pathway remained so far elusive. In response to this unmet need, we developed novoStoic, which allows for seamlessly blending known reactions with *de novo* steps to construct balanced and thermodynamically feasible pathways that maximize performance objectives such as product yield or conversion profit margin (Kumar et al., Nat. Commun., published online January 12, 2018. https://dx.doi.org/10.1038/s41467-017-02362-x). Reaction rules used to define *de novo* steps are inferred by a prime factoring algorithm (rePrime) by analyzing known reactions cataloged in databases. A mixed-integer linear programming-based algorithm (i.e., novoStoic) is then applied to identify a ranked list of pathways with respect to performance criteria such as product yield or pathway length while keeping to a minimum the number of putative reactions. We demonstrated novoStoic for the design of short chain alcohols, drug precursors, and the biodegradation of multi-ring aromatics.

…invoke putative steps only when necessary…allows for seamlessly blending known reactions with de novo *steps.*

What's Next?

novoStoic is currently used in a variety of projects, including the identification of pathways for funneling multiple lignin-derived aromatic monomers and dimers to advanced biofuel molecules. Moving forward, novoStoic can be augmented by ranking designs based on enzyme catalytic efficiency, intermediates toxicity, and production host compatibility.

The Inevitable Complexity of Ecological Networks

Ricard Solé and Sergi Valverde, Universitat Pompeu Fabra; and Ricard Solé, Santa Fe Institute

Principles

Ecological networks are made of tangled webs of interacting species. These webs display some remarkable universal features, such as common laws of species abundance or the ways in which interactions are organized in seemingly hierarchical or "nested" matrices. What is the origin of these universal patterns?

Although an optimization scenario resulting from natural selection might seem the most reasonable explanation, we have shown that a very different explanation might provide an alternative view (Valverde et al., Nature Ecol. Evol. *2*, 94–99). We used a simple graph growth model lacking population dynamics and essentially introducing a simple set of rules of speciation (copy) and divergence (rewiring). The *in silico* model is capable of reproducing most quantitative properties found in real networks, suggesting that their architecture is a "spandrel"—that is, a byproduct of the basic (almost mathematical) rules of growth.

Our result suggests that many structural regularities associated to ecological (and perhaps other) networks have little to do, at least at some level, with natural selection

What's Next?

Our result suggests that many structural regularities associated to ecological (and perhaps other) networks have little to do, at least at some level, with natural selection. Statistical tests will be needed to validate our proposal and compare it with other types of network models. It is also important to analyze how selection operates on top of the web generated by growth rules. One particularly relevant case study to further expand our work is the human microbiome. For example, are there fundamental evolutionary constraints that also shape the ways bacteria create their webs of interactions? Is something different due to the fact that this is an ecosystem living inside a whole organism?

Quantifying Positive and Negative Selection in Cancer

Donate Weghorn and Shamil Sunyaev, Harvard Medical School

Principles

Cancer evolves through so-called driver events, many of which can be traced back to mutations. Besides these tumor-promoting mutations under positive selection, certain genes are protected from protein-altering changes through negative selection. Detection of negative selection is disproportionately more difficult and, thus, is usually ignored. Identification of both kinds of selection relies on accurate estimation of the background mutation probability, which is highly heterogeneous across the cancer genome.

We developed a probabilistic model to infer both positive and negative selection, taking into account heterogeneity of background mutation rate (Weghorn and Sunyaev, Nat. Genet. *49*, 1785–1788). For each cancer type and each gene, the model was used to compute the expected number of nonsynonymous mutations in the gene under neutral evolution (or, to be more exact, its neutral distribution). Comparing this to observed mutation data yields an estimate of selection, measured as an excess (positive) or deficit (negative) of nonsynonymous changes. Despite the elusive statistical nature of negative selection, we found a significant signal at the genome-wide level for several cancer types, as well as enrichment with genes with likely cell-essential function, as determined from a CRISPR knockout screen.

We developed a probabilistic model to infer both positive and negative selection, taking into account heterogeneity of background mutation rate.

What's Next?

Identification of negative selection opens up a new avenue to uncover cancer vulnerabilities and, thus, potential targets for therapy. Signal stemming from oncogenes that are activated by changes in non-coding regions may identify non-canonical cancer drivers. The hierarchical model framework that incorporates the genome-wide distribution of background mutation density represents a broadly applicable approach for this class of problems.

Chaperones Shape Protein Production Costs

Zoltán Farkas and Csaba Pál, Biological Research Centre, Hungarian Academy of Sciences

Principles

Proteins are involved in essential cellular functions and are therefore vital for life, but at the same time, they might consume limited cellular resources or have negative side effects. Therefore, protein production is a tightly coordinated process that is intimately linked to cellular needs.

Recently, we asked how the fitness costs of a nonfunctional fluorescent protein overproduction vary across genetic backgrounds and environmental stresses in baker's yeast (Farkas et al., eLife *7*, e29845). By integrating genome-wide genetic interaction and environmental stress screens, we identified several cellular systems—mainly involved in protein production itself—that buffer the harmful effects of protein overproduction. Perhaps most significantly, impairment of the Hsp70-associated chaperone network—either by subjecting yeast cells to proteotoxic conditions or via genetic inactivation of chaperone regulators—rendered yeast cells hypersensitive to gratuitous protein overproduction. Moreover, protein burden significantly enhanced the propensity for protein aggregation in chaperone-deficient cells. We also demonstrated that by exceeding a critical threshold of protein concentration, even a globular protein could perturb the complex formation of a specific chaperone regulator and thereby unbalance normal proteostasis.

…impairment of the Hsp70-associated chaperone network…rendered yeast cells hypersensitive to gratuitous protein overproduction.

What's Next?

We conclude that the proteostasis surveillance system can shape protein production costs. Our work is relevant to understand how evolution optimizes allocation of cellular resources in changing environments. Moreover, our findings may also be used to better design heterologous expression systems in the future.

Cell Systems Call

A Window onto Archaea Cell Size

Amy Schmid, Duke University; and Ethan Garner and Ariel Amir, Harvard University

Principles

The growth and division of single cells is highly orchestrated, ensuring the viability of cellular progeny. While decades of research on cell growth has led to insights on the way the cell cycle is regulated in bacteria and eukaryotes, little is known about cell growth and division in archaea, the third domain of life. Due to the extreme environments inhabited by many archaeal organisms, addressing this question in single cells was previously difficult. In the current study, custom microchambers were developed to visualize the real-time growth of individual *Halobacterium salinarum* cells, an archeon that lives in hyper-saline environments (Eun et al., Nat. Microbiol. *3*, 148–154). Using single-cell data, the team quantitatively analyzed the size-control strategy used by this organism, demonstrating that single cells grow exponentially, adding a constant volume from birth to division. Surprisingly, despite billions of years of evolutionary divergence, the same strategy is shared with budding yeast and various bacteria, though *H. salinarum* shows more variability across cells.

> *…custom microchambers were developed to visualize the real-time growth of individual* Halobacterium salinarum *cells…[revealing that] the same [cell-size-control] strategy is shared with budding yeast and various bacteria.*

What's Next?

The molecular mechanisms underlying cell growth and division across different domains of life remain unclear, as are the couplings between cell division and other cellular processes. For example, how is the timing of cell division and DNA replication coordinated in archaea? How is cell cycle progression affected by the conditions of the natural, often extreme, environments of archaea? These questions are now accessible with the imaging and analytical tools developed in this study.

Shining Light on the Demise of Single mRNAs

Evelina Tutucci and Robert H. Singer, Albert Einstein College of Medicine

Principles

Following single mRNAs in living cells from birth to death reveals critical information about how different steps of gene expression, from transcription to degradation, are coordinated. For the imaging of mRNAs, we previously developed the MS2 system to endogenously tag an mRNA with MS2 stem-loops bound by the MS2 coat protein (MCP) fused to fluorescent proteins. Previous stem-loops were resistant to degradation in yeast leading to accumulation of MS2 fragments and to erroneous conclusions about mRNA localization and turnover.

In our new work, we re-engineered the MS2 system for monitoring decay of single mRNAs in living *S. cerevisiae* (Tutucci et al., Nat. Methods *15*, 81–89). The new reporter (MBSV6) combines (1) MS2 stem-loops with reduced affinity for MCP, (2) an optimized MCP expression, and (3) an increased distance between each of the MS2 stem-loops. These modifications resulted in an RNA aptamer that degrades simultaneously with the tagged mRNA, allowing us to determine exactly when mRNA degrades.

> *The new reporter (MBSV6)… degrades simultaneously with the tagged mRNA, allowing us to determine exactly when mRNA degrades*

What's Next?

MBSV6 reveals mRNA localization during the stress response in yeast induced by glucose depletion. In contrast to previous reports, a coordinated recruitment of mRNA in cytoplasmic structures, such as P-bodies, was not observed. Further work will be focused on the mechanisms regulating mRNA degradation in the cytoplasm. More generally, MBSV6 will be broadly applicable to interrogate the life cycle of rapidly turning over mRNAs.