© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

SYSTEMS BIOLOGY

Tracking the protein–metabolite interactome

A method combining limited proteolysis with mass spectrometry systematically detects protein–metabolite interactions.

The genome, the transcriptome and the proteome all receive a lot of attention, but there are also multitudes of small molecules in a cell, collectively making up the chemically diverse metabolome. Metabolites wear many hats, serving as enzyme substrates and products, cofactors, allosteric regulators, and as mediators of protein-complex assembly. But despite their essential roles in biological processes, their interactions with proteins—which are often transient and low-affinity—remain largely a mystery.

"Understanding how these interactions occur on a global scale is essential to understand mechanisms of cellular adaptation and ecosystems' dynamics," says Paola Picotti of ETH Zurich in Switzerland. She and her colleagues recently designed a method to systematically discover which proteins bind to a metabolite of interest. The approach may also be useful for drug discovery, as a way to identify druggable sites in proteins and to test for off-target effects.

To identify protein–metabolite interactions in *Escherichia coli*, Picotti's team treated a whole-cell lysate with a metabolite of interest then added a low amount of the broad-specificity protease proteinase K for a short period of time, all under native conditions. This 'limited proteolysis' approach generates structure-specific protein fragments—metabolite binding can block proteinase K cleavage at locations which would otherwise be severed. Switching to denaturing conditions, the researchers then used the enzyme trypsin to completely digest the metabolite-treated sample and a reference untreated sample, generating peptides for label-free quantitative mass spectrometry analysis, and compared the resulting differential peptide spectral patterns. "The peptide-level resolution of our method enables pinpointing metabolite binding sites and effects on binding interfaces in case of protein complexes," notes Picotti. "This information is very powerful because it enables predicting the nature of the interaction, especially if the

SYNTHETIC BIOLOGY

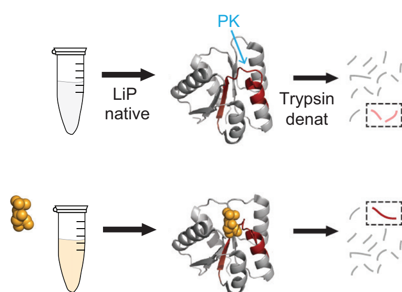
GENE FACTORIES MADE OF DROPLETS

How to make large synthetic gene libraries out of some beads, an oligo array and an emulsion.

Sriram Kosuri, a professor at the University of California, Los Angeles, is a builder of genes. Much like assembly algorithms construct gene sequences from a jumble of fragments, Kosuri physically stitches together short synthesized DNA oligomers, or oligos. As a postdoc in George Church's group at Harvard University, he multiplied the scale of gene synthesis by sequestering pools of cheap, microarray-generated oligos in individual microwells. Low background and small volumes made the approach efficient, but "it became very clear that you weren't going to get cost reduction from the oligos alone, because the enzymatic costs of doing all of those things in 384 wells would just add up," says Kosuri. His lab's new method, DropSynth, substantially drops the price of full-length gene libraries.

Oligos can be produced by the thousands on microarrays via chemical synthesis, but their length is limited to around 230 nucleotides. Kosuri's team came up with the idea of assigning unique barcodes to every oligo destined for a particular gene, so that it can be physically retrieved by a bead containing a complementary sequence. The beads are partitioned into microdroplets in an emulsion, and overlapping oligos are joined using polymerase cycling addition, which benefits from the high concentration of reagents in the tiny droplet. The protocol is straightforward: "Once you have barcoded beads, it's just a vortexed emulsion PCR," says Kosuri.

The small volumes reduce expenses to about $2 per gene, depending on the cost of oligo library production, and even less if the barcoded beads are reused (one batch of beads is enough for hundreds or thousands of DropSynth reactions). Assembly is limited to about 700 base pairs, largely by the probability that every oligo in a reaction is error free. Currently, fewer than half of individual oligos are error free, leading to only 1–5%



Limited proteolysis (LiP) coupled with mass spectrometry reveals protein–metabolite interactions. Image reprinted with permission from Piazza *et al.* (Elsevier, 2018).

structure of the proteins involved are known."

The researchers applied their method to assess 20 different *E. coli* metabolites, including amino acids, organic acids, sugar phosphates and nucleotides. They identified nearly 1,700 protein–metabolite interactions, of which more than 80% were novel, a particularly impressive result given that *E. coli* has probably the best-characterized metabolic network of any organism. The data provide clues about the biological activities of 76 proteins of completely unknown function.

A major challenge that Picotti's team faced was method validation. "Since our approach identifies a multitude of novel interactions for the metabolites we studied, an obvious question was: are they real?" says Picotti. Since it was not realistic to validate all novel interactions, they assessed method performance in several ways. They applied the approach to well-studied metabolites and a highly specific drug as a benchmark. They compared their results to a database of known enzyme-ligand interactions. They developed a scoring system to estimate the confidence of an interaction based on multiple lines of evidence. Finally, they performed biological experiments to validate several novel interactions.

Though the method appears to be unbiased with regards to metabolite chemistry, the researchers note that it will be difficult to identify metabolite interactions with low-abundance proteins. The compartmentalization found in eukaryotic cells could also make such analyses in higher-level organisms more complicated. Despite these challenges, the method has great potential to provide new insights into the essential roles of metabolites in the cell.

Allison Doerr

RESEARCH PAPERS

Piazza, I. *et al.* A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* **172**, 358–372 (2018).

assembled sequences that are perfect. Kosuri credits postdoc Calin Plesa and graduate student Angus Sidore for driving the project, and for extensive optimizations that have dropped the error rate on synthesized genes since their results were published. "We're now about 10–15% perfect in the lab," he says.

The researchers generated nearly 13,000 *DHFR* and *PPAT* gene sequences, up to 670 base pairs long, from across the bacterial tree of life. They submitted assemblies of *PPAT*, target for antibiotic development, to a complementation assay to find which sequences confer protection against selection. Error is not a big problem for the lab's applications, which only need a single error-free assembly per gene. It is easy to add duplicate reactions—what Kosuri calls "multiple shots on goal"—to increase the probability of attaining error-free sequences; and in some cases, errors come in handy. For *PPAT*, Plesa devised a way to analyze the fitness of mutated sequences, which revealed a trove of insights into protein function, including a number of gain-of-function mutations. "I was skeptical," says Kosuri. "I usually try to throw away everything that's not perfect."

The team is actively attempting to further improve scale, reduce error and increase assembly length and success rate, and is working to get the most out of large-scale gene synthesis. Everyone in the lab uses synthesis to explore "some kind of sequence-to-function relationship, whether that be transcription, or splicing, or protein function, or protein-ligand interactions, or protein-protein interactions," says Kosuri. He is thinking of ways to make bead libraries even more accessible, and has generated a website, dropsynth.org, and discussion group devoted to the method. The new scale of gene synthesis involves a way of thinking that is not common among biologists, but one that he is hopeful will spread.

Tal Nawy

RESEARCH ARTICLES

Plesa, C. *et al.* Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* **359**, 343–347 (2018).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.