# Metabolomics: from small molecules to big ideas

Monya Baker

The focus of metabolomic studies is shifting from cataloging chemical structures to finding biological stories.

In 2004, when Alan Saghatelian wanted to know what reactions a particular enzyme catalyzed *in vivo*, he had to plow through a stack of paper two feet high. Standard *in vitro* techniques to uncover enzymes' functions would have been much more straightforward, but such studies work only when the right substrates are screened and even then might represent reactions that only occur in a test tube. To get a read on enzymatic activity in actual cells, Saghatelian, then a postdoc in Benjamin Cravatt's laboratory at The Scripps Research Institute, developed an 'untargeted metabolomics' strategy to find previously unidentified substrates. Saghatelian lysed cells with and without the enzyme, ran the lysates through a mass spectrometer and compared the small molecule profiles found in each sample.

This approach, which they termed discovery metabolite profiling, yielded substrates unlikely to have been found *in vitro* and led to the discovery of a new class of mammalian lipids as substrates for an enzyme considered well characterized[1]. Identifying the substrates, though, was slow going. "I had to manually look at each mass range [of chromatograms from the spectrometer], looking for ions that had changed between the two samples," says Saghatelian, now at Harvard University. Things are different now, he says. Not only are mass spectrometers better at resolving and measuring compounds, instrument vendors all now sell software for comparing samples automatically.

The word 'metabolomics' was coined in 1998. It refers to biochemistry writ large, the study of the entire set of small molecules in a biological sample. In 2001, when a handful of biochemists organized themselves into the Metabolomics Society, their chief goal was to produce a catalogue of the metabolites that existed in any particular
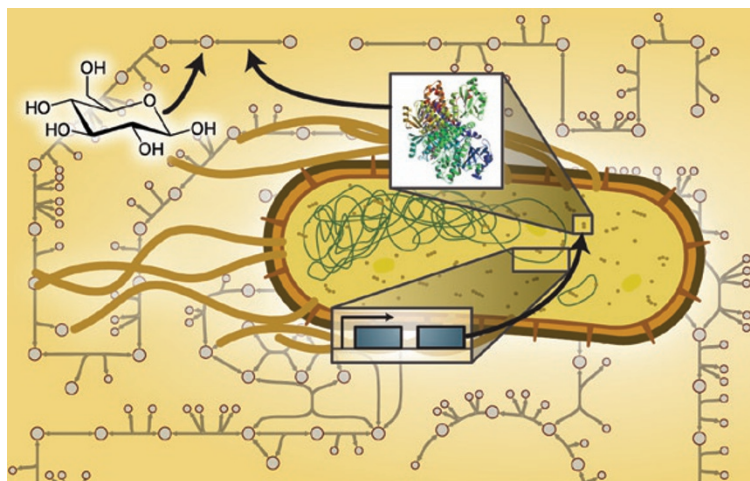
Genomic, proteomic and metabolomic data can all be integrated using genome-scale metabolic network reconstructions.

organism. The assumption was that a list of human metabolites could be as useful as a list of human genes. Such a list could help researchers probe multiple pathways simultaneously, and collections of metabolites consistently associated with biological traits could become biomarkers for identifying diseases and optimal treatments.

Researchers have only recently been able to integrate all the necessary steps to produce profiles consistently. "It's been a decade-long progression to measure more metabolites more reliably," says Joshua Rabinowitz at Princeton University. Profiling the complete metabolome remains an impossible task for now, but with dependable methods in hand, he says, the field is shifting from cataloging metabolites to asking broader biological questions about how metabolites reflect and affect cell function.

The known unknowns

Small molecules are less tractable to cataloging than are the objects of other 'omics'.

Unlike genes, transcripts and proteins, metabolites are not encoded in the genome. They are also chemically diverse, consisting of carbohydrates, amino acids, lipids, nucleotides and more.

Extraction, separation and analytic techniques that work for one class of metabolites are often useless for others. With nucleic acids and proteins one detection technique will usually suffice, but the metabolomic community must rely on a suite of detection techniques, including capillary electrophoresis–mass spectrometry, liquid chromatography–mass spectrometry, gas chromatography–mass spectrometry and nuclear magnetic resonance, all paired with various enrichment and separation technologies.

"Over the last 10 years, we have learned that there is no one platform; there is no one method," says Rima Kaddurah-Daouk at Duke University, one of the scientists who founded the Metabolomics Society. In fact, she says, it is hard even to pin down how many compounds a sample contains.

Josh Rabinowitz wants to remake the classic biochemistry diagrams to include quantitative information.

Most bacteria are believed to produce a few hundred metabolites; a plant may produce a few thousand. Animals contain perhaps a few thousand they make themselves as well as nutrients and pollutants from their environment and substances produced by microbes in the gut.

"We use three different detection methods, and we still miss stuff," says Michael Snyder of Stanford University, describing his work to detect metabolites that interact with proteins. Though researchers in Snyder's laboratory have now developed techniques to look at hydrophilic compounds, their initial studies looked at hydrophobic ones not because these were more likely to be interesting but because these were technically easier to analyze[2]. Even so, he says, developing these assays required years of optimization and quality control. Graduate student Xiyan Li often ran assays six times to make sure everything was working and highly reproducible.

Scientists in a few highly dedicated laboratories can look at all sorts of compounds, from volatile compounds to complex lipids, but that capacity requires extensive equipment and well-trained researchers. Members of Oliver Fiehn's lab at the University of California, Davis, for example, have ten different kinds of mass spectrometers and five protocols geared to various compound classes.

And detection is just part of the challenge. Many compounds can be detected
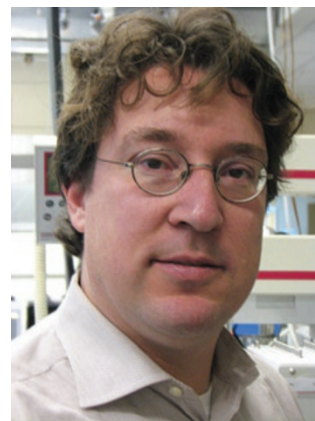
Oliver Fiehn is creating software to show how different mass spectrometry setups will portray the same molecule.
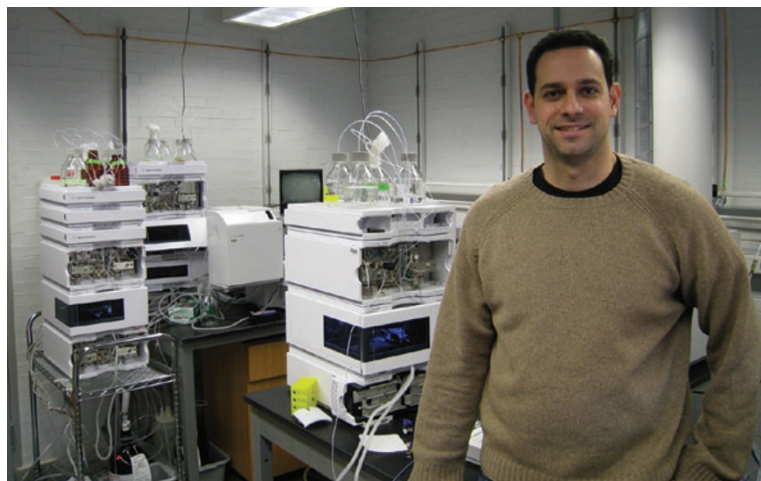
Alan Saghatelian says metabolomic profiling can uncover enzymes and pathways with therapeutic relevance.

but not identified. Researchers can make informed guesses and usually deduce compounds' molecular formulas, but typically only about 25% of observed compounds can be tentatively identified. "Compound identification is improving the fastest as it had the furthest to come," says Steven M. Fischer, senior applications specialist at Agilent Technologies. Recently, software from Gary Siuzdak and Colin Smith at The Scripps Research Institute has automated data analysis, making metabolomic analysis much more accessible. Still, confirming an unknown metabolite's identity often requires not just puzzling over spectra and running samples against known standards but also synthesizing compounds. Such unknowns can discourage researchers from looking broadly at metabolites. "If you're not a synthetic chemist, you're not likely to want to do these kinds of experiments," says Saghatelian. "If you find something that hasn't been identified yet, you don't know what to do with that. When you find a new metabolite, you can't just knock it out like a gene or a protein; you have to figure out the pathways that make it and break it down, and what it binds to."

But profiling studies can have big payoffs. The metabolites present in a cell offer a better snapshot of what is happening at a given moment than genes or even proteins, says Saghatelian. "What's nice about these metabolic profiling experiments is that you can measure some changes you can't see in genome sequencing, but it's still clear that a pathway is important."

Time to apply

Combining metabolomic profiling with other omics techniques has helped researchers identify genetic variation that predicts how a patient will respond to certain antidepressants, says Fiehn, one of several authors of a study showing how metabolomics can inform genomic studies[3]. Metabolomic profiling found that levels of the amino acid glycine differed between individuals who stopped responding to antidepressants and those who continued to benefit from them. This led researchers to analyze genetic variation in a range of metabolic genes in these individuals, focusing on genes associated with glycine synthesis and degradation. The team ultimately identified genetic variations that could help predict how an individual will respond to certain antidepressants. Without metabolomic profiling, Fiehn says, his colleagues would have been drowning in data. "It lets you know what to look for," he says.

In a different approach, Jeremy Nicholson of Imperial College London is applying analytical tools developed for genetic research to metabolomics. Metabolite-wide association studies, he says, could be equally or more useful than genome-wide association studies to classify disease and predict responses to treatment.

Metabolites themselves may have a more important biological role than many scientists have assumed. Whereas the goal of hundreds of studies has been to understand regulatory networks by hunting for protein-protein and protein-DNA interactions, few studies have been designed to look at protein-small molecule interactions. In the first high-throughput study to do so, researchers led by Stanford

Table 1 | Bioinformatic tools for metabolomics

Name and web address	Lead developer, affiliation	Description
BiGG database http://bigg.ucsd.edu/	Bernhard Palsson, University of California, San Diego	'In silico resonstructions' that integrate published data from genome-scale metabolic networks
BinBase http://sourceforge.net/projects/binbase/	Oliver Fiehn, University of California, Davis	Tool to process and analyze data from different mass spectrometry platforms
BioCyc http://biocyc.org/intro.shtml/	Peter Karp, SRI International	Databases describing biochemical machinery of many organisms, plus software to construct, query and publish pathway databases online
Chemical Translation Service http://cts.fiehnlab.ucdavis.edu/	Oliver Fiehn, University of California, Davis	Software to allow queries across metabolomic publications and databases that use different compound identifiers
Human Metabolome Database http://www.hmdb.ca/	David Wishart, University of Alberta	Database containing nearly 8,000 metabolites; each entry contains 110 fields covering clinical, biochemical and other data
MassBank http://www.massbank.jp/	Takaaki Nishioka, Keio University	Repository of high-resolution mass spectral data for metabolites
METLIN http://metlin.scripps.edu/	Gary Siuzdak, The Scripps Research Institute	A database and repository of mass spectral data on over 25,000 metabolites
XCMS/metaXCMS http://metlin.scripps.edu/xcms/ http://metlin.scripps.edu/metaxcms/	Gary Siuzdak, The Scripps Research Institute	Open-source software to identify and analyze metabolites from raw mass spectrometry data; metaXCMS allows comparisons across several sample classes
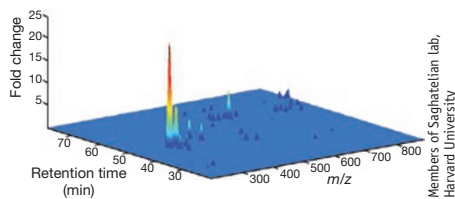
A three-dimensional plot can be used to compare ratios of various metabolites across samples.

University's Snyder isolated dozens of proteins from yeast and then carefully collected and analyzed hydrophobic metabolites associated with each protein[2]. Ergosterol, the yeast equivalent of cholesterol, bound many proteins in stoichiometric ratios that suggested a regulatory role. Though no such interactions had previously been described, 21 of 100 kinases were associated with hydrophobic small molecules, and additional work on one kinase showed the association was required for enzymatic activity. Preliminary work indicates that the regulatory roles for metabolites translate to human cells, says Snyder: "our research results suggest a highly connected network of interactions in which the small metabolites add an extra dimension of regulatory information."

In retrospect, such results should have been expected; the small molecule cyclic AMP (cAMP) has long been known to regulate many pathways. Other researchers are now looking for other small molecules that might have similar roles. Whereas Snyder's strategy was to look for metabolites that bound a defined set of proteins, Anne-Claude Gavin and colleagues at the European Molecular Biology Laboratory, Heidelberg, selected 56 lipids and intermediate metabolites and searched for the proteins that bound them. They identified over 500 protein-lipid interactions, over two-thirds of which were unknown or unexpected[4]. In follow-up studies they used live-cell imaging to validate the functional interactions of a subset of metabolites, the mysterious sphingolipids, revealing a potential new type of protein-binding domain.

Snyder calls such work interaction proteomics, which is different from classic metabolomic profiling experiments. "It's not your typical experiment to look for a biomarker," he says. "We are trying to work out pathways." In fact, metabolic profiling is increasingly moving from identifying potential biomarkers to putting them in context. In some cases, this work has even

indicated potential therapeutic interventions. In one oft-cited example, researchers at Agios Pharmaceuticals investigated mutations in genes for the enzyme isocitrate dehydrogenase that are frequently observed in glioblastomas. Metabolic profiling showed that the mutations bestowed a new catalytic function on the enzyme[5]. Notably, this mutated active site was not present in benign cells.

Metabolomics' beginnings in biochemistry offer a distinct advantage for validating protein targets, says Princeton University's Rabinowitz, a co-author on the study[5]. Decades of work to understand how products of one enzyme become substrates for the next provide crucial context for interpreting metabolomic data. And researchers who want to investigate the role of particular metabolic enzymes in disease can frequently draw on small-molecule tools, providing a fast-track for drug-discovery efforts.

Instead of focusing on metabolite amounts, many researchers prefer to look for changes in metabolic activity. Metabolite amounts can change for multiple reasons, explains Rabinowitz. A downregulated pathway might produce less of a particular metabolite; an upregulated pathway might consume more, so opposite mechanisms could produce the same effect. To distinguish between these possibilities, Rabinowitz developed a technique called kinetic fluxomics, which uses isotope-labeled precursors such as [^{13}C]glucose or [^{13}C-^{15}N] glutamine to gauge pathways' activities over time.

One such study showed that viral infection boosted rates of fatty-acid synthesis. Adding drugs that reduced fatty-acid synthesis slowed replication of cytomegalovirus as well as influenza A, the cause of most serious viral influenzas[6]. As there are many steps in lipid synthesis, says Rabinowitz, there are potentially many ways to interfere with it and so block viral replication. Researchers in other labs are also finding that unanticipated pathways are involved with disease, says Rabinowitz. "No one thought that lipid metabolism had anything to do with viruses, and there are other pathways that people thought had nothing to do with certain diseases that are actually apparently quite central to them."

Building pathways

"Metabolism, not in the old sense of discovering molecules, but in terms of doing

systems analysis, is really emerging now," says Bernard Palsson, at University of California, San Diego. A metabolic profile is really a disassembled jigsaw puzzle of cells' activity. Omics and informatic tools can help researchers put that puzzle together.

Global metabolic analysis is both easier and more complex than the research that revealed the Krebs' cycle some seven decades ago, says Rabinowitz. "We can see in the mass spectrometer all the things that had to be inferred or painstakingly measured before," he says. Now, however, researchers are also looking at metabolism in a broader way, not only what molecules are present and how they interconvert but understanding how they are controlled.

Palsson's approach involves collating genes known to catalyze biochemical reactions and building them into genome-scale metabolic *in silico* reconstructions, which are being used as models to predict the effects of drugs. At this point, though, models are generally built using gene expression and proteomics data, says Nathan Lewis, a graduate student in Palsson's lab. Metabolomic data are used for validating and extending the model rather than constructing it.

Peter Karp of Stanford Research Institute (SRI) International's Artificial Intelligence Center also starts with genomic data to construct databases describing biochemical machinery of many organisms. Experimental data provided by collaborators are computationally analyzed to identify metabolites that change in sync with gene
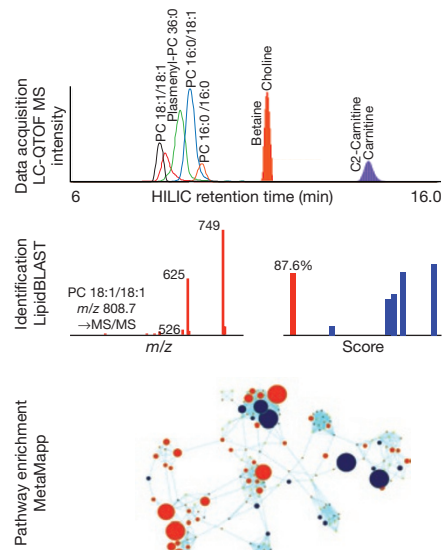
Identifying metabolites in biological samples can help researchers home in on pathways.

BOX 1   METABOLOMICS *IN SITU*

Where a metabolite is can be much more important than how much is there. "If the metabolite is in the vacuole, it is not exposed to nuclear or cell-surface receptors," says Wolf Frommer at the Carnegie Institution for Science, who has developed genetically encoded sensors that can be placed precisely in the endoplasmic reticulum, nuclear membrane and elsewhere to identify metabolites. Last year, researchers in his laboratory reported using fluorescence resonance energy transfer (FRET) and microfluidics to monitor intracellular glucose and ATP in yeast cells exposed to varying amounts of nutrients[7]. The sensors can also be used to measure metabolite uptake and release in mammalian and plant cells; moreover, FRET is much simpler to implement than many biologists seem to believe, says Frommer. "You don't even need a microscope," he explains. Whereas the technique can be used to monitor only one or two metabolites at once, Frommer anticipates that studies can be miniaturized and parallelized so that hundreds of metabolites can be monitored in time and space.

For tissue, mass spectrometry approaches can be used. Gary Siuzdak at Scripps Research Institute recently introduced nanostructure-initiator mass spectrometry, an extraction method that can be used to probe metabolites in complex tissues. Researchers in Siuzdak's lab recently used it to detect cholesterol in the cerebrum, cerebellum and brainstem of mice used to study a neurocognitive disorder that results from faulty cholesterol synthesis[8]. The implementation, he says, is trivial: "tissues can be placed directly on the nanostructure-initiator mass spectrometry chip without any preparation and then analyzed immediately." Because there is no matrix deposition as in matrix-assisted laser desorption/ionization and thus no matrix background, spectra are easier to interpret. The difficulty, says Siuzdak, is that the chips are not commercially available yet.

For metabolomics in general, the past 5 to 10 years have been characterized by the advent of more accurate machines allowing more accurate identification of the contents of sample lysates, says Arthur Castle, Program Director of Metabolomics and Informatics at the US National Institutes of Health. Emerging technologies show not just compounds' identities but also their location. "Imaging is not that big yet," he says, "but I expect it to get big."
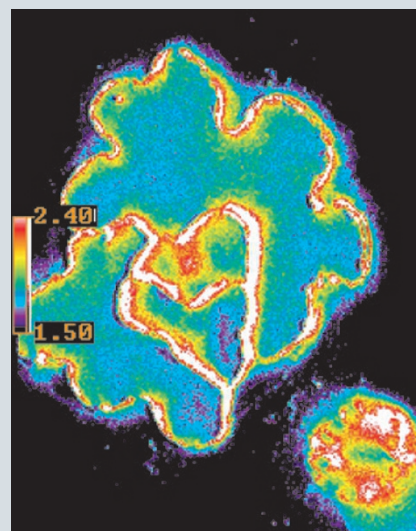
Glucose levels in the cytosol of leaf epidermal cells of *Arabidopsis thaliana*, a model plant, as visualized and quantified with a FRET sensor.

expression. This information is then used to flesh out maps of metabolic pathways. Separately, Karp uses metabolomic data to search for correlations between metabolites and drug responses.
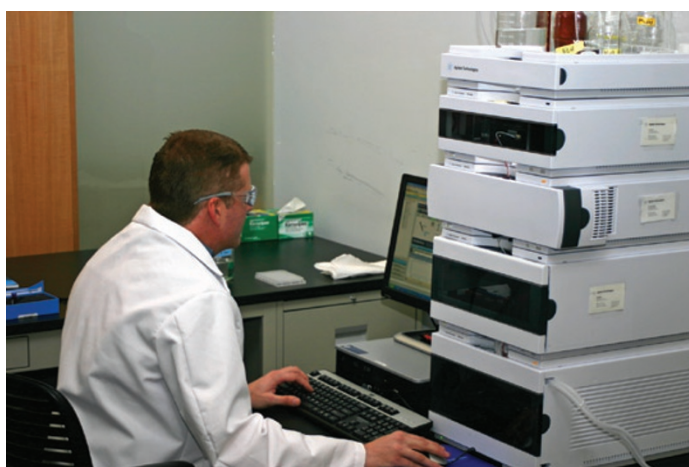
Unfortunately, many metabolites from collaborators' data sets cannot be resolved into chemical structures, Karp says. "We can't map them to our pathways right now, so we disregard them." The ability to identify the unknown metabolites would speed his and others' research, he says, and several collaborative efforts are underway to reduce the number of unknowns and to put them in context (**Table 1**). Other techniques are being developed to precisely locate metabolites in cells and tissues (**Box 1**).

Ultimately, researchers need to know not only what metabolites are where but also which are relevant. "There are a lot of peaks, but not all of them are important metabolites," says Rabinowitz. Establishing a set of molecular standards to distinguish metabolites from procedural artifacts, he says, would be somewhat equivalent to having an organism's genetic sequence; it will be useful, but he is not stopping to wait for it. "One thing we've learned from sequencing the genome is that you should try to push forward to interesting science stories as soon as possible."

1. Saghatelian, A. *et al. Biochemistry* **43**, 14332–14339 (2004).
2. Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M. & Snyder, M. *Cell* **143**, 639–650 (2010).
3. Ji, Z., *et al. Clin. Pharm. Therap.* **89**, 97–104 (2011).
4. Gallego, O. *et al. Mol. Syst. Biol.* **6**, 430 (2010).
5. Dang, L. *et al. Nature* **462**, 739–744 (2009).
6. Munger, J. *et al. Nat. Biotechnol.* **26**, 1179–1186 (2008).
7. Bermejo, C., Haerizadeh, F., Takanaga, H., Chermak, D. & Frommer, W.B. *Biochem. J.* **432**, 399–406 (2010).
8. Patti, G.J. *et al. Neuroscience* **170**, 858–864. (2010).

Monya Baker is technology editor for *Nature* and *Nature Methods* (m.baker@us.nature.com).

More accurate mass spectrometers and data-analysis software encourages metabolic profiling.