Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Microbial metabolomics: innovation, application, insight

Bree B Aldridge[1,2] and Kyu Y Rhee[3,4]

Most textbooks depict metabolism as a well understood housekeeping function of cells. However, organisms vary in their metabolic needs according to the specific niches they reside in and selective pressures encountered therein. Recent advances in analytical chemistry have begun to reveal an unexpected diversity in the composition, structure and regulation of metabolic networks. Here, we review key technological developments in the area of metabolism and their impact on our understanding of the fundamental roles of metabolism in cellular physiology.

Addresses
[1] Department of Microbiology & Molecular Biology, Tufts University School of Medicine, Boston, MA 02111, USA
[2] Department of Biomedical Engineering, Tufts University, Medford, MA 02155, USA
[3] Division of Infectious Diseases, Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA
[4] Department of Microbiology & Immunology, Weill Cornell Medical College, New York, NY 10065, USA

Corresponding author: Rhee, Kyu Y (kyr9001@med.cornell.edu)

Current Opinion in Microbiology 2014, **19**:90–96

This review comes from a themed issue on **Novel technologies in microbiology**

Edited by **Emmanuelle Charpentier** and **Luciano Marraffini**

For a complete overview see the Issue and the Editorial

Available online 11th July 2014

http://dx.doi.org/10.1016/j.mib.2014.06.009

Introduction

In biology, paradigm shifts frequently follow new technologies and their ability to replace inference with observation. Accordingly, recent advances in analytical chemistry and cellular imaging have spurred a renaissance in the long studied field of metabolism. Though once viewed as a well understood and invariant housekeeping function of all cells, metabolism has re-emerged as an unexpectedly complex and integral component of nearly all physiologic processes [1,2].

From a conceptual perspective, metabolism is the biochemical fuel of all cellular processes. Yet, organisms reside in diverse, often polyphyletic, niches where they encounter a diverse range of selective pressures. Growing evidence has established that organisms can vary in their ability to regulate both the levels and configurations of a given set of metabolic enzymes [3]. The ways in which this regulatory adaptability is to navigate the specific needs of a given cell however remains unresolved.

From an analytical perspective, studies of metabolism have been historically limited to molecules for which radiometric or indirect biochemical detection methods could be devised, or bio-informatically annotated genes and enzymes associated with their synthesis or activity used as surrogates. This is because the diversity of metabolites is atomic, rather than molecular. Moreover, metabolites function within the context of reactions whose identities, rates and directions remain incompletely defined and, where known, serve even more complex and less defined physiologic networks [4]. Thus, while genetic and biochemical approaches afforded genome-scale insights, such studies have provided only fragmented or inferential views of a given organism's metabolic network and capabilities.

Tools of the trade

Tools and methods to detect and quantify metabolites date to the earliest days of biochemistry [5]. However, only recently have technical and methodologic innovations in mass spectrometry and NMR (expertly reviewed elsewhere [1,6–10]) made it possible to measure hundreds, if not thousands, of metabolites simultaneously in the absence of biological preconception, an area recently coined metabolomics. That said, no single method/tool has proven capable of embracing the full chemical and quantitative diversity of the metabolome (operationally defined here as the collection of all small molecules (<1 kDa) of a given system). For example, while NMR-based methods are unmatched for their ability to structurally elucidate and provide absolute quantitation of metabolites, with minimal sample preparation (and in some cases in intact cells/organisms), they generally exhibit a sensitivity 100–1000-fold less than mass spectrometric methods and can be computationally limited in their ability to resolve complex mixtures. Liquid chromatography-coupled mass spectrometry (LC–MS), in contrast, is unparalleled for both its sensitivity and dynamic range over NMR and other spectroscopic methods, such as ultraviolet (UV) and infrared (IR) spectroscopy; as well as its unique capacity to resolve and identify metabolites in their native state from complex mixtures containing thousands of analytes that can vary over a 10,000-fold difference in abundance. Among mass spectrometers, triple quadrupole instruments offer excellent sensitivity for measurements of pre-specified analytes with the capacity for identification by fragmentation, while high resolution full scan detectors (such as time-of-flight or

Orbitrap instruments) enable unbiased quantitation of known and unanticipated metabolites with accompanying mass measurements often useful in determining their atomic composition and empirical formulae [1,7]. Recognized limitations of LC–MS however are its bias against volatile metabolites, need for specialized, and often multiple, chromatographic methods, and limited capacity for structural identification and absolute quantitation. Gas chromatography-coupled mass spectrometry (GC–MS) enables quantitation of volatile and uncharged metabolites that are hard to measure by LC–MS, as well as isomeric compounds, such as sugars and lipids, that are notoriously difficult to resolve by LC–MS and often beneath the limit of detection by NMR. Disadvantages of GC–MS are its requirement for extensive chemical derivitization and, thus, *a priori* knowledge of the chemical properties, if not identities, of the molecules of interest. Mindful of these many, but complementary, analytical trade-offs, cell-wide or organism-wide profiling of metabolites has become achievable using a combination of carefully selected platforms.

Analytical advances aside, the impact of these technologies has been further augmented by parallel advances in upstream sample handling and downstream data analysis. Robust sample handling methods to minimize artifactual perturbation and maximize metabolite coverage have thus been developed in a systematic, and platform-specific manner [11–15]. Key parameters for optimization include experimental reagents and apparatus, sample manipulation, metabolic quenching and recovery, and, for trace analytes, concentration; each of which was optimized on a quantitative scale and, where relevant, has been validated against biological benchmarks such as the Atkinson energy charge [16]. Methods for absolute metabolite quantitation have been similarly developed with the use of isotopic internal standards, typically added at the time of metabolome quenching to account for the same degradative and adsorptive losses as their endogenous biological counterparts [17].

Data analysis algorithms to enable standardized and nearly automated detection, indexing, quantification and statistical comparison of targeted and untargeted metabolomic datasets have similarly emerged in a platform-specific manner. Identification of novel metabolites remains a fundamental challenge. However, numerous metabolite databases and resources, including the ability to track even unidentified molecules using proxy identifiers (such as accurate mass-chromatographic retention time identifiers), have steadily begun to emerge [18–23].

Metabolomic metrics

An early, and still somewhat surprising, finding of metabolomics studies (irrespective of biological system) was the seemingly large (in some cases as high as 60%) proportion of detected molecules not included in existing databases or metabolite repositories [24,25]. Accounting for even the artifactual impact of metabolites arising from biological or analytical degradation or chemical adduction, growing evidence indicates that the number of physiologic metabolites is far greater than those accounted for by currently known biochemical pathways. Accordingly, an emerging application of metabolomics technologies has been to expand the chemical inventory of cellular phenotypes. For example, using an LC–MS-based platform which coupled chemical class-specific chromatography to high resolution mass spectrometry, Layre and colleagues reported the detection, indexing and comparative quantitative profiling of over 5000 molecular features (spanning 7 broad lipid categories, 23 generic chemical classes, 43 subclasses and 58 alkyl families) from the laboratory strain of *Mycobacterium tuberculosis* (Mtb) H37Rv to those of the clinical Mtb strain Beijing HN878 and the saprophytic *Mycobacterium smegmatis* with few or no false-positive molecules [26]. Interestingly, these studies not only resulted in the discovery of potential strain-specific biomarkers that could be used for chemotaxonomic classification, but more importantly expanded the inventory of mycobacterial lipids by nearly an order of magnitude over the previously largest database of known mycobacterial lipids.

In a technical tour-de-force, Bennett and colleagues made use of the same multiplexing capability of high resolution mass spectrometry to quantify the absolute intracellular concentrations of more than 100 metabolites in *Escherichia coli* using carefully calibrated isotopic internal standards [27••]. This work unexpectedly revealed that the measured metabolome was quantitatively dominated by a surprisingly small number of compounds, with the 10 most abundant compounds comprising 77% of the total molar concentration and the less abundant half of measured metabolites, which spanned most biological classes, comprising only 1.3%. In addition to its unusually broad quantitative view of a cellular metabolome, this information was used to infer the apparent occupancy of enzymes with defined K_m values for the quantified substrates, more than 50% of which were found to be saturating for substrate or cofactors such as ATP and NAD^+ during conditions of logarithmic growth (Figure 1).

Metabolomic matchmaking

From a more gene-centric perspective, we and others have exploited the analytical breadth of LC–MS-based platforms to expand the repertoire of cellular biochemical reactions. One such application, termed activity based metabolomic profiling (ABMP), enabled the discovery of enzymatic activities for genes of unknown function by incubating the corresponding recombinant protein with a mixture of potential co-factors and a highly concentrated preparation of the metabolome of the homologous organism. By monitoring for the matched, time and enzyme-dependent depletion and accumulation of putative substrates and products [28•], this assay allowed for the label