Cellular metabolomics of *Escherichia coli*

Joshua D Rabinowitz

*Escherichia coli* is among the simplest and best-understood free-living organisms. It has served as a valuable model for numerous biological processes, including cellular metabolism. Just as *E. coli* stood at the front of the genomic revolution, it is playing a leading role in the development of cellular metabolomics: the study of the complete metabolic contents of cells, including their dynamic concentration changes and fluxes. This review briefly describes the essentials of cellular metabolomics and its fundamental differentiation from biomarker metabolomics and lipidomics. Key technologies for metabolite quantitation from *E. coli* are described, with a focus on those involving mass spectrometry. In particular emphasis is given to the cell handling and sample preparation steps required for collecting data of high biological reliability, such as fast metabolome quenching. Future challenges, both in terms of data collection and application of the data to obtain a comprehensive understanding of metabolic dynamics, are discussed.


“What is true of *Escherichia coli* must also be true of elephants,” opined the Nobel Laureate Jacques Monod in 1954 [1]. This claim has proved to be remarkably apt in many aspects of biology, but perhaps none more than metabolism. The core structures of cellular metabolites, ranging from components of glycolysis and the citric acid cycle to intermediates in nucleotide biosynthesis, are identical across all known organisms. The organization of these metabolites into pathways of covalent reactions is also impressively similar between species. Sequencing of the full genomes of organisms ranging from *E. coli* to humans has led to the assignment to each metabolic reaction of the enzyme sequences responsible for its catalysis. This information has opened the door to cellular metabolomics: the quantitative study of the full network of cellular metabolism. Particular areas of interest include the concentrations of the full scope of metabolites, the fluxes of their associated transformations and the dynamic regulation of both.

**What differentiates cellular metabolomics?**

Metabolomics refers to the measurement of the full small-molecule contents of biological systems and the attempt to relate the resulting data to the underlying biological phenomena. Its birth and growth as a field of study has been catalyzed by the successes of genomics, as well as by breakthroughs in mass spectrometry (MS) that facilitate highly parallel measurement of small molecules.

Three fundamental divisions within metabolomics are biomarker metabolomics, lipidomics and cellular metabolomics. Biomarker metabolomics is the search for novel markers of disease or drug effects via analysis of bodily fluids, typically plasma or urine [2,3]. Lipidomics is the study of the full molecular constituents of cell membranes [4-6]. Cellular metabolomics is the study of the small molecules formed as cells convert incoming nutrients into usable energy and larger structural elements. Features shared across all branches of metabolomics are the need for highly parallel analysis of small molecules and the generation of large quantities of data.

From an analytical and data analysis perspective, an important distinguishing feature of cellular metabolomics is its focus on known compounds:

- High water-solubility: for example, components of central carbon metabolism (glycolytic,
Rabinowitz

pentose phosphate pathway and citric acid cycle intermediates) and biosynthetic metabolites (amino acids, nucleotides, vitamins and their precursors) [7]

- Inherently transient nature
- Low abundance: metabolites comprise only approximately 3% of E. coli cell dry weight [8]
- Well-established biological role, that is, known position in the overall network of metabolism

In addition, while biomarker metabolomics may be content to find consistent trends of diagnostic value, cellular metabolomics and lipidomics aim to capture the true contents of living cells. Given the fast turnover of many metabolites (especially in E. coli), generation of analysis-ready samples that accurately reflect the actual cellular metabolome is of paramount importance. Before discussing this critical issue in further detail, an introduction to E. coli and its special utility for cellular metabolomics is provided.

Why Escherichia coli?

E. coli is the prototypic enteric bacteria. Enterics are major constituents of normal gut flora. They are also important human pathogens: E. coli is a leading cause of urinary tract infections; enterohemorrhagic E. coli strains cause life-threatening bloody diarrhea; and Pseudomonas aeruginosa causes severe chronic lung infections in patients with cystic fibrosis.

Enteric bacteria are nonphotosynthetic Gram-negative rods capable of generating ATP both aerobically and anaerobically. They are generally competent to live on simple minimal media containing only salt, sugar, ammonia, phosphate and sulfate. Their pathways of central metabolism and biosynthesis are remarkably similar to those of yeast [9] and also, to a somewhat lesser extent, mammals, despite radically different intracellular compartmentation.

The metabolic pathways of a broad spectrum of organisms are well captured in several web-accessible databases, for example, the Kyoto Encyclopedia of Genes and Genomes (KEGG) [10] and MetaCyc [11], with information regarding E. coli especially reliable and comprehensive. The EcoCyc [101] database (a subset of the larger MetaCyc that focuses specifically on E. coli) is especially accessible to novice users [12,13].

In addition to being well-studied metabolically, E. coli is among the most intensively studied organisms from a genetic perspective. A diversity of strains has been fully sequenced [14-17], and microarrays are available for complete genomic sequence analysis of novel mutants (tiling arrays) [18,19] and for genome-wide mRNA expression analysis (expression arrays) [20,21]. Knockouts of every E. coli gene have been made [22] and libraries containing knockouts of every nonessential gene are available [23].

Compared with eukaryotes, a major simplifying feature of bacteria is the comparative lack of intracellular compartmentalization: the vast majority of E. coli metabolism is thought to occur in a single intracellular space, the cytosol. Owing to the small size of E. coli and the efficiency of diffusion over short distances, the E. coli cytosol can, to a first approximation, be thought of as a well-mixed vessel. Such an approximation is valuable for facilitating efforts to model metabolism.

Thus, E. coli is a useful model for cellular metabolism because it has many metabolic features in common with higher organisms, and is also especially easy to grow, genetically manipulate and model quantitatively.

Approaches to metabolite quantification

The foundation of metabolomics is the ability to measure many metabolites in parallel. For cellular metabolomics, key analytical challenges are the low abundance of most cellular metabolites and the difficulty of separating these metabolites from more abundant metabolites and other biomolecules. A fundamental distinction among analytical approaches to cellular metabolomics is between those that can measure metabolites directly within cells and those that require breaking up the cells to yield analysis-ready samples.

Methods that are capable of measuring metabolites directly within cells are conceptually appealing, especially if they have adequate time-resolution to capture directly metabolic dynamics. The two primary approaches of this sort are fluorescence and nuclear magnetic resonance (NMR) spectrometry. Fluorescent-based measurement of metabolites can be conducted using genetically encoded nanosensors based on bacterial periplasmic binding proteins [24]. This approach, while providing good temporal and spatial resolution [25,26], currently has the disadvantage of measuring only a single metabolite at once [27]. NMR has the theoretical potential to measure many metabolites simultaneously in live cells. However, NMR of live cells continues to face challenges in terms of temporal and spatial resolution, sensitivity, and ability to distinguish low-level metabolites against a background of much more abundant species.

Accordingly, cellular metabolomics continues to rely largely on methods that lyse cells to release metabolites prior to their analysis. Several recent reviews examine the full range of technologies appropriate for analysis of the resulting samples [28-30]. The most promising of these involve MS, with samples either analyzed directly, for example, using matrix-assisted laser desorption/ionization (MALDI) MS [31], or after chromatographic separation. Two major advantages of chromatographic separation are its ability to differentiate, based on retention time, isobaric compounds that cannot be readily distinguished by MS (e.g., sugar isomers), and to reduce ion suppression by abundant mixture components that would otherwise compete with less abundant analytes for entry into the mass spectrometer.

A variety of chromatography MS-based metabolomics assays have been developed over the past 5 years. Assays involving gas chromatography (GC) have been used to quantify a large number of compounds, yielding outstanding results for small organic molecules and, after derivatization, good results also for amino acids, sugars and monophosphates compounds [32,33]. GC is less well suited, however, to measuring highly polar analytes, such as multiply phosphorylated compounds, or larger metabolites, such as folates.
Assays involving liquid-phase separation have generally yielded data on a smaller subset of compounds, but often provide information about a greater fraction of the most biologically critical metabolites, such as nucleotide triphosphates and redox-active species [34-39]. Liquid chromatography (LC) assays also typically avoid the need for preanalysis sample derivatization.

Different liquid-phase separation techniques each have their own advantages. Reversed-phase chromatography provides good results for a broad spectrum of analytes [40], especially anionic compounds (with an amine-based ion pairing reagent) [41], but poor retention and separation of most hydrophilic zwitterionic compounds, such as nonderivatized amino acids [42]. Capillary electrophoresis and hydrophilic interaction chromatography are applicable to a broad spectrum of cationic and anionic analytes, but provide suboptimal resolution of closely related compounds (e.g., different triphosphate species) [34,38,43]. Ion exchange chromatography provides excellent resolution of sugar isomers, but risks ion suppression owing to the nonvolatile salts present in the buffer system [44].

A critical step in LC/MS methods is conversion of the output from the LC system into gas-phase ions. In pharmaceutical analysis, both electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) are popular approaches, with APCI often providing better qualitative reproducibility and robustness. For compounds that are naturally charged in solution, such as water-soluble cellular metabolites, ESI generally offers substantially superior sensitivity, however. For this reason, most LC/MS assays for cellular metabolomics employ ESI.

Once gas-phase ions are obtained, an appropriate MS or tandem MS (MS/MS) technique must be employed to generate the actual data. Major options include time-of-flight, ion trap and triple quadrupole MS. Time-of-flight provides high mass accuracy and the ability to identify both previously known and unknown analytes in parallel. Disadvantages are suboptimal sensitivity and dynamic range for known analytes. Ion traps operated in data-dependent MS/MS mode can provide spectra for both known and previously unknown analytes. If coupled to a high mass accuracy detector (e.g., an ion cyclotron resonance [45] or orbitrap mass spectrometer [46]), they provide high mass resolution product ion spectra, and thus are outstanding for identification of unknown metabolites.

Triple quadrupole mass spectrometers operated in multiple reactions monitoring (MRM) mode enable quantitation of known metabolites with predetermined fragmentation patterns. At any given instant, the mass spectrometer detects only a single analyte using a selected reaction monitoring (SRM) scan: the first quadrupole selects for the parent ion mass; the second quadrupole conducts collision-induced dissociation; and the third quadrupole selects for the predetermined product ion mass (FIGURE 1). Each SRM scan takes approximately 0.1 s, enabling quantitation of numerous metabolites during a single LC run. While providing no information on unknown analytes, MRM scanning (which can also be conducted on an ion-trap instrument at the expense of slightly lower sensitivity) generally offers the best quantitative performance sensitivity and reproducibility for known analytes. As optimized MRM parameters become available for increasing of metabolites, the power of MRM-based methods is steadily growing [5,38]. A potential shorthand name for these methods, given their focus on quantitation of known metabolites, is metabolome quantitation MS.

**Preparation of biologically reliable, mass spectrometry-ready samples**

Given the current reliance of cellular metabolomics on analytical methods requiring prior release of metabolites from cells, sample preparation is a critical element of cellular metabolomics methodology. Three core components of the sample preparation process are growth of the cells, quenching of their metabolism and extraction of metabolites. For cells grown in liquid culture (as is common for physiological studies of *E. coli*, as well as most other microbes), different approaches have been taken to metabolome quenching (TABLE 1). All of them either generate a sample contaminated with extracellular material or risk artifacts owing to metabolome alterations (if cells are collected prior to metabolism quenching) or metabolite losses (all known means of metabolism quenching also cause metabolite leakage from bacteria) [47].

My laboratory has recently developed an alternative approach that involves growing cells directly on filters on top of an agarose-media support (FIGURE 1) [68]. Nutrients diffuse up through the agarose and filter to the cells, which grow at a similar rate to cells in comparable liquid culture media. As the cells only sparsely cover the filter and are grown on the filter too briefly for gross colonies to form, inhomogeneities in nutrient access between cells are minimal. The filter culture methodology enables quick quenching of metabolism simply by transferring a filter from an agarose plate into cold organic solvent. This step entails minimal risk of metabolome alteration while advantageously separating cells from the bulk of their extracellular environment. The filter culture methodology also enables a quick change of the nutrient environment by transfer of a filter to a plate with a different medium composition.

Once metabolism is quenched, metabolites must be extracted. The metabolite extraction step faces three key hurdles:

- Obtaining a high yield of metabolites
- Preventing metabolite losses via decomposition
- Minimizing sample contamination

The most serious type of contamination involves macromolecules or cellular metabolites degrading to give fragments that appear to be other metabolites. For example, protein might degrade to free amino acids or ATP might degrade to form adenosine, leading to serious errors in experimental interpretation.

Extraction procedures should be evaluated based, in part, on their avoidance of such degradation. One means of identifying metabolite decomposition is to spike the extraction solvent with isotope-labeled metabolites and then determine, by MS,
whether the spiked compound was lost and also whether it produced degradation products [49]. Similar isotope-tracer approaches should also be able to reveal metabolites that form from macromolecule decomposition.

To date, there is no perfect method of metabolome extraction that balances high efficiency with minimal decomposition. One of the most common approaches historically has been perchloric acid extraction. Systematic studies in both E. coli and yeast, however, have recommended cold methanol (or methanol:water) over other tested solvents (including hot alcohols, chloroform mixtures, aqueous perchloric acid or aqueous potassium hydroxide), based on better overall metabolome yields and presumptive lower risk of metabolite degradation compared with procedures involving high temperatures or extremes of pH [50,51]. We have recently found, however, that nucleotide triphosphates can decompose into nucleosides and other related products during extraction of E. coli with cold methanol:water mixtures [49]. The decomposition reactions are catalyzed by residual cellular material from the quenched E. coli, as the nucleotides alone are stable in methanol or methanol:water for several days at under 4°C.

These findings highlight the need for improved extraction procedures. For many compounds, we find extraction with 80:20 acetonitrile:water at acidic pH and cold temperature promising for E. coli [RABINOWITZ JD, KIMBALL E, UNPUBLISHED RESULTS]. Others have recently recommended heating without solvent addition [52]. Further evaluation of these and other novel methods is warranted.

**Figure 1. Escherichia coli sample preparation and analysis by metabolome quantitation mass spectrometry.** (A) Filter-based cell culture, fast metabolism quenching and cold metabolite extraction. (B) LC triple quadrupole tandem mass spectrometry analysis of the resulting extract with the aim of quantifying known intracellular metabolites. Figure adapted in part from [48].

HPLC: High-performance liquid chromatography.
<table>
<thead>
<tr>
<th>Brief description</th>
<th>Culture type</th>
<th>Quenching step</th>
<th>Risk of metabolome alteration</th>
<th>Risk of metabolite losses</th>
<th>Extent of sample contamination</th>
<th>Difficulty of metabolome measurement</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>No harvest</td>
<td>Liquid</td>
<td>Drop cells plus media directly into quenching solvent</td>
<td>Low</td>
<td>Low</td>
<td>High (extracellular metabolites and salt mixed with cellular metabolites)</td>
<td>High (sample diluted and contaminated)</td>
<td>Reliable for nonexcreted metabolites, but analysis difficult</td>
</tr>
<tr>
<td>Quench then harvest</td>
<td>Liquid</td>
<td>Drop cells plus media directly into quenching solvent; then centrifuge cells to isolate pellet</td>
<td>Low</td>
<td>High (bacteria rapidly leak metabolites upon exposure to all known quenching solvents)</td>
<td>Low</td>
<td>Low</td>
<td>Artifacts possible owing to metabolite leakage, but promising if absence of leakage can be proven for specific solvents and/or analytes</td>
</tr>
<tr>
<td>Harvest then quench</td>
<td>Liquid</td>
<td>Centrifuge or filter cells; then add quenching solvent directly to cell pellet or filter-bound cells</td>
<td>High (metabolome may change during harvesting)</td>
<td>Moderate (cell concentration step may result in consumption of some metabolites)</td>
<td>Low</td>
<td>Low</td>
<td>Artifacts possible owing to metabolome alterations during harvesting, but data for longer lived metabolites potentially useful</td>
</tr>
<tr>
<td>Wash then quench</td>
<td>Liquid</td>
<td>Centrifuge or filter cells; wash cells; then add cold solvent to cell pellet or filter-bound cells</td>
<td>Very high (metabolome will likely change dramatically during washing, if not before)</td>
<td>Very high (washing may result in metabolite consumption or leakage)</td>
<td>Low</td>
<td>Low</td>
<td>Artifacts excessively likely</td>
</tr>
<tr>
<td>Filter culture</td>
<td>Cells grown on filter on top of agarose-media support</td>
<td>Place cell-loaded filter directly into cold solvent</td>
<td>Low</td>
<td>Moderate (filter may absorb some metabolites)</td>
<td>Moderate (some media transferred with filter)</td>
<td>Moderate (sample less concentrated than from cell pellet)</td>
<td>Reliable for broad spectrum of analytes</td>
</tr>
</tbody>
</table>
Role of isotopic standards in metabolite quantification

To improve quantitation and avoid risks of artifacts owing to ion suppression, MS-based assays commonly use isotope-labeled internal standards (IS). Such standards (differing from the analyte of interest only in their mass) can be differentiated by MS, but otherwise show identical behavior to the actual analyte, rendering the analyte-to-IS ratio a reliable measure of analyte abundance. For metabolomic studies, it is neither economically feasible nor scientifically promising (owing to long-term instability of complex metabolite mixtures) to produce a comprehensive mixture of isotope-labeled IS via organic synthesis. One alternative is to use a mixture of 10–20 purified, labeled standards, which together capture the major chemical properties of cellular metabolites. Another is to synthesize the IS biologically; for example, by uniformly feeding *E. coli* 13C-glucose (42,44,53) or 15N-ammonia (55).

Spiking uniformly labeled cell extracts with known quantities of unlabeled standards can enable absolute metabolite quantitation, based on the ratio of the labeled-to-unlabeled metabolite signal and the known concentration of the spiked, unlabeled metabolite. For reliable quantitation, the standards must be dissolved shortly before use or otherwise their stability in solution confirmed in advance. As cellular metabolites can be lost at any step in the sample processing procedure, most reliable quantitation will be achieved if the spiked metabolite is added as early as possible (e.g., directly in the quenching solution) rather than later in the process.

Flux measurement

Although the focus of the present review is measurement of metabolite concentrations, not fluxes, and given the importance of both concentrations and fluxes to obtaining a complete view of cellular metabolism, a brief mention of experimental and computational techniques for flux measurement is warranted. For a thorough review of flux measurement methods, see (54).

The best established experimental approach to flux measurement, metabolic flux profiling (MFP), involves measuring the steady-state isotope-labeling pattern of cellular metabolites or proteic amino acid after feeding cells a mixture of isotope-labeled and unlabeled glucose (or other carbon source) (55,56). MFP has proved to be a valuable tool for dissecting fluxes in central carbon metabolism in both wild-type and mutant microbes (57–60). Recently, it has also been applied to study the effect of growth rate on carbon metabolic fluxes in *E. coli* (61). A variant of MFP involves measuring the labeling kinetics, rather than the steady-state labeling pattern, of metabolites (62). This kinetic flux profiling approach has been applied to explore fluxes outside of central carbon metabolism, where steady-state labeling patterns are often uninformative (e.g., nitrogen-labeling patterns of amino acids contain little information, as most amino acids contain only a single nitrogen). Kinetic flux profiling also has the potential to capture dynamic changes in metabolic fluxes.

One of the great computational successes of metabolomics has been in steady-state flux prediction for *E. coli* using the technique of flux balance analysis (FBA). FBA applies knowledge of the stoichiometry of a metabolic reaction network to identify all possible steady-state network fluxes (63). Mathematically, the fluxes are unknowns in a system of linear algebraic equations, with each equation associated with a particular metabolite. The number of unknowns (fluxes) exceeds the number of equations (metabolites), and hence the system is underdetermined (i.e., a wide range of fluxes is possible). To select among these possible fluxes, a clearly defined objective is needed. For *E. coli*, maximal growth per carbon consumed has proved to be a useful objective function. Using it, an optimal set of fluxes to meet the objective can be calculated by linear optimization (64). Remarkably, the growth rate and nutrient consumption of *E. coli* have been shown to evolve in the laboratory to match predictions of FBA (65,66).

Examples of *Escherichia coli* metabolite concentration data & its analysis

Studies of *E. coli* with a metabolomic flavor substantially predate the genomic era. In 1971, Lowry and colleagues used enzyme assays to quantitate 16 key intracellular metabolites from *E. coli*, mapping their responses to changing carbon and nitrogen sources (67). Over the past decade, a variety of additional papers investigating the *E. coli* metabolome have been published, revealing the potential for large metabolome changes in response to changing nutrient conditions (58,68,69). These studies also provided valuable information regarding the response of central carbon metabolites to a glucose pulse (70,71) and to certain enzyme mutations (72,73).

Recently, my laboratory examined the response of *E. coli* and *Saccharomyces cerevisiae* to carbon and nitrogen starvation using the filter-culture and metabolome quantitation MS methodology highlighted in Figure 1. The quantitative dynamics of metabolite levels upon nutrient removal was measured for over 60 known metabolites (Figure 2) (48). Consistent with prior results, we found dramatic changes in metabolite levels, with greater than tenfold changes found for approximately 50% of the metabolites.

Response patterns were remarkably similar between *E. coli* and *S. cerevisiae*, as revealed visually by the display of the data after hierarchical clustering in heat map format (Figure 2) (48). This data analysis and display approach, borrowed from genomics (74), is quite useful for getting a general sense of major patterns in cellular metabolomic data. In our case, subsequent analysis of the types of metabolites in the different clusters revealed strong enrichments for specific classes of metabolites. For example, the cluster consisting of metabolites that decreased in all starvation conditions (marked 'Generically decreased' in Figure 2A) was substantially enriched for biosynthetic intermediates, which is consistent with downregulation of de novo biosynthesis occurring reliably during nutrient starvation. This generic response of decreased biosynthetic intermediates may eventually prove to be the metabolomic analogue to the stress response, which pervades much of microbial transcriptome data (75).

Matrix decomposition techniques, such as singular value decomposition (SVD) and principal component analysis (PCA), provide another approach to finding major trends in
complex metabolomic data sets [76]. In SVD, a matrix of data is broken down into characteristic vectors that attempt to concisely describe much of the relevant information. These characteristic vectors can be thought of as prototypical response patterns. The first characteristic vector is the response pattern, which, when multiplied by a series of metabolite-specific scalars, best recapitulates the full matrix. Note that the step of multiplication by metabolite-specific scalars implies that the sign of the characteristic vector (increasing vs decreasing) is unimportant. Only the pattern across the columns matters. PCA is SVD of the covariance matrix; roughly speaking, the primary difference between PCA and SVD is that PCA focuses on capturing variance between conditions rather than typical response patterns across them.

Application of SVD to the starvation data shown in Figure 2A yielded three characteristic vectors, which together account for a remarkable 83% of the information in the matrix (Figure 2B). Each vector had a clear biological interpretation: the first represented a generic starvation response (same across all starvation types), the second, a nutrient-specific response (opposite in carbon vs nitrogen starvation) and the third an organism-specific response (opposite in E. coli vs yeast). Application of PCA to a set of starvation data from the unicellular green alga Chlamydomonas yielded two major principal components, which clearly separated the five different starvation conditions tested [77].

The data analysis approaches of clustering and matrix decomposition are valuable tools, but fail to capitalize fully on existing literature knowledge regarding metabolism. For example, they do not take into account the known positions of cellular metabolites within the covalent reaction network of metabolism. One simple approach to incorporating this information is to overlay metabolomic data on top of metabolic maps. However, more advanced approaches to identifying quantitative relationships between fluxes, metabolite levels, isolated enzyme activities and gene expression data are needed. Flux directions and metabolite levels are linked by thermodynamic constraints: flux can proceed down a pathway only if the free energy of each pathway step (which depends on the products-to-reactants ratios, Q, according to $\Delta G = \Delta G^0 + RT \ln Q$) is negative. This inequality constraint may prove useful both for limiting feasible flux directions given known levels of cellular metabolites and for predicting metabolite concentration ranges given known flux directions [78].

**Expert commentary**

Computational (e.g., FBA) and experimental (e.g., MFP) methods for flux determination have already had a substantial impact on understanding of E. coli physiology. Methods for metabolome concentration measurement are only now beginning to produce data of comparably high biological value. The greatest barrier to generation of such data is producing samples that contain nontrivial levels of metabolites while accurately reflecting true intracellular metabolomic composition. If appropriate samples can be produced, a diversity of MS-based techniques can adequately analyze them. Accordingly, future work should focus on integrated protocols that take into account every laboratory step from cell culture to metabolome quenching to sample analysis. These protocols should be evaluated not only against standard analytical quality metrics (e.g., measurement reproducibility), but also for their ability to replicate key biological results. Table 2 provides a list of exemplary high reliability metabolome results for E. coli that are consistent with known flux directions and regulatory interactions, and that have been independently verified by multiple investigators.
Five-year view
Analytical methods for *E. coli* metabolomics are likely to steadily improve over the next 5 years. With continuous progress in chromatographic separation, ionization technology and mass spectrometers, effective methods for quantitating most known components of core *E. coli* metabolism (~500 compounds) will be developed. To obtain the best analytical coverage of the full water-soluble metabolome, a combination of separation approaches will probably prove most effective. As long as each individual approach is relatively easy to set up and fast to run, then analysis of each sample on a few different methods should be logistically bearable. Data from multiple instruments can then be fused computationally as needed [82].

Beyond sample analysis, an area of critical need is sample preparation. The filter culture method shown in Figure 1 represents a decent choice for present studies, but is inconvenient for measuring detailed metabolic dynamics because one filter is required per sampling time point. In addition, filter culture conditions are less controlled than those of bioreactors or chemostats. Accordingly, it would be desirable to find a way to sample continuously from liquid culture, while avoiding the problem of metabolome leakage. This may require some combination of optimization of the quenching solvent (although all cold solvents may disadvantageously lead to membrane ice crystal formation and thus cell leakage) and more rapid separation of the quenched cells from their surroundings (e.g., using a fast filtration technology). If better means of sample preparation prove to be hard to develop, methods for measuring metabolites directly within cells will grow in importance. Both fluorescence-based and NMR methods hold promise.

Beyond better data collection methods, there is a critical need for better methods of data analysis, presentation and integration. Work in this area is rapidly expanding and I anticipate many breakthroughs at the interface between experimental and computational metabolomics in the near future. One critical area will be the identification of inter-relationships between metabolic fluxes, metabolite concentrations and activities of particular enzymes. Another area will be integration of genomic, proteomic and metabolomic data to elucidate pathways involving different classes of biomolecules. High-quality dynamic experiments should prove especially valuable for determining causal relationships: distinguishing metabolites that regulate the expression of genes from genes that dictate the levels of metabolites.

Data display and integration efforts feed into another area of great promise: development of predictive models of metabolic dynamics. For glycolysis, predictive models have existed for some time and continue to improve [71]. However, for much of the rest of metabolism, they are only now beginning to be developed [83]. Given the unique suitability of bacteria to modeling via ordinary differential equations (owing to their relative lack of subcellular organism), development of predictive models will likely become a major emerging theme in *E. coli* metabolomics [84]. In building such models, a key challenge is identification of parameter values (e.g., Michaelis-Menten constant [Km], inhibitory constant [K] and maximal volume [vmax]) for enzymes in vivo. Closed-loop, experimental computational algorithms hold promise for this purpose [85].

As both measurement and data analysis capabilities improve, cellular metabolomics will begin to exert major impact on both bioengineering and pharmaceutical research. The ability to quickly assess the full metabolic consequences of overexpression or knockout of a particular gene will be valuable for guiding genetic engineering efforts aimed at optimizing production of an industrially valuable product. Combined genomic and metabolomic analysis may enable rational selection of gene targets for manipulation [86]. Examination of the full metabolic effects of antibiotic therapy (in both sensitive and resistant microbes) may point to previously unappreciated mechanisms underlying drug efficacy or resistance. Knowledge of the full metabolic effects of antibiotic agents, especially antimicrobials, may point to new combinations of drugs that maximize therapeutic efficacy while minimizing side effects and/or risk of resistance development.

**Acknowledgements**
Thanks to David Botstein, Matthew Brauer, Wenyun Lu, Elizabeth Kimball, Jie Yuan, Bryson Bennett and Shawn Campagna for their intellectual contributions. Thanks also to the Beckman

---

**Table 2. Exemplary high reliability metabolomic results for *Escherichia coli***

<table>
<thead>
<tr>
<th>Result</th>
<th>Directly measured by</th>
<th>Other evidence supporting its biological veracity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP $\rightarrow$ ADP $\rightarrow$ AMP* (absolute ATP* in range of $\sim$3–6 μmol g&lt;sub&gt;CDW&lt;/sub&gt;⁻¹)</td>
<td><a href="#">RABINOWITZ JD</a></td>
<td>Required to drive other cellular phosphorylation reactions</td>
</tr>
<tr>
<td>Glutamate –ten-times glutamine* (absolute glutamate* in range of $\sim$50–100 μmol/g&lt;sub&gt;CDW&lt;/sub&gt;)</td>
<td>[50,82,83]</td>
<td>Consistent with isotope-labeling kinetics of glutamate and glutamine pools upon switching into $^1$N-ammonia media</td>
</tr>
<tr>
<td>Glutamine falls profoundly upon nitrogen limitation</td>
<td>[48,80,81]</td>
<td>Low glutamine triggers transcription of N-response genes</td>
</tr>
<tr>
<td>FBP falls and PEP rises markedly in response to glucose limitation</td>
<td>[48,87]</td>
<td>Required for gluconeogenesis to be thermodynamically favorable</td>
</tr>
</tbody>
</table>

*Note: Future metabolomic methods should be validated, in part, based on their ability to recapitulate these results. *

*In freely growing cells (i.e., not limited by carbon or nitrogen availability).

CDW: Cell dry weight; FBP: Fructose-1,6-bisphosphate; PEP: Phosphoenolpyruvate.
Cellular metabolomics of *Escherichia coli*

Foundation, the National Institutes of Health Center for Quantitative Biology at Princeton University (P50 GM071508), the National Science Foundation Dynamic Data-Driven Application System—Small Multidisciplinary Research Projects (DDDAS-SMRP) program and the American Heart Association Scientist Development grant program for their financial support.

### Key issues

- *Escherichia coli* is a valuable model for cellular metabolomics: the comprehensive study of the small molecules involved in the generation of usable energy and macromolecule building blocks by cells.

- For quantitation of the cellular metabolome, the leading analytical technology is chromatography mass spectrometry (MS).

- The selection between gas chromatography MS and liquid chromatography (LC) MS depends on the particular metabolites of greatest interest.

- For quantitation of known metabolites by LC tandem MS (MS/MS), multiple reaction monitoring-based analysis (metabolome quantitation MS) provides the best sensitivity.

- Generating an analysis-ready sample that accurately reflects the true intracellular metabolic contents is a key challenge in cellular metabolomics.

- Fast quenching of metabolism is essential.

- Future method development efforts should focus on the integration of sample preparation and analysis, and should be judged, in part, based on their ability to recapitulate well-characterized metabolic responses.

- Data analysis and interpretation is another key challenge.

- Visualization of the data in clustered heat map format is often informative.

- More sophisticated data analysis approaches that integrate knowledge of metabolite concentrations, fluxes and structure of the metabolic network are beginning to emerge.

### References

Papers of special note have been highlighted as:

- of interest
- of considerable interest


- Matrix-assisted laser desorption/ionization mass spectrometry (MS)-based approach to *Escherichia coli* metabolomics.


- Powerful gas chromatography MS-based measurement technology that helped launch metabolomics as a field.


- Capillary electrophoresis MS-based approach to metabolomics.


Hydrophilic-interaction chromatography MS-based approach to *E. coli* metabolomics.


First large-scale use of biologically derived 13C-labeled metabolites as internal standards for metabolomics, as well as ion-exchange chromatography MS-based approach to metabolomics.


- Bacteria leak metabolites when exposed to cold quenching solvents, suggesting that quench-then-harvest approaches are risky for bacterial metabolomics.


Integration of filter culture methodology and metabolome quantitation MS to determine metabolomic response of *E. coli* and yeast to carbon and nitrogen starvation; also discusses clustering and singular value decomposition analysis of cellular metabolic data.


- Method for detecting decomposition in metabolic extracts, and its application to demonstrate instability of nucleotide triphosphates in methanol:water extracts of *E. coli*.


Cold methanol:water recommended for extraction of metabolites from *E. coli*. 196


79 Use of known flux directions to constrain feasible metabolite concentration ranges.


82 Exceptionally rigorous analysis of intracellular glutamate and glutamine levels.


**Website**

EcoCyc homepage
www.ecocyc.com

**Affiliation**

Joshua D Rabinowitz, MD, PhD
Princeton University, Department of Chemistry & Lewis-Sigler Institute for Integrative Genomics, 241 Carl Icahn Laboratory,
Princeton, NJ 08544, USA
Tel.: +1 609 258 8985
Fax: +1 609 258 3565
joshr@princeton.edu