Understanding Microbial Metabolism

Diana M. Downs

Department of Bacteriology, College of Agriculture and Life Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706, email: downs@bact.wisc.edu

Key Words
genetics, genomics, biochemistry, complex systems

Abstract

Metabolism encompasses the biochemical basis of life and as such spans all biological disciplines. Many decades of basic research, primarily in microbes, have resulted in extensive characterization of metabolic components and regulatory paradigms. With this basic knowledge in hand and the technologies currently available, it has become feasible to move toward an understanding of microbial metabolism as a system rather than as a collection of component parts. Insight into the system will be generated by continued efforts to rigorously define metabolic components combined with renewed efforts to discover components and connections using in vivo–driven approaches. On the tail of a detailed understanding of components and connections that comprise metabolism will come the ability to generate a comprehensive mathematical model that describes the system. While microbes provide the logical organism for this work, the value of such a model would span biological disciplines. Described herein are approaches that can provide insight into metabolism and caveats of their use. The goal of this review is to emphasize that in silico, in vitro, and in vivo approaches must be used in combination to achieve a full understanding of microbial metabolism.
INTRODUCTION

This article is not written as a “review” of metabolism, but as a presentation of methods, approaches, and thought processes that can be used productively to understand microbial metabolism. This manuscript has been written from the perspective that metabolism is a complex system and as such is more than simply the sum of its parts. By this criterion, all biochemical processes both defined and unknown, in addition to the regulatory, metabolic, and kinetic interactions between them, comprise metabolism. Most bacterial physiologists and biochemical geneticists would agree that what they possess is not a textbook-style catalog of metabolic facts, but rather an ability to integrate diverse experimental results with past observations and published data to generate insight into the metabolic network. In the past, this ability was facilitated by accumulated observations and anecdotal tales traded among microbial physiologists. The decline in researchers who focus on metabolism has unfortunately decreased this resource.
Understanding Microbial Metabolism Requires Bench Science

Like any complex system, metabolism is far more than the sum of its parts and thus understanding metabolism will not be achieved simply by defining and cataloging metabolic processes (25). Recent technological advances have facilitated efforts to consider metabolism as a system. Mathematical models and metabolic maps generated from various in silico analyses allow visualization of known metabolic connections and attempt to predict behaviors on the basis of these models (13, 39, 48, 118, 142). It is important, however, to recognize that the potential of these models to be descriptive is proportional to the knowledge of the components and connections being modeled. To significantly advance the theoretical models requires additional data about protein function and metabolic interactions provided by bench science. Thus the role of basic science is to rigorously define metabolic components and connections to allow comprehensive model building with theoretical approaches. Solid knowledge of the system parameters is needed to minimize the assumptions required to build a useful model (138). The necessary data are likely to be obtained by implementing classical and modern technical approaches to understand metabolic processes (e.g., a biochemical pathway) in isolation and extending in vivo analyses to identify and characterize connections between processes.

MODEL ORGANISM(S) FOR METABOLIC STUDIES

Historically, model systems and model organisms have been used to generate knowledge of general biological paradigms. This representative approach compensates for the fact that the number of organisms worthy of study far exceeds available resources. An ideal organism for studying microbial metabolism has two essential characteristics: (a) sequenced genome and (b) rigorous in vivo genetic capabilities.

For decades *Escherichia coli* and its close relative *Salmonella typhimurium* have been used as the model bacteria for metabolic studies (95). Metabolic paradigms differ enough between gram-positive (i.e., *Bacillus subtilis*) and gram-negative (i.e., *E. coli*, *S. typhimurium*) bacteria that a model for each is desirable (135). In addition to the bacterial systems, it is critical to have a model system for eukaryotic and archaeal metabolisms. *Saccharomyces cerevisiae* is a eukaryote with a long history of genetic analyses (17), and genetic systems in archaeal species are quickly advancing in their ability to support in vivo analyses (122). With these model organisms an understanding of four systems of microbial metabolism can be obtained, serving as a solid foundation for work in other organisms.

Core Metabolism

The need for a minimal number of model systems to address metabolism is based on the assumption that metabolism has evolved as a system at the cellular level. As such there is virtue in studying fewer systems in more detail to define network integration. This strategy is defensible particularly with the highly conserved central metabolism that includes components of the central dogma, pathways for carbon catabolism, biosynthesis of amino acids, nucleotides, vitamins, and cofactors. Focusing on a few model organisms for these studies allows researchers to continually develop techniques to advance in vivo analyses focused on metabolism as a system. Admittedly, at least in the initial stages, this scenario supports the understanding of a system at the expense of organismal diversity.

Heterologous Systems and Unique Metabolic Processes

Characterizing metabolic processes unique to a subset of organisms (e.g., photosynthesis, iron oxidation, nitrogen fixation, methylotrophy) provides a different challenge. Often the relevant organisms do not have in vivo...
technologies that allow integrated systems analyses. Components of these metabolic processes can often be studied in heterologous systems. Not surprisingly, however, various processes and pathways are integrated into the metabolism of the native host in a way that cannot be mimicked in heterologous hosts (53, 92, 109, 155). Although limited in providing insight into a native metabolism, heterologous systems can identify differences between metabolic networks and support efforts to identify changes in the network required to allow incorporation of a heterologous component.

STRENGTHS AND LIMITATIONS OF TOOLS TO PROBE METABOLISM

Even in well-studied model organisms our knowledge of metabolism is far from complete. Understanding metabolism requires defining the functional contribution of components and their connections in vivo, a goal that requires broad interdisciplinary efforts. The critical premise in these studies is that each component/connection (i.e., gene product, metabolite) contributes to the function of the whole system; therefore monitoring the complete cell is the best way to investigate function. Technical advances of the past decades have increased the feasibility of using a whole-cell (i.e., systems) mindset to approach metabolism. These advances include the ability to (a) monitor various parameters on a global scale via various "omic" technologies, (b) generate and identify diverse mutations, and (c) overexpress and purify components. Data can be generated rapidly with these technologies, reducing the intellectual downtime spent overcoming technical barriers to obtaining needed data. Perhaps the most exciting result of this progress has been the increased feasibility of probing the gray areas of metabolism, i.e., those areas dominated by subtle and indirect effects. In fact, an increased understanding of these areas is critical to expanding our knowledge of metabolism beyond a simple description of the parts.

A concern with emerging technologies is the perception that they eliminate the need for a solid knowledge of basic principles. A single technical approach will not provide understanding of metabolism at a sophisticated level; rather the strengths and limitations of each approach must be recognized. Any approach to metabolism should be based on a strong background in the basic principles outlined in the central dogma.

Genomic/Bioinformatics Approaches

In the past several years technologies to monitor global changes of various parameters have increased at an extraordinary rate. Various specific techniques fall under the rubric of omics and bioinformatics approaches. At a minimum, these techniques result in the ability to monitor the global "phenotype" of transcripts (47), proteins (113, 114), and metabolites (19, 101, 136). Advancing bioinformatics capabilities provides the means to cluster these data in informative ways and mine sequence data to identify probable homologs/orthologs/paralogs, promoters, regulatory binding sites, signaling networks, and enzymatic motifs (7, 67, 87, 140). Global approaches are diverse, but in general they have similar strengths and weaknesses with respect to understanding metabolism. The major strength of high-throughput technologies is the wealth of data obtained in a single experiment, often providing a snapshot of cellular activity. These techniques benefit from computational tools designed to process and present data in a way that is manageable, allowing an otherwise overwhelming amount of data to be viewed as interpretable patterns (30, 31, 64). The patterns generated (e.g., the definition of a regulon, a pattern of metabolic flux, or an evolutionary tree) provide a level of detail that would otherwise require a large number of traditional
experiments. Data from omic studies are valuable in generating testable hypotheses about functional roles and connections in metabolism.

What genomics and bioinformatics technologies do not provide on their own is fundamentally new functional information, or definitive mechanistic data. In general these technologies are broad sweep studies in which relevance is often defined by a technical parameter. For instance, microarray studies monitoring expression of genes by mutation or environmental perturbation require a change in expression greater than twofold to be considered relevant. This cutoff reflects a technical rather than a biological definition of relevance because changes in expression and/or activity of less than twofold can have significant consequences on metabolism (18, 35, 86, 125).

Biochemical Approaches

Rigorous biochemical analyses of cellular processes will remain a critical aspect of studies on metabolism for the foreseeable future. Biochemical analysis of enzymes and other processes in a defined in vitro system has been considered the hallmark of understanding a molecular process. For instance, a biosynthetic (or catabolic) pathway is considered defined if it can be reconstituted in vitro with purified components (21, 69). A similar attitude holds true for defining the function of individual enzymes (70, 127, 139).

The strength of a biochemical approach is that it tests directly the function of a component and thus eliminates complications caused by indirect effects present in the natural system. However, a biochemical approach has limitations that must be acknowledged in our effort to consider metabolism a system as opposed to a collection of component parts. In vitro biochemical analyses identify what could happen in the cell. In vitro conditions are optimized to detect the activity of interest; thus implemented conditions are unlikely to reflect the cellular environment with respect to salt concentration, substrate concentration, pH, cofactor levels, and redox state (141). Without complementary in vivo data, biochemical results do not define the cellular role of a protein, nor do they detect side reactions, regulatory controls, or accessory components that could be relevant in vivo. Although critical, biochemical results simply provide a piece of data that must be integrated with results from other approaches to understand metabolism.

Classical and Molecular Genetic Approaches

In efforts to understand metabolism, a biochemical genetic approach driven by phenotypic analysis allows the investigator to “listen” to the cell without being biased by anticipated results or a previous dogma. In vivo phenotypic changes are a direct link to the metabolism of the cell, since a phenotype is a measure of the system’s function. When designed and rigorously monitored, phenotypes are sensitive enough that small metabolic changes can mean the difference between growth and no growth. The general properties of a genetic approach are common to complex system analysis (38); deductive logic is used to assign function to component parts and to understand the integration of these parts. In this scenario, a component part is removed or altered (i.e., mutated), and any behavioral change of the system as a whole (i.e., phenotype) is monitored. The resulting behavioral change is analyzed in the context of additional information to hypothesize a function for the removed piece. Genetically, there are two primary approaches to probe metabolism: (a) make targeted mutations and monitor the resulting phenotype, and (b) screen for a desired phenotype and define the mutations causing it.

Critical to the latter approach is the use of unbiased screens, which can lead to the identification/understanding of proteins and
processes independent of precedent. Because of its unbiased nature, current jargon might define this approach as "discovery science." However, the need to identify (i.e., discover) fundamentally new paradigms is emphasized by the realization that our predictions (or those of a computer) of function or metabolic potential are based on precedent. It is unlikely that precedents for all paradigms involved in metabolism have been described.

Phenotypes caused by the removal of gene products have been characterized extensively by decades of genetic analyses and more recent global efforts and are available at the E. coli online resource (39a, 56), the B. subtilis Genome Database (8a), and the Saccharomyces genome database (122a). Collective results from these studies have contributed to our present understanding of the basic framework of central biochemical pathways in the cell. To gain insight beyond this basic framework requires exploitation of more sophisticated genetic analyses. A critical component of these efforts must be the ability to generate and characterize mutations that alter, not remove, components of metabolism. In this task, molecular technologies complement, rather than replace, classical genetic approaches and thought processes. Mapping a mutation or purifying a mutant protein is no longer an insurmountable technical task. While presented here as a simplistic notion, a genetic approach (opposed to genetic techniques) requires rigorous, creative thought and integration of known facts to generate testable hypotheses.

Like other approaches, when used in isolation, genetic analyses have significant weaknesses. Ironically, the strengths of a genetic approach (i.e., breadth, unbiased results) are also its weaknesses. While genetics has the potential to uncover new functions and subtle interactions, it does not allow conclusions about mechanism nor does it easily differentiate between direct and indirect effects. Similar to the global genomic approaches, genetic analysis breaks new ground and raises questions. Pursuit of these questions with appropriately rigorous experimental approaches is a critical step in understanding metabolism.

Caveats in Interpreting Data from Metabolic Studies

Metabolic studies can generate results that are difficult to interpret if one adheres to the simplified view of cellular processes that pervades the textbooks and common presentation of biochemical pathways. Several simplifying tenets must be re-evaluated to more realistically interpret metabolic results.

Wild-Type Strains Are Not Metabolically Stable

Experimental biology depends on comparison to a standard, typically called the wild type. The assumption is that the wild type (strain, protein) provides a stable background with which mutant derivatives can be compared. With an isolated component, this assumption is generally valid. However, when experiments involve analysis of metabolism, this assumption is flawed. Spontaneous mutation rates are such that two cultures of E. coli originating from the same single colony will be genetically different after outgrowth (85, 89). The integrated nature of metabolism makes it probable that some number of these spontaneous mutations affects the system in detectable ways. This realization has implications not only for experiments in the same lab day to day, but for the comparison and reproducibility of results in different labs theoretically using the same strain background. Many examples of the unselected divergence of strains affecting experimental outcomes have been reported (8, 54, 63, 88, 112).

Given the inevitability of these genotypic variants, a critical question is whether a clonal population is genetically homogenous enough to allow interpretation of results. The validity of interpretations from metabolic studies (e.g., growth analyses) can be increased by (a) the generation and use of isogenic strains
and (b) the application of statistical methods. The former minimizes the time since the relevant strains have diverged, thereby decreasing the probability that a mutation affecting the relevant phenotype will arise. The latter point recognizes that the number of potential variables demands statistical analyses to validate conclusions. Increased use of statistical methods requires that results of growth analyses are presented in a quantitative rather than a qualitative way (i.e., growth rates/final densities versus plus and minus signs). In this feature microbial metabolism can be equated to biological disciplines (i.e., ecology, population genetics, epidemiology) in which it is not possible to control all parameters, and therefore statistical analyses are routinely reported.

The potential for genetic variability puts the onus on the investigator to control critical aspects of the experiment and utilize statistical methods to support conclusions. Without such rigor, it is unlikely that in vivo genetic analyses will be fully appreciated by the scientific community. The presence of uncontrolled mutational changes suggests that results of metabolic experiments define a potential rather than an absolute function of the system. From this perspective, the plasticity of the network and its ability to adjust to perturbations (known and unknown) are reflected in the conclusions (138). Whereas it will soon be feasible to identify the nucleotide changes that accumulate on a genomic scale (68), attributing a functional change of the network (i.e., growth properties) to each mutation is not yet possible.

**Enzymes Can Be Promiscuous**

Enzymes have been described as selective and efficient machines (80). From a biochemical perspective, efficiency may be proportional to selectivity, but in the context of a metabolic network, relaxed selectivity could result in low-level formation of a product that has a critical role in the system. The many enzymes that catalyze more than one reaction (81, 99, 110, 143, 156) are certain to be an underestimate of this potential. Few, if any, studies directly address the possibility of multiple activities for enzymes. The anticipated prevalence of promiscuous enzymes in metabolism has implications for interpreting genetic experiments. Minor activities can be amplified by multi-copy expression, regulation, or mutations affecting the enzyme directly. Thus alternative or multiple activities of a gene product should be considered when deciphering the mechanism by which a mutation causes a phenotype.

**Small-Flux Changes Can Be Amplified**

The integrated nature of complex systems allows perturbations in one area to radiate and possibly amplify their effect in other areas distant from the site of initial impact (66a). Small changes in gene expression, flux distribution, or enzyme activity exert large-magnitude changes on the growth phenotype (i.e., metabolism) of the organism (35, 40, 58, 86, 125). These examples emphasize the need to probe a phenotype without a bias for the magnitude of change that caused the effect. The current understanding of metabolism is not sufficient to preemptively determine a cut-off for relevance.

**Metabolites Can Be Required for Nutrition and/or Regulation**

Regulation is a critical component of metabolism. While the prevalence of transcriptional regulation and the significance of the resulting regulons are widely appreciated, allosteric and metabolic regulation also permeate the system. Cellular metabolites (i.e., intermediates or byproducts of a biochemical pathway) can affect metabolism in positive or negative ways (35). When a cell requires a metabolite for growth, the compound is commonly assumed to satisfy a nutritional need. However, exogenous metabolites can also allow growth by exerting a regulatory effect. For instance aminimidazole...
carboxamide ribotide (AICAR) (6, 16, 133) and α-ketobutyrate (28, 29, 82, 83) produced endogenously by the histidine and isoleucine pathways, respectively, can have a deleterious effect on growth. Supplementation with the relevant amino acids reverses the effect by allosterically inhibiting the respective pathway (6, 83). When exogenous supplementation generates a growth phenotype, it is valuable to consider both nutritional and regulatory mechanisms to explain it. Additional experiments can distinguish between the two possibilities by using mutations that eliminate allosteric interactions (i.e., feedback-resistant alleles).

**A SUCCESSFUL MODEL SYSTEM THAT ILLUSTRATES APPROACHES TO UNDERSTANDING METABOLIC COMPLEXITY**

As a result of decades of biochemical and genetic studies, direct metabolic connections such as branched pathways and transcriptional regulons have been extensively characterized. Fewer indirect connections, such as those defined by common metabolic pools, cofactor requirements, or flux distributions, have been carefully described. In general, basic metabolic research programs have continued to focus on specific components in metabolism. Many of these studies have programmatic or self-imposed boundaries that limit the scope of experiments. In contrast, the work in my laboratory has sought to understand metabolic interactions without constraining the direction these studies could take. In the past several years this approach has generated and refined a model network that radiates from the thiamine biosynthetic pathway in *Salmonella enterica*. Serving as a node in the metabolic network, thiamine biosynthesis has proved to be advantageous for dissecting metabolic connections and defining the general role of uncharacterized proteins in the context of this network.

**Basic Biosynthetic Pathway for Thiamine**

Thiamine pyrophosphate (TPP), the biologically active coenzymic form of the vitamin, is generated by the independent synthesis and ultimate condensation of a thiazole and pyrimidine moiety (14). While the general pattern of synthesis is conserved, differences exist between the specific pathways contained in bacteria and fungi (66, 157), and between bacterial species (94). The current view of the biosynthetic pathway in *S. enterica* is shown in **Figure 1a**. A diagram depicting significant features of this model system is shown in **Figure 1b**. First, two pathways (B, C) converge to form products that are condensed to generate the final product (9, 146). Salvage enzymes (E, F) provide additional inputs into each of these pathways (107, 117). Further, one of the converging pathways (B) is a branch off the well-characterized purine biosynthetic pathway (A) (45, 46, 96–98) that is inhibited by a byproduct of the histidine biosynthetic pathway (D) (6) that feeds directly into purine biosynthesis (65, 154). This diagram illustrates several issues that must be solved by the cell to integrate gene regulation and flux distribution of what might be considered three relatively simple biosynthetic pathways. Regulation of each pathway has been well characterized (75, 76, 120, 121, 147, 148), but questions about the integration of these systems, such as how thiamine synthesis is ensured under conditions that repress the purine pathway, are being addressed (6, 27, 108).

**Conditional Growth Facilitates Genetic Approach**

Thiamine is an essential nutrient, and therefore cell growth depends on the ability to either synthesize the vitamin or obtain it exogenously. The result relevant for metabolic studies is that if the bacteria are growing in the absence of exogenously provided thiamine, synthesis must be occurring endogenously. This feature allows growth to be a sensitive
assay for flux through the biosynthetic pathway to thiamine.

Contrary to expectations for a branched pathway (Figure 1), the first step, phosphoribosylpyrophosphate amidotransferase (PurF, EC 2.4.2.24), was dispensable for thiamine but not purine biosynthesis (37) under some conditions. Biochemical labeling studies, and the requirement for the remaining purine genes, showed that this thiamine-independent growth reflected phosphoribosylamine (PRA) generation in the absence of PurF (43, 108). Taken together these results led to the conclusion that purF mutants were conditional thiamine auxotrophs, and predicted that additional mechanisms to generate PRA must exist in the cell. The conditional auxotrophy of purF mutants is the feature that has been exploited genetically to identify mutations that affect thiamine synthesis directly and indirectly.

Consideration of metabolism as a system allows the conclusion that any process found (by mutation) to affect thiamine synthesis is integrated with this pathway and as such defines a metabolic connection. Figure 2 summarizes the conditions used and the loci connected to thiamine synthesis on the basis of this assumption. Characterization of these loci has not only defined connections between metabolic processes but led to the identification and characterization of open reading frames (ORFs) of unknown function.

Low Flux Generates Instability

A purF mutant was found to reflect the general situation of low flux through the purine/thiamine pathway owing to the fact that the alternative mechanism(s) of PRA formation was less efficient than the PurF enzyme (108) (Figure 3, scenarios 1 and 2). The ability to reduce flux proved to be critical because it decreases the natural robustness of the system and thus increases the sensitivity of the network to small changes. With low flux the demand for purines cannot be met (thus the standard addition of adenine to the medium), and the quantity of thiamine synthesis is determined by the efficiency of converting the limited flux available to thiamine. Significantly, function of the system (i.e., the ability to synthesize thiamine) is easily monitored by growth analysis.

Analysis of the phenotypic consequences of altering flux distribution at the purine/thiamine branch point provided valuable insight into the balance of the system. Figure 3 summarizes the several flux scenarios created by environmental and genetic manipulation, and indicates the resulting effect on thiamine synthesis. Studies showed that in a purF purR double mutant aminoimidazole ribotide (AIR) was diverted from 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) formation toward purine biosynthesis because expression of the pur genes, specifically purE (108), was derepressed (121) (Figure 3, scenario 2 versus 3). Although the resulting flux shift did not overcome the purine requirement, it eliminated the ability of the cell to generate thiamine. This result provided a tool to differentiate conditions (media or genetic) in which thiamine was required owing to insufficient flux from those conditions of no flux (i.e., no PRA) (23). For instance, in some conditions that are nonpermissive for thiamine synthesis in a purF mutant (i.e., glucose medium), blocking an outlet of AIR with a lesion in purE restored thiamine-independent growth (108). This concept is illustrated in Figure 3, scenario 4 versus 5. These and other metabolic studies indicate that system robustness can be generated by excess substrate (or flux) masking inefficiencies in a pathway.

Metabolic Connections Identified by Lesions Disrupting PurF-Independent Thiamine Synthesis

Independent of flux distribution, the efficiency of converting limited PRA to thiamine can be affected in a number of ways. Several
mutations that prevent thiamine synthesis in a purF genetic background have been identified over the years (10, 11, 23, 36, 41, 106). In general the lesions indirectly disrupt one or more of three components of the thiamine pathway such as (a) the ability to generate PRA, (b) the efficiency of AIR-to-HMP conversion and (c) the ability to generate thiazole.

**Loci involved in PurF-independent PRA formation.** Loci required directly for PurF-independent PRA formation were identified by mutations that were not suppressed by blocking PurE (Figure 3, scenario 5). The mutations of this class had no detectable effect on thiamine synthesis when a wild-type PurF was present, consistent with a direct role in the formation of PRA and thus redundant with PurF. This class of mutants consisted primarily of mutations that blocked the oxidative pentose phosphate (OPP) pathway (41, 108). Additional studies with these mutants led to the prediction of a cellular enzyme able to generate PRA from ribose-5-P (41). Reverse genetic approaches have identified an activity in crude extracts that uses ribose-5-P and glutamine (or asparagine) to generate PRA (A. I. Ramos & D. M. Downs, manuscript submitted). The identification of a low-level activity with the substrate predicted from nutritional analyses supported the use of in vivo nutritional analyses to hypothesize biochemical processes.

Mutants unable to generate PRA have not been identified in *Salmonella*. Although mutations that prevent PRA formation under one or more conditions are readily isolated, each of these mutants can grow in the absence of thiamine by changing the growth condition or genetic background. In other words, in every mutant that is unable to generate PRA, an additional mutation can be selected that restores growth. This property suggests the presence of an extensive metabolic network that can compensate for specific perturbations. One explanation of these results is that multiple cellular enzymes contribute to PRA formation.
When mutational analyses that screened for loss of function (i.e., inability to generate thiamine) failed to identify all contributors to PRA formation, another strategy was used. Selections to detect amplification, rather than elimination, of the relevant functions identified additional components involved in PRA formation. One enzyme identified in these studies was anthranilate synthase phosphoribosyltransferase (AS-PRT), a multimeric enzyme complex (encoded by the trpDE genes) that catalyzes the first two reactions in the synthesis of tryptophan (74, 150, 151). A mutation in TrpD (TrpD<sup>P362L</sup>) was isolated, resulting in a mutant AS-PRT complex that contributed to the synthesis of tryptophan and thiamine (115). The above lesion enhanced the PRA-forming ability of the mutant complex and led to the finding that when overproduced (by multicopy or derepression with a trpR mutation) the wild-type enzyme generated the same phenotype (i.e., the ability to generate PRA independent of PurF). An important result of this work was the identification of a previously unknown activity inherent in the native TrpDE enzyme. This result is significant in the context of metabolic potential and is one example supporting the notion that mutant analysis can lead to a deeper understanding of the natural state of the system.

**PRA formation provides a model metabolic node.** As a result of the work done with thiamine synthesis, multiple metabolic inputs to PRA have been identified and are schematically represented in Figure 4. The current view of the complexity is contrasted with the simple view of the pathway presented previously (Figure 4, inset). The inputs to PRA in addition to PurF include a promiscuous enzyme (115), a role of metabolic environment in enzyme activity (B. A. Browne, A. I. Ramos & D. M. Downs, manuscript submitted), and a new activity (A. I. Ramos & D. M. Downs, manuscript submitted). Although PRA is a

---

**Figure 3**

Effect of flux distribution on thiamine synthesis independent of PurF. Five different scenarios of flux distribution through the purine/thiamine pathway are schematically shown. In each case, the shading of the arrows represents the relative flux. Only the darkest arrows support synthesis of the product sufficient to allow growth. The basis for scenarios 1 through 5 is described in the text.
single metabolite in a branched pathway, at least three mechanisms of formation, independent of the classically defined pathway, have been identified. These results highlight the complexities that permeate metabolism and yet remain uncharacterized. Realizing that potentially thousands of such metabolic nodes exist in metabolism, a relevant question is how will these nodes be identified and what assays will allow analysis of both inputs and outputs of each to ensure that the true complexity is uncovered?

**Loci defining a connection of thiamine synthesis and Fe-S cluster metabolism.**

Other lesions that prevent thiamine synthesis in a purF mutant do so by decreasing the efficiency of converting AIR to HMP. This class of mutants requires thiamine only when flux through the common pathway is reduced by genetic or nutritional means. Analyzing mutants of this class led to the conclusions that histidine biosynthesis (via the purine intermediate AICAR) (128), CoA levels, and Fe-S cluster metabolism were each affecting the conversion of AIR to HMP (6, 35a). The largest mutant subclass with this behavior consists of loci that affect the metabolism of Fe-S clusters. Mutations in gdxA, rscC, apbE, and apbC were isolated in genetic screens for thiamine auxotrophy in a purF background (11, 12, 62, 106). Subsequently, isc mutations were added to this class by reconstruction experiments (130). This study illustrates the use
of both convergent and divergent thought in metabolic analysis. It was recognized that lesions in loci of unknown function (apbE, apbC, rseC) generated a phenotype that could also be caused by lesions in loci of known function (gshA, isc).

The resulting hypothesis implicating the unknown loci in the general process affected by the known functions led to the prediction and subsequent demonstration that the previously uncharacterized loci were involved in Fe-S cluster metabolism (convergent logic) (131, 132). The conclusion that Fe-S cluster metabolism is integrated with thiamine synthesis predicts that other genes involved in Fe-S cluster metabolism will also affect thiamine synthesis. Phenotypic analyses of both CyaY (71, 153) and YggX (26, 61, 111) in the context of thiamine synthesis have supported this prediction (35a, 144) (divergent logic). Results of these approaches define connections and provide an important basis for probing function of unknown gene products in the context of the metabolic network.

Mutants compromised in Fe-S cluster metabolism have a decreased efficiency of converting AIR to HMP (35, 35a). ThiC, the only enzyme known to be involved in the conversion of AIR to HMP, is without sequence motifs indicative of a cluster (84). The conversion of AIR to HMP is sensitive to the superoxide-generating compound paraquat, consistent with the involvement of a cluster that can be oxidized (35a). Together, several results have indicated a role for Fe-S cluster metabolism in the conversion of AIR to HMP. The lack of an obvious target for this effect emphasizes the difficulty that arises when in vivo data suggest a conclusion not readily supported by an accepted paradigm.

Mutations compromised in Fe-S cluster metabolism generate an additional defect in thiamine synthesis if the yggX gene is not functional. In this case the requirement is for the thiazole moiety of thiamine (62, 92, 130). The thiazole pathway contains the ThiH enzyme, a member of the S-adenosylmethionine (SAM) radical family of proteins (134) that is characterized by labile Fe-S clusters. Together these results led to the simple hypothesis that mutants compromised in Fe-S cluster synthesis, repair, or both accumulate less active ThiH owing to the oxidation of the cluster (61). This model has been supported by numerous genetic studies, and the purification of ThiH and its various mutant forms will address the biochemical predictions of the model (63, 92, 132). The synergistic effect of YggX on various processes has been exploited to probe its function and the integration of multiple gene products involved in Fe-S cluster metabolism and oxidative damage repair (26, 61, 111). Note that studies on this interesting protein were initiated after it was determined that a spontaneous mutation in this gene had altered phenotypic properties between strains.

Using Phenotypic Analysis to Probe Function of Unknown ORFs

One of the key challenges in understanding metabolism is defining the function and role of uncharacterized ORFs. Isolation of mutants to probe metabolic questions can identify lesions in (a) known genes, (b) genes of unknown (but predictable) function, or (c) genes with no predictable function. Characterization of each gene in the context of the phenotype for which it was isolated provides information about metabolism. For instance, the first scenario may indicate a previously unrecognized connection between processes in metabolism (22, 23, 42). In the second case, phenotypic results may support the predicted function (34, 137) or suggest a different activity (149).

Particularly challenging are those mutations that identify loci for which computational analyses cannot predict a general or specific function. In some ways these ORFs are the most exciting because they may represent fundamentally new functional paradigms in metabolism. The downside of this result is the difficulty in designing subsequent steps in...
their analysis due to lack of an obvious direction to take. An approach to uncharacterized ORFs distinct from that above requires lesions in the relevant gene to cause multiple phenotypes. In this case, varied phenotypes can be considered together to propose a function for the gene product by assuming that each phenotype has a common mechanistic basis. This thinking is well illustrated by the history of describing transcriptional regulators (i.e., Crp), whose global role was initially suggested by the multiple nutritional requirements produced by a null mutation.

**YjgF: a model uncharacterized ORF.** Using lesions that enhanced PRA formation in the absence of both *purF* and a functional OPP pathway, we identified the *yjgF* locus (44). Homologous proteins had a strong presence in the eukaryotic literature, having been attributed a number of interesting characteristics including provocative patterns of accumulation in different tissues and poorly defined activities in vitro (100, 104, 116, 119, 123). Subsequent to its isolation in the context of thiamine synthesis, we and others noted a connection between the *yjgF* locus and branched-chain amino acid biosynthesis in bacteria (60, 124) and yeast (78). Somewhat surprisingly, this protein continues to be the target of intense structural studies, with the publication of more than seven high-resolution structures (20, 32, 33, 91, 93, 105, 129, 145). While the structures consistently reveal a homotrimer, the single study that incorporated in vivo results into the experimental design provided additional information. A study by Parsons et al. (105) with the *Haemophilus influenza* protein indicated that the protein had a weak affinity for intermediates of the branched-chain amino acid pathway. This study offers an example of a productive convergence of physical approaches and in vivo data, providing insight off which future studies can build.

Phenotypes of strains lacking *yjgF* in *Salmonella* so far indicate a connection to at least branched-chain amino acid biosynthesis, thiamine biosynthesis, and carbon catabolism. Pursuit of the connection to branched-chain amino acids, and results of the structural study by Parsons et al. (105), led to a model in which YjgF bound to α-keto acid metabolites, preventing them from mediating a putative toxic effect (124). This model focused on a role of YjgF in branched-chain amino acid biosynthesis, yet homologs of YjgF are present in organisms unable to synthesize these amino acids. This phylogenetic distribution emphasized that YjgF had a more general role in metabolism. In addition, a model for YjgF function had to incorporate the finding that *yjgF* mutations restore PRA formation in a *purF* mutant.

A general model for YjgF function consistent with all *yjgF* mutant phenotypes described in *S. enterica* is illustrated in Figure 5 (38). Simply, this model suggests that the role of YjgF is to bind and eliminate a class of

![Figure 5](https://www.annualreviews.org/doi/10.1146/annurev.micro.60.011605.113601)

**Figure 5**

Schematic of the general model for YjgF function in the cell. The features of this model are described in the text. Metabolite X represents a class of small molecules predicted to be a low-yield product of various metabolic enzymes. These metabolites have one of two fates: (a) In the presence of YjgF they are sequestered/removed or (b) in the absence of YjgF these metabolites are hypothesized to have effects elsewhere in metabolism, including, but not limited to, damaging enzymes and serving as a substrate for reactions not normally occurring in the metabolic environment of a wild-type strain.
metabolites that can have deleterious effects on other components of metabolism. In addition to having the benefit of biochemically testable predictions, the model incorporates a number of premises about metabolism that have been discussed throughout this review. First, enzymes are promiscuous and as such can generate side products that are toxic. Characterized examples of this process include radical oxygen species generated by cellular enzymes (50, 72, 126); keto acids produced in a number of reactions and toxic to various metabolic processes (15, 59, 82); and methylglyoxal, a metabolite that accumulates during uncontrolled carbohydrate metabolism (4, 49, 79).

Second, this model suggests YjgF (potentially like many uncharacterized ORFs) has a supporting role in metabolism. Lesions in yjgF fail to result in a detectable phenotype under laboratory growth conditions and strain backgrounds. Recent results are consistent with a role for YjgF in affecting the metabolic status in the cell (B. A. Browne, A. I. Ramos & D. M. Downs, manuscript submitted). Data suggest that the TrpDE enzyme mediates PRA formation in a yjgF mutant background via a mechanism distinct from that used in the wild-type strain (B. A. Browne, A. I. Ramos & D. M. Downs, manuscript submitted). Therefore, as a general scenario, the model depicted in Figure 5 is consistent with all current data and provides a framework to consider this protein in further studies. YjgF falls in the class of proteins for which no significant functional insight has been provided by bioinformatic analyses, suggesting the exciting possibility that this protein represents a new functional paradigm.

An important aspect of the work on YjgF is the illustration that the in vivo genetic analysis in a model system can provide insight into the functional and cellular role of a protein that is highly conserved and widely distributed. It is advantageous to design experiments in a way that considers the relevant protein in a cellular context (i.e., multiple distinct phenotypes) rather than purses the explanation of a single phenotype. A model becomes more precise as it meets additional constraints placed on it.

Expanding Our Understanding of the Network Through Suppressor Analysis

An effective genetic means to probe metabolic function and define connections is suppression analysis. Use of this strategy is intertwined throughout the examples described above. If metabolic processes can be modified to compensate for a defect in the system, then the component allowing modification can be identified by the site of suppressor mutations. Such mutations can change a phenotype by altering the function of a protein or, alternatively, causing metabolic rerouting. When selecting for suppression of a defect, one can find examples in which the suppressing mutation directly mediates a change (i.e., an altered enzymatic activity that can provide a missing product) (115), or a mutation that indirectly mediates the suppression (i.e., a change in a global gene expression that compensates on a system-wide level) (35, 125). Often a repeating cycle of isolating gain- and loss-of-function mutations can result in significant metabolic insight. This strategy is similar in concept to following a thread through a maze. Suppressor analysis is most valuable when no bias is introduced into the selection of mutations and the complete range of mutants obtained is characterized. The advantage of this approach is that it can provide a breadth of useful metabolic data. The disadvantage of this approach is that the presence of diverse information requires the investigator to maintain and integrate a large and diverse dataset in their thinking. This broad integrative mindset is advantageous in piecing together metabolism, but it is often an overwhelming prospect. The realization that mutations suppressing mutations can be isolated ad nauseum illustrates the challenge metabolic investigators face. An investigator setting out to understand metabolism must
strike a balance between the depth that provides conclusive mechanistic details and the breadth required to describe and characterize a system.

Identification of Additional Nodes for Understanding Metabolism

While the thiamine-centric model system continues to provide insight into metabolism, other model systems need to be advanced. Additional model systems are needed to demonstrate that the approach that has been successful with thiamine synthesis is of general use as a means to further understand metabolism. The strategy of probing metabolism by phenotypic analysis means that all connections cannot be identified from one starting point.

Characteristics of the biosynthetic pathway for coenzyme A (CoA) (via pantothenate) suggest it could be approached in a manner similar to that for thiamine synthesis. Pantothenate is an essential vitamin, and in *Salmonella* (and most bacteria) it is synthesized by the condensation of two moieties that are independently synthesized, pantoate and β-alanine (24, 73). In a feature analogous to thiamine, pantoate is made from a branch off the pathway for valine, integrating synthesis of this vitamin with branched-chain amino acid biosynthesis. The potential of this system is due to the redundancy between the ketopantoate reductase (encoded by the *panE* gene) (52) of the described biosynthetic pathway and the acetohydroxy acid isomeroreductase (encoded by the *ilvC* gene) (110), an enzyme required for the biosynthesis of all branched-chain amino acids. Early studies identified this redundancy on both a biochemical and genetic level (52, 90, 110). What is critical for studies on metabolic integration is that cells lacking PanE are prototrophic but have a 10-fold reduction in the internal CoA levels (51). Thus a system of low flux like that exploited in the thiamine system can be generated. To describe metabolism, it is necessary to repeat a network-generating process from other starting points, ultimately connecting the subsystems defined in each study. This is an exciting time because multiple systems for metabolic integration can be investigated and ultimately combined to further our understanding of metabolism as a system.

EMERGING AND CONTINUING AREAS OF FOCUS IN UNDERSTANDING METABOLISM

Significant interest is focused on describing metabolism in a variety of organisms. These efforts have been prompted by and facilitated with the evolution of genomic and bioinformatics tools. The ability to reconstruct metabolism or define metabolic potential of an organism from the genome sequence is now offered by a number of programs (57, 77, 102, 103, 152). Technologies for mining genomic data continue to progress and offer a means to fill in missing but predicted functions; suggest function by coinheritance; cluster genes by expression, location, and proposed function; and predict metabolic roles and regulatory characteristics. The integration of physical and biological sciences is further evidenced by the evolving efforts to model biological systems with in silico analyses. Continued description of components with experimental science will facilitate the generation of more comprehensive models that provide not only a catalog of data but also a valuable predictive feature.

Recent studies combining in silico models with experimental measurements (i.e., 13C-labeling studies in vivo) are providing a framework (i.e., the model) that defines measurable parameters diagnostic of the status of the system (i.e., phenotype) (5, 40). If one of these parameters deviates from the predicted level as a consequence of a mutational or environmental change, the potential sites of disruption can be suggested by the model. Thus, the model can be used to generate a testable hypothesis to define the translation of a perturbation through the system. This combinatorial
approach is similar to classical in vivo approaches in that it initially focuses on the whole system while ultimately directing study to a specific target. Although the models are necessarily simplified, studies of this type have significant potential because of the recognition that experimental and theoretical results need to be correlated.

CONCLUSIONS
The most pressing need in the field of metabolism is the identification of new functional paradigms and the resulting expansion of the data on which computational programs and mathematical modeling efforts are based. The number of genes that cannot be functionally assigned, despite modern computational programs, emphasizes that fundamentally new functional insights are needed.

Further analysis of metabolism does not require the use of specific tools beyond what biomedical researchers routinely employ to understand metabolic processes. Rather, it requires that the experimental approach is undertaken with a global systems perspective that allows one to follow metabolic results to the point at which relevant connections become clear. Understanding microbial metabolism requires a long-term investment by the investigator and the scientific community. Perhaps this investment starts with the realization that technology is not likely to generate a magic bullet that results in a quick understanding of metabolic complexity and gene function. The proven success of classic logical thought combined with continuing technological advances makes this an era of great potential and excitement for understanding microbial metabolism.

SUMMARY POINTS
1. Microbial metabolism is best appreciated when considered a complex system that is more than the sum of its parts.
2. The current ability to mathematically model (and thus understand) metabolism is limited by our knowledge of component parts and the connections between them.
3. An in vivo, phenotypically driven approach provides the means to suggest the role of uncharacterized ORFs in the context of metabolism.
4. Understanding microbial metabolism requires the integration of in silico, in vitro, and in vivo approaches.

ACKNOWLEDGMENTS
This work was supported by competitive grant GM47296 from the NIH and MCB0096513 from the NSF. Funds were also provided from a 21st Century Scientists Scholars Award from the J.M. McDonnell fund.

LITERATURE CITED
1. Deleted in proof
2. Deleted in proof
3. Deleted in proof


34. Dougherty M, Downs DM. 2003. The *stm4066* gene product of Salmonella enterica serovar Typhimurium has aminoimidazole riboside (AIRs) kinase activity and allows AIRs to satisfy the thiamine requirement of pur mutant strains. *J. Bacteriol.* 185:332–39
39a. *E. coli* online resource. [http://ecoli.aist-nara.ac.jp/](http://ecoli.aist-nara.ac.jp/)


73. Jackowski S. 1996. Biosynthesis of pantothenic acid and coenzyme A. See Ref. 95, 1:687–94


118. Reed JL, Vo TD, Schilling CH, Palsson BO. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904GSM/GPR). *Genome Biol.* 4:R54


