Chapter 7
Systems Analysis of Microbial Adaptations to Simultaneous Stresses

Ross P. Carlson, Olusegun J. Oshota, and Reed L. Taffs

Abstract Microbes live in multi-factorial environments and have evolved under a variety of concurrent stresses including resource scarcity. Their metabolic organization is a reflection of their evolutionary histories and, in spite of decades of research, there is still a need for improved theoretical tools to explain fundamental aspects of microbial physiology. Using ecological and economic concepts, this chapter explores a resource-ratio based theory to elucidate microbial strategies for extracting and channeling mass and energy. The theory assumes cellular fitness is maximized by allocating scarce resources in appropriate proportions to multiple stress responses. Presented case studies deconstruct metabolic networks into a complete set of minimal biochemical pathways known as elementary flux modes. An economic analysis of the elementary flux modes tabulates enzyme atomic synthesis requirements from amino acid sequences and pathway operating costs from catabolic efficiencies, permitting characterization of inherent tradeoffs between resource investment and phenotype. A set of elementary flux modes with competitive tradeoffs properties can be mathematically projected onto experimental fluxomics datasets to decompose measured phenotypes into metabolic adaptations, interpreted as cellular responses proportional to the experienced culturing stresses. The resource-ratio based method describes the experimental phenotypes with greater accuracy than other contemporary approaches and further analysis suggests the results are both statistically and biologically significant. The insight into metabolic network design principles including tradeoffs associated with concurrent stress adaptation provides a foundation for interpreting physiology as well as for rational control and engineering of medically, environmentally, and industrially relevant microbes.

R.P. Carlson (✉) • O.J. Oshota • R.L. Taffs
Chemical and Biological Engineering Department, Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717-3920, USA
e-mail: rossc@coe.montana.edu

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*, Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_7,

Keywords Elementary flux mode • Evolution of metabolism • Nutrient limitation • Resource-ratio theory • Stoichiogenomics

Abbreviations

^{13}C	carbon 13
$C_{fumarate}$	concentration of fumarate
C_{inv}	carbon investment
Cmole	carbon mole
Cmol glc	Cmol glc, carbon moles of glucose
Cmol X	Cmol X, carbon moles of biomass
$C_{op,X}$	carbon operating cost for growth
CμM	micromoles of carbon per liter of cytosol
$\boldsymbol{E}$	matrix containing ecologically important elementary modes
[E]	enzyme concentration
EFM	elementary flux mode
EFMA	elementary flux mode analysis
FBA	flux balance analysis
k_{cat}	catalytic turnover number
K_m	half-saturation constant
MFA	metabolic flux analysis
N_{inv}	nitrogen investment
$N_{inv,X,1:1}$	nitrogen investment for growth, minimalist flux-to-enzyme approach
$O_{2,op,X}$	oxygen operating cost for growth
$\boldsymbol{v}$	vector containing fluxes
v_{max}	maximum enzyme-catalyzed reaction rate
$\boldsymbol{w}$	vector containing weighting factors

7.1 Introduction: Resources and Life

Life is driven by mass and free energy extracted from the environment. The immense diversity of life highlights the variety of these mass and energy forms as well as the array of successful strategies for exploiting them. Metabolic networks channel these resources under an array of environmental constraints and stresses using competitive sets of gene products (Elser et al. 2007). The discipline of ecology has explicitly and implicitly studied mass and free energy flows on a continuum of size scales for more than 100 years. Central themes of many ecological analyses include resource investment strategies, energetic efficiencies of different physiologies and the tradeoffs between the two. These theories, developed and tested on macroscale systems, are useful for decoding the rapidly accumulating omics databases representing microbial-scale systems.

Scarcity of resources over evolutionary time is thought to have influenced the elemental make-up of microbes (Dekel and Alon 2005; Elser et al. 2011; Makino et al. 2003; Sterner and Elser 2002; Zinn et al. 2004). For instance, elemental availability is believed to have influenced the evolution of microbial DNA and amino acid sequences (Bragg and Hyder 2004; Bragg and Wagner 2009). The chromosome sequence of *Pelagibacter* is biased toward low nitrogen containing codons, presumably as a response to the extreme nitrogen scarcity in oceanic ecosystems (Giovannoni et al. 2005). Proteins involved with carbon, nitrogen, and sulfur acquisition contain less of the respective element than average proteins in both prokaryotic and eukaryotic microorganisms (Baudouin-Cornu et al. 2001; Elser et al. 2011). Highly expressed proteins are also thought to be influenced by 'elemental sparing' where the relative material synthesis costs of limiting resources are reduced (Bragg et al. 2012). Strategic use of limiting resources also extends to enzyme cofactors. Respiratory chains require large investments of iron; microbes native to low iron environments often shift from iron containing enzymes to non-iron requiring enzymes such as from ferredoxin to flavodoxin under conditions of iron scarcity (Erdner and Anderson 1999).

It is hypothesized that not only are macromolecule sequences and cofactor requirements influenced by resource scarcity, but also the structure and regulation of entire biochemical networks. Selective pressures have eliminated microbes that fail to allocate limiting resources to cellular functions providing the most competitive return on investment (Carlson 2007, 2009; Dhurjati et al. 1985; Molenaar et al. 2009; Wessely et al. 2011). The return on investment can reflect a multitude of fitness properties including toxin production or resistance, attachment to surfaces or chemotaxis, osmolyte synthesis or salvage, and high-yielding or high-rate extraction of free energy. The role of high-yielding, metabolic efficiency on cellular physiology has been established both theoretically and experimentally (Carlson and Srienc 2004b; Varma and Palsson 1993; Westerhoff et al. 1983).

This chapter details a theory for decoding the multiplicity of microbial strategies to extract and channel mass and energy flows under simultaneous environmental stresses, with a focus on resource scarcity. Elucidating tradeoffs between resource availability and microbial physiology provides a theoretical basis for systemic analyses of omics data, and a rational basis for controlling microbes in medical, environmental and industrial applications.

7.2 *In Silico* Analysis of Metabolic Systems

Notable advances have been made in understanding the basis for metabolic functioning, but there is still need for guiding network design principles (Huang 2000; Papp et al. 2009). These principles need to be investigated through computer-based analyses integrating the numerous metabolic components and their interactions into testable models. For instance, any broad search of the interconnected fluxes of atoms, electrons, and energy in large-scale reaction networks is simply beyond manual feasibility.

Applications of biochemical kinetic models, typically systems of coupled ordinary differential equations describing the evolution of concentrations in time, have illuminated microbial processes for decades (e.g. Kargi and Weissman 1987; Nielsen and Villadsen 1992; Smallbone et al. 2010; Steuer et al. 2006; Straight and Ramkrishna 1994; Varner 2000). Unfortunately, these approaches when applied to metabolic networks are typically either simplistic in terms of the number of processes considered, or in terms of the mechanistic model, or require extensive sets of condition-specific kinetic parameters. Published kinetic parameters, as compiled by databases like BRENDA (Chang et al. 2008; Schomburg et al. 2004), can vary over several orders of magnitude for the same enzyme, requiring expert knowledge for appropriate incorporation into biological models and interpretation of results (Teusink et al. 2000).

An alternative class of successful *in silico* methods, circumventing requirements for large condition-sensitive parameter sets, is known as stoichiometric modeling. Stoichiometric models extract systemic information from conservation relationships and molecular-level network structure, frequently deduced from omics datasets (Reed and Palsson 2003; Trinh et al. 2009). There are three major stoichiometric modeling approaches: metabolic flux analysis (MFA), linear programming methods (also known as flux balance analysis or FBA), and metabolic pathway analysis (including elementary flux mode analysis or EFMA). All three methods define a high-dimension solution space of physiologically permissible metabolic flux distributions based on a stoichiometric matrix specifying system conservation relationships. This solution space is often represented as a pointed convex cone, although the shape can vary depending on the reversibility properties of the reactions (Llaneras and Picó 2010; Wagner and Urbanczik 2005). The three approaches differ in how metabolic flux distributions are selected from the solution space. Detailed descriptions and comparisons of these methods can be found elsewhere (Blank and Kuepfer 2010; Reed and Palsson 2003; Schilling et al. 1999; Trinh et al. 2009).

The continuous solution space defined by stoichiometric methods such as FBA or EFMA contains an effectively infinite number of mathematical possibilities. Identifying biologically and ecologically relevant flux distributions in this continuum requires selection criteria. Maximization of biomass yield on substrate is a widely assumed basis for identifying a cellular physiology. The measure has an appealing simplicity, reflecting an ecologically reasonable structure-function relationship for metabolic networks. This criterion has been utilized successfully to predict and interpret microbial behaviors, including *Escherichia coli* grown in glucose-limited chemostats at modest dilution rates (Carlson and Srienc 2004b; Fong et al. 2003; Fong and Palsson 2004; Varma et al. 1993; Varma and Palsson 1994). However, biomass yield maximization does not always adequately describe metabolic behaviors such as batch growth, or chemostat growth under conditions of nitrogen limitation (Carlson 2007, 2009; Papp et al. 2009; Schuetz et al. 2007; Schuster et al. 2008, 2011).

Because of its role in the current chapter, EFMA will be discussed briefly here. A more detailed description can be found in Chap. 2 of this book which is dedicated to EFMA. EFMA begins with the construction of a stoichiometric matrix representing

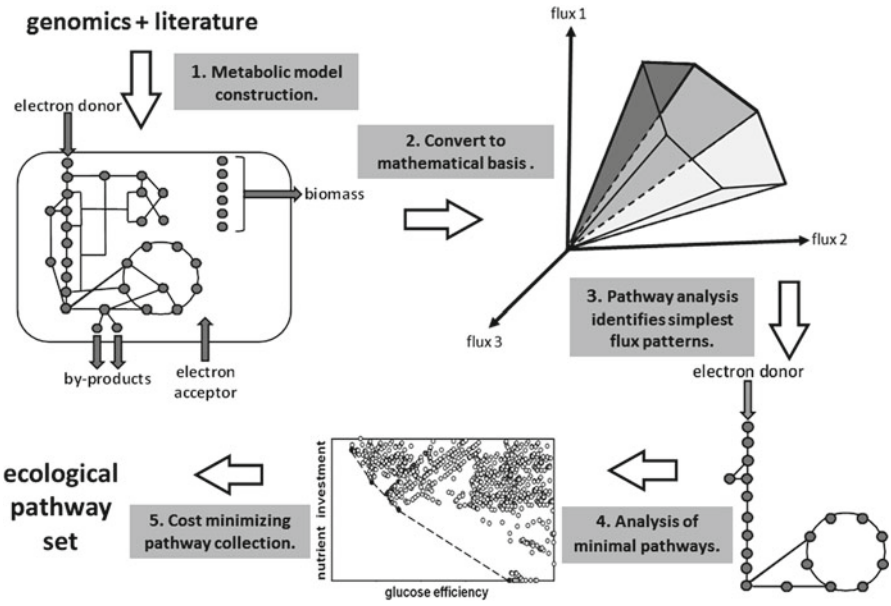

Fig. 7.1 Schematic representation of elementary flux mode analysis (EFMA) model building and cost-benefit analysis. EFMA model building and cost-benefit analyses involve the following steps: (1) metabolic model construction using genomic information and literature resources; (2) model conversion into a mathematical basis, whereby the reaction stoichiometries and reversibilities constrain steady-state cellular phenotypes to flux distributions within a space known as a flux cone; (3) EFMA decomposition of a reaction network into its simplest steady-state pathways, called elementary flux modes (EFMs); (4) cost-benefit analysis translates the EFMs into resource costs based on biochemical and genomic information; (5) minimizing combinations of these resource costs identifies competitive EFMs, collected as ecologically relevant pathway sets

the set of biochemical reactions to be considered. These are frequently compiled based on enzymes annotated from a genomic or metagenomic dataset (Fig. 7.1). The mathematical representation of the model encompasses all thermodynamically relevant system flux distributions. The complete set of enzymatically unique, minimal steady-state pathways spanning this permissible space is known as the elementary flux modes (EFMs) (Gagneur and Klamt 2004; Klamt et al. 2005; Schuster and Hilgetag 1994; Schuster et al. 2000). EFMs allow straightforward investigation of a network's metabolic potential from the bottom up (Klamt and Stelling 2003; Llaneras and Picó 2010; Papin et al. 2004; Trinh et al. 2009). These simplest pathways represent building blocks for steady-state metabolism, and their linear combinations allow system-level analyses of individual microbes, pure cultures, and consortia (e.g. Carlson et al. 2002, 2005; Carlson 2007, 2009; Klamt et al. 2008; Poolman et al. 2004; Taffs et al. 2009; Trinh et al. 2006, 2008; Wlaschin et al. 2006; Zhao and Kurata 2009).

7.3 Analyzing Metabolic Networks with Molecular Level Resource-Ratio Theory

The shortcomings of the often used 'biomass maximization on substrate' criterion highlight the need for additional theoretical frameworks for explaining microbial behaviors. The ecological fields are rich in theoretical approaches for analyzing fitness in various environments. In addition, decades of economic research have highlighted the importance of strategic resource allocation. These theoretical frameworks are well-suited for describing and interpreting many levels of biological organization, from molecular-level metabolic systems to entire ecosystems (Carlson and Taffs 2010). Resource-ratio theory, from the discipline of ecology, is a unifying approach for understanding shared resource competition and investment (Tilman 1980). A recent meta-analysis found that experimental tests supported predictions based on resource-ratio theory 75% of the time, making it one of the most successful theories in ecology (de Mazancourt and Schwartz 2010; Miller et al. 2005).

An approach similar to resource-ratio theory has been applied to *in silico* metabolic networks to quantify the fitness of the genome encoded metabolic potential subjected to pairs of scarce resources (Carlson 2007, 2009). It is hypothesized that evolutionary selection has favored regulation schemes directing resource investment toward effective, stress-specific metabolic pathways. The methodology examines tradeoffs between possible genome-encoded physiologies, represented as EFMs. Cells with regulation schemes permitting effective use of genomic potential would gain a fitness advantage under the relevant stress. To date, the approach has been applied to combinations of two limiting resources which can be either catabolic or anabolic in nature (Carlson and Srienc 2004b; Carlson 2007, 2009). This metabolism-focused application is a departure from traditional resource-ratio theory, which analyzes competition between different organisms for shared resources, rather than comparing potential phenotypes available to a single organism. The approach also shares conceptual similarities with the efficient frontier curve traditionally utilized in economic risk-return analysis and recently applied to biological systems (Kitano 2010).

Case studies in molecular level resource-ratio analysis decomposed *in silico* stoichiometric models of *E. coli* into EFMs (Carlson and Srienc 2004b; Carlson 2007, 2009). As mentioned previously, EFMs are minimal biochemical pathways comprised of metabolite transport and chemical reactions; the enzyme-based steps require an investment of anabolic resources like nitrogen or iron. EFMs, in this context, represent theoretical proteomes for which investment requirements can be tabulated. Calculation of investment requirements necessitates an assumed relationship between fluxes and enzyme concentrations. This relationship varies depending on specific enzyme properties, as well as the chemical environment. Two scenarios are proposed as bounds on the pathway-level flux-to-enzyme concentration relationships. The first approach is a minimalist relationship between pathway enzymes: the concentration ratio of every enzyme pair is set to one. This scenario would utilize varying metabolite pool size and activity modulation including allosteric regulation to achieve a specified flux distribution (this flux-to-enzyme model

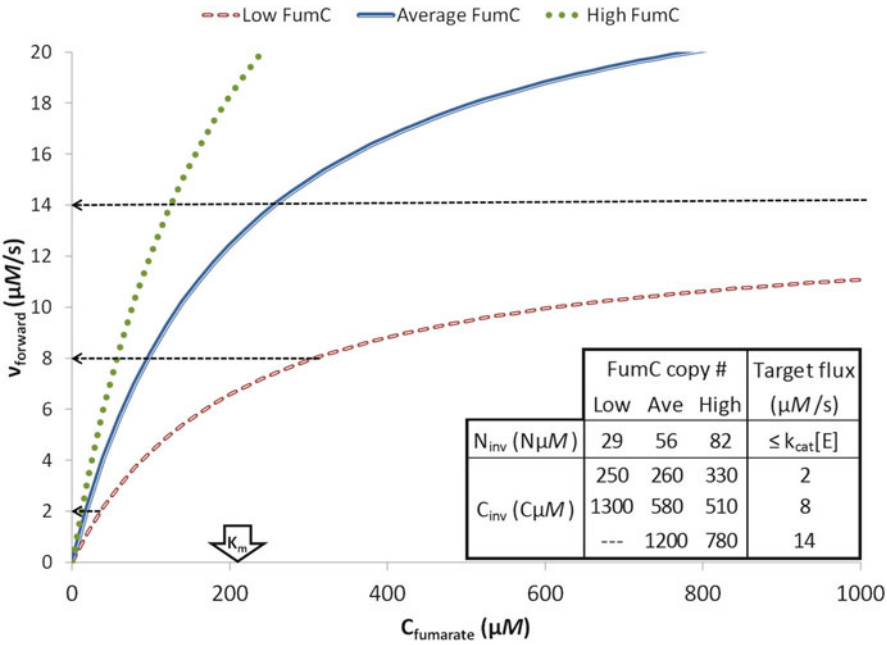

Fig. 7.2 Forward flux through the *E. coli* fumarase FumC (ordinate) as a function of fumarate concentration (abscissa), for three enzyme concentrations (*curved traces*). Each of the dashed *horizontal lines* represents a target flux, i.e. a fitness objective. The intersections of these lines with the various traces represent alternative investment strategies equally capable of achieving the fitness objective. The carbon and nitrogen investments (N_{inv} and C_{inv}) are quantified in the inset table; they represent a sum of both enzyme and metabolite pool investments. Michaelis-Menten kinetics were assumed, and the reverse reaction was neglected. Kinetic parameters for *E. coli* FumC were taken from the literature (Estévez et al. 2002), and the substrate concentration associated with half-saturation of FumC is indicated on the x-axis with an *arrow*. The plotted range of fumarate concentrations is well within intracellular measurements of *E. coli* from the literature (Bennett et al. 2009). The 'low', 'average', and 'high' enzyme concentrations were estimated based on intracellular copy numbers measured in *E. coli* (Taniguchi et al. 2010), assuming a cylindrical cell 2 μm long with a 1 μm diameter

was previously termed the first order method). The second flux-to-enzyme approach assumes a flux proportional relationship. Enzyme concentration ratios for any pathway enzyme pair are proportional to flux ratios through the corresponding reactions. The proportional scenario would use differences in enzyme concentration as the major controller of flux (this model was previously termed the zeroth order method). A comparison of the minimal and proportional flux-to-enzyme approaches to experimental fluxomic data suggests the minimalist approach is a better approximation for *E. coli*, although many additional, potentially relevant, relationships are imaginable (Carlson 2009).

The two flux-to-enzyme approaches are compared in Fig. 7.2 using experimental data for the *E. coli* fumarase enzyme FumC. A flux, representing a fitness objective, can be realized by a number of different proteomes, depending on the associated

metabolome. Two central concepts are illustrated in Fig. 7.2 by the intersections of the fitness objective fluxes (dashed lines) and the enzyme kinetics traces. First, each target flux can be driven by multiple combinations of enzyme and substrate concentrations. Second, low enzyme concentration may not support a high target flux, no matter how high the substrate concentration. These ideas illuminate tradeoffs between alternative investment strategies for achieving the same phenotype.

A straightforward tradeoff exists between investment of substrate into enzymes and into metabolites. Steady state metabolite pools are an investment because stationary concentrations represent substrate that is unavailable for other processes. The optimal strategy in terms of this tradeoff isn't universal: it depends in complex fashion on kinetic parameters, enzyme and substrate stoichiometry, and the target flux. Comparing the second and third lines of the inset table from Fig. 7.2 shows that it takes less carbon investment to drive very small fluxes through FumC with a large fumarate pool, but larger fluxes can be driven more economically with a large enzyme pool. This effect, which may seem counterintuitive initially, is a result of extreme differences in investment per molecule of enzyme versus per molecule of substrate.

Tradeoffs are environment specific and investment of an abundant resource into metabolite pools may not represent a fitness burden. The first line from the Fig. 7.2 inset table illustrates this point: if nitrogen limitation is the only relevant stress, it may be competitive to express the smallest FumC concentration capable of supporting a given target flux, even if that requires a heavy investment of carbon into metabolite pool. Additional environmentally-dependent tradeoffs are also noteworthy. In an environment that changes rapidly, the target flux may increase faster than a microbe can reorganize its proteome. In those environments, it may be competitive to overbuild enzymes for the current condition. As an example, the second line of the inset table from Fig. 7.2 implies that, in a carbon-limited environment with a low flux target, it is effective to drive that flux via the metabolite pool ($C_{inv} = 250$ vs. 330 CμM). This conclusion, however, rests on the assumption that the flux target won't suddenly and drastically increase. If it did, the final line of the inset table from Fig. 7.2 demonstrates that the new flux target might exceed the capacity of the small enzyme pool, providing a competitive advantage to the 'overbuilt' phenotype.

7.4 Constructing Resource Pair Tradeoff Curves

EFM anabolic resource investment requirements can be analyzed in concert with a catabolic physiological fitness metric termed an operating cost. EFM operating costs are the amount of substrate, typically an electron donor or acceptor, required to synthesize a specified amount of cellular product, such as one Cmole of biomass or one mole of ATP: equivalent to inverted yields. The operating cost reflects the net conversion efficiency of the pathway. The relationship of investment requirements against operating costs ranks EFMs efficiently based on combined resource use minimization.

Minimizing combined EFM resource costs for the same cellular product identifies a tradeoff curve, also known as a 'cost minimization envelope' (Carlson and Srienc 2004a, b). Figure 7.3 illustrates two tradeoff curves for the *E. coli* central metabolism. Figure 7.3a is an electron donor (glucose) vs. electron acceptor (oxygen) tradeoff curve while Fig. 7.3b is an electron donor (glucose) vs. anabolic resource (nitrogen) investment cost tradeoff curve. The plots are graphical illustrations of resource-ratio planes defined by the genome encoded metabolic potential. Each circle represents one distinct EFM. The EFM coordinates are the costs in terms of the two plotted resources to synthesize one Cmole of biomass. Physiologies that minimize requirements for resources are hypothesized to represent ecologically and economically competitive strategies. The leftmost EFM along the electron donor axis (abscissa in Fig. 7.3a, b) minimizes the cost of biomass synthesis on glucose; however, this reduction is offset by a higher requirement for the resource plotted on the ordinate. High tradeoff curve resource-ratios (paired resource/electron donor) permit increasingly efficient extraction of substrate free energy, up to a system constrained maximum. Moving right along a tradeoff curve maps competitive physiologies as the resource-ratio of either oxygen or nitrogen decreases relative to glucose. Lowering the resource-ratio increases glucose operating costs (Fig. 7.3).

Tradeoff curves represent combinations of enzymatic steps that provide the highest possible free energy yield based on the interaction between network stoichiometry and resource availability ratios. When the ratio of oxygen or nitrogen to carbon is small, tradeoff analysis indicates it is competitive for *E. coli* to down regulate the citric acid cycle, extracting easily accessible free energy from glycolysis while secreting partially oxidized metabolic intermediates (e.g. acetic or lactic acid). Retaining the 'lost' material and energy through further oxidation would represent a poor investment of the scarce resource. When the same resource ratios are high, the EFMs populating the left extreme of the tradeoff curve indicate it becomes ecologically competitive to completely oxidize glucose. The slopes of a tradeoff curve formalize the diminishing return of free energy extraction from substrate relative to efficient use of the second resource. These specific results are based on the physiological potential of *E. coli*, and general results depend on an organism's genome-encoded potential.

Physiological responses consistent with resource-ratio theory can have tradeoff properties that seem initially contradictory. For instance, the phenomena of microbial overflow metabolisms have been reported under a variety of conditions, including both nutrient excess and scarcity. The widespread occurrence of this metabolic strategy suggests there are fundamental adaptive principles guiding microbial responses (El-Mansi and Holms 1989; El-Mansi 2004; Majewski and Domach 1990; Neijssel et al. 1996; Straight and Ramkrishna 1994; Teixeira de Mattos and Neijssel 1997). The presented resource-ratio approach provides a single theory to explain overflow metabolism for both conditions based on small resource availability ratios. Under conditions of unrestricted environmental resources, the cellular resource ratio is constrained by the ratio of the corresponding transporter capacities (V_{max}). An imbalance in transporter capacity can create nutrient limitation stress responses even when

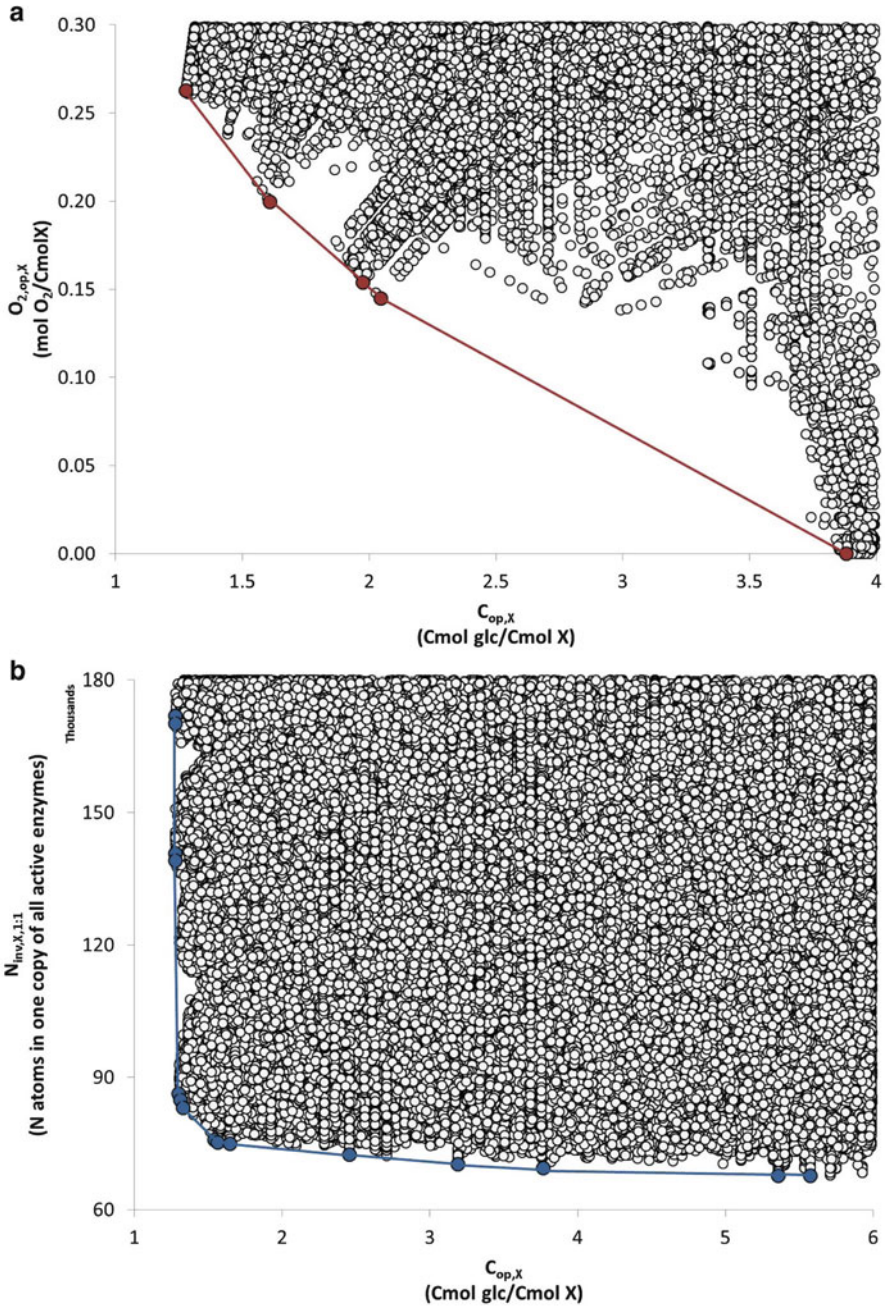

Fig. 7.3 Tradeoff curves for biomass synthesis in *E. coli* MG1655. Each *circle* represents an elementary flux mode (EFM) producing one carbon mole of biomass; *dark filled circles* and *line segments* represent flux distributions minimizing the combined costs plotted on the respective axes. Both abscissae represent glucose operation costs ($C_{op,x}$): the amount of glucose required to

the limiting resource is present at high levels, relative to enzyme half-saturation values. Under conditions of nutrient scarcity, however, availability depends on environmental supply and transporter affinity (K_m). The resource-ratio-based analysis of metabolic potential predicts a physiology consistent with an overflow metabolism in both cases.

Additional *in silico* studies have investigated alternative enzyme investment proxies, including the minimization of total flux or number of biochemical steps (de Figueiredo et al. 2009; Hoffmann et al. 2006; Holzhütter 2004; Poolman et al. 2004; Stelling et al. 2002), a summation of enzyme set molecular weights or volume (Beg et al. 2007; Vazquez et al. 2008a, b; Wessely et al. 2011), and a number of alternatives (Schuetz et al. 2007). While the approaches don't explicitly consider resource-ratio theory, they do consider allocation of resources, and sometimes arrive at similar predicted physiologies. An explicit consideration of resource investment has advantages when the resource is not uniformly distributed across proteins like the cofactor iron. Recent simulations and experiments in our laboratory demonstrated that the iron-limited physiological response is distinct from nitrogen limitation in *E. coli*, and that these differences map well to a resource-ratio interpretation.

7.5 Decomposing Fluxomics Data into Simultaneous Stress Adaptations

Microbes live in complex environments and are exposed frequently to simultaneous demands (Elser et al. 2007). Adaptations to maximize fitness require allocating limiting resources in optimal proportions between multiple cellular responses. These adaptations are proposed to be predictable using economic, evolutionary, and ecological theory (Bloom et al. 1985; Carlson and Taffs 2010; Kitano 2010; Papp et al. 2009; Perrin and Sibly 1993). In fact, dynamic modeling methods have been applied for decades to study simultaneous environmental pressures (Bader 1978; Kooijman 2000; Molenaar et al. 2009; Straight and Ramkrishna 1994), but these models are not yet able to fully capitalize on the omics revolution. Continuing advances in stoichiometry-based network models, however, are enabling systems-wide modeling efforts to understand microbial resource allocation.

A recent study used the stoichiometric modeling approach EFMA to decode potential *E. coli* adaptations to multiple stresses (Carlson 2009). EFMs from the *E. coli* model were translated into resource costs, and a set of ecologically relevant EFMs were assembled from cost minimizing tradeoff curves (Fig. 7.3). Each individual

Fig. 7.3 (continued) build one Cmole of biomass using a pre-existing proteome. The ordinate in subfigure (**a**) represents oxygen operation cost ($O_{2,op,x}$): a similar metric for oxygen. The ordinate in subfigure (**b**) represents nitrogen investment cost ($N_{inv,x}$): the amount of nitrogen contained in the proteome supporting each EFM flux distribution. The plotted investment costs were calculated using the minimalist flux-to-enzyme method

EFM consists of a unique enzyme pattern, providing a basis for flux fingerprinting. The flux distributions were projected onto experimental ^{13}C fluxomics datasets (Schuetz et al. 2007) to decompose measured phenotypes into metabolic responses, interpreted as proportional to the experienced culturing stresses.

The mathematics behind this analysis used experimentally measured ^{13}C-based flux distributions ($\vec{v}$) and the set of ecologically competitive elementary modes (E) identified through paired cost minimizations. The fluxomic data described the *E. coli* central carbon metabolism using ten degrees of freedom; the EFMs were projected into the same ten dimensional space to permit comparison. The proportional contribution of each individual EFM to the overall phenotype is given by a weighting vector $\vec{w}$ such that:

$$\vec{v} = E\vec{w} \tag{7.1}$$

where E is a matrix containing the elementary modes as column vectors. The vector $\vec{w}$ was determined using nonnegative least squares analysis to solve the following problem:

$$\text{Min}_{\vec{w}} \left\| E\vec{w} - \vec{v} \right\|; \forall i : w_i > 0 \tag{7.2}$$

Only positive elements were considered for $\vec{w}$ to limit considerations to biologically meaningful solutions; biomass- and ATP-synthesizing EFMs are not reversible. The magnitude of each weighting element represents the proportional contribution of the corresponding EFM to the mathematical description of measured cellular physiology.

In the described study (Carlson 2009), the *E. coli* metabolic network contained 197,018 distinct EFMs; the set E consisted of 38 distinct pathways identified from tradeoff curve analysis. Only three or four of these pathways were needed to best describe each of the four experimentally measured metabolic phenotypes ($\vec{v}$). These phenotypes comprised unrestricted batch growth, two carbon-limited chemostats (dilution rates of 0.2 and 0.4 h^{-1}), and a nitrogen-limited chemostat (dilution rate of 0.2 h^{-1}). The best fit of $\vec{v}$ for each of the four experimental distributions included EFMs from multiple tradeoff curves, suggesting that every culture investigated was responding to simultaneous stresses. For instance, the batch growth flux distribution was best described as a combination of metabolic strategies likely to be competitive for oxygen and nitrogen limitation, as well as an optimal biomass synthesis pathway. The description is attractive considering the often-reported overflow metabolism associated with *E. coli* unrestricted batch growth (e.g. Meadows et al. 2010; Xu et al. 1999).

In addition to its promising qualitative properties, the accuracy of the *in silico* description was compared quantitatively to an extensive collection of FBA-based flux distributions using the Euclidean distance metric. The set of ecologically selected EFMs described the experimental data with better accuracy than any of the tested FBA-based distributions. The best FBA descriptions required different objective functions for different culturing conditions. The most accurate reported FBA-based

description for the experimental batch growth data utilized the maximization of ATP yield per flux unit objective function (Schuetz et al. 2007).

The nonnegative least squares-based predictions were tested for statistical and biological significance by comparing the ability of 10,000 randomly selected EFM sets (of the same size) to describe the experimentally measured phenotypes. Not one random set described the four tested experimental datasets as accurately as the EFM set selected by minimization of combined costs. A perturbation analysis was performed to characterize the sensitivity of the approach to experimental uncertainty in the measured flux distributions. The predictions were remarkably stable, with the magnitude of the weighting factors falling into a tight distribution and no more than 6% of 10,000 simulations selecting EFMs different than the initial batch growth analysis (Carlson 2009).

7.6 Future Directions

An important consideration in stoichiometric modeling of metabolism is the robustness of phenotype. For every metabolism our group has investigated, the resource-ratio space very near the tradeoff surface includes many additional EFMs (Fig. 7.3). Laboratory chemostat experiments have shown that yield differences as small as 0.5% are selectable (Dykhuizen and Hartl 1980), but selection pressures vary in both magnitude and kind over evolutionary time, suggesting that even greater margins in 'optimal' metabolic behavior are likely fixed in nature. While perhaps not 'competitive' in strictly-controlled, non-varying environments, versatility and resilience matter in the long game, and tradeoffs between robustness and optimal performance are well known in nature as well as engineering (Csete and Doyle 2002; Kitano 2010). Our laboratory is experimenting with different margins of optimality and clustering approaches to find an appropriate balance between the concepts of metabolic robustness, resource requirements and cellular performance for selecting important metabolic phenotypes.

Not all investments produce a benefit that is easy to compare objectively between EFMs. For example, extensive theoretical and experimental work supports the notion that rate-based strategies can outcompete yield-based strategies in certain environments (MacLean 2008; Schuster et al. 2008). Unfortunately, quantifying the relative benefit of rate differences between pathways would obviate the main benefit of stoichiometric modeling: wide applicability without extensive parameterization. Similarly, some investments can be difficult to tabulate; cofactor requirements for enzymes are frequently unknown and potentially flexible with likely fitness consequences, causing difficulties in analysis of trace element scarcity. Differential investments in non-metabolic functions (e.g. chaperones, sensory proteins, or even ribosomal synthesis) can require special effort to be integrated into the tradeoff curve framework, although simple dynamic models are being developed for that purpose (Molenaar et al. 2009). With eukaryotic systems, organelles and their associated biochemical pathways have additional investment requirements that

should be accounted for to adapt the presented methodology. Questions remain: for instance, what additional membrane investment cost should be considered for organelle enzymes, or is this investment negligible when compared to the outer cellular membrane synthesis costs? Detailed ecological models exist for cellular resource budgets, accounting for uptake and growth machinery, storage compounds, and baseline physiologic functions. These are important considerations: while the metabolic enzyme collection is flexible in microbial generalists, the resources invested into other processes are critical throughout microbiology and presumably non-uniform. In addition microbes often express enzymes that are not required for their current circumstances. In a form of 'hedging metabolic bets', microbes may divert a fraction of their resources to a backup plan if the environment suddenly changes. Economic models associated with hedge funds surely will be useful for interpreting what fraction of a resource pool should be invested into risk-management strategies. For these reasons, development of suitable proxies for various types of investments and benefits will continue as an open area in the field of metabolic analysis. The best proxies will be calculable without extensive laboratory work, but effective in selecting a set of modes that is responsive to a specific stress.

Finally, microbial communities can be highly regulated and stable structures based on extensive trading of resources (Miller et al. 2010; Pfeiffer and Bonhoeffer 2004; van der Meer et al. 2005; Wintermute and Silver 2010). Stoichiometric analysis of community metabolism is a growing field (Dias et al. 2008; Stolyar et al. 2007; Taffs et al. 2009; Zhuang et al. 2011; Zomorrodi and Maranas 2012), and aspects of the presented analysis have been extended to coevolved microbial consortia to examine whether members forgo individual optimality for community-related benefits (Taffs et al. 2009). Resource-ratio theory has also recently been expanded to interpret resource exchanges between species (de Mazancourt and Schwartz 2010). Refining this theory to allow stoichiometry-based modeling of exchanges between and relative abundance of interacting microbes will play an important role in deciphering geochemical cycles. Highly efficient natural systems will also provide design blueprints for robust artificial consortia in bioprocess applications (Bernstein et al. 2012).

7.7 Conclusion

Stoichiometric network models are important theoretical tools, facilitating systems-wide modeling efforts for interpreting how microbes allocate resources in response to different environmental demands. This chapter explored microbial phenotypes that maximize fitness in multi-factorial environments through an economic analysis of metabolism. The economic analysis identified costs associated with every genome-encoded phenotype, which were used with ecological resource-ratio theory to characterize phenotypic tradeoff surfaces. These metabolic strategies support the notion of fundamental adaptive principles to the regulation of metabolism. For instance under the seemingly opposite scenarios of unrestricted and scarce resources, the resource-ratio based analysis predicts an overflow metabolism is ecologically

competitive. The presented schemata, developed primarily to understand metabolic network structure and function, provide a rational basis for investigations in environmental ecology, as well as for control of microbial processes in medical and industrial applications.

References

Bader FG (1978) Analysis of double-substrate limited growth. Biotechnol Bioeng 20:183–202

Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D (2001) Molecular evolution of protein atomic composition. Science 293:297–300

Beg QK, Vazquez A, Ernst J, De Menezes MA, Bar-Joseph Z, Barabási AL, Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc Natl Acad Sci U S A 104:12663–12668

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol 5:593–599

Bernstein HC, Paulson SD, Carlson RP (2012) Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. J Biotechnol 157:159–166

Blank LM, Kuepfer L (2010) Metabolic flux distributions: genetic information, computational predictions, and experimental validation. Appl Microbiol Biotechnol 86:1243–1255

Bloom AJ, Chapin FS, Mooney HA (1985) Resource limitation in plants – an economic analogy. Annu Rev Ecol Syst 16:363–392

Bragg JG, Hyder CL (2004) Nitrogen versus carbon use in prokaryotic genomes and proteomes. Proc R Soc Lond Ser B 271:S374–S377

Bragg JG, Wagner A (2009) Protein material costs: single atoms can make an evolutionary difference. Trends Genet 25:5–8

Bragg JG, Quigg A, Raven JA, Wagner A (2012) Protein elemental sparing and codon usage bias are correlated among bacteria. Mol Ecol 21:2480–2487

Carlson RP (2007) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics 23:1258–1264

Carlson RP (2009) Decomposition of complex microbial behaviors into resource-based stress responses. Bioinformatics 25:90–97

Carlson R, Srienc F (2004a) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. Biotechnol Bioeng 85:1–19

Carlson R, Srienc F (2004b) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. Biotechnol Bioeng 86:149–162

Carlson RP, Taffs RL (2010) Molecular-level tradeoffs and metabolic adaptation to simultaneous stressors. Curr Opin Biotechnol 21:670–676

Carlson RP, Fell DA, Srienc F (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol Bioeng 79:121–134

Carlson R, Wlaschin A, Srienc F (2005) Kinetic studies and biochemical pathway analysis of anaerobic poly-(R)-3-hydroxybutyric acid synthesis in *Escherichia coli*. Appl Environ Microbiol 71:713–720

Chang A, Scheer M, Grote A, Schomburg I, Schomburg D (2008) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res 37:D588–D592

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. Science 295:1664–1669

De Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25:3158–3165

De Mazancourt C, Schwartz MW (2010) A resource ratio theory of cooperation. Ecol Lett 13:349–359
Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592
Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27:1–9
Dias JML, Oehmen A, Serafim LS, Lemos PC, Reis MAM, Oliveira R (2008) Metabolic modelling of polyhydroxyalkanoate copolymers production by mixed microbial cultures. BMC Syst Biol 2:59
Dykhuizen DE, Hartl DL (1980) Selective neutrality of 6pgd allozymes in *Escherichia coli* and the effects of genetic background. Genetics 96:801–817
El-Mansi M (2004) Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications. J Ind Microbiol Biotechnol 31:295–300
El-Mansi EM, Holms WH (1989) Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. J Gen Microbiol 135:2875–2883
Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett 10:1135–1142
Elser JJ, Acquisti C, Kumar S (2011) Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. Trends Ecol Evol 26:38–44
Erdner DL, Anderson DM (1999) Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnol Oceanogr 44:1609–1615
Estévez M, Skarda J, Spencer J, Banaszak L, Weaver TM (2002) X-ray crystallographic and kinetic correlation of a clinically observed human fumarase mutation. Protein Sci 11:1552–1557
Fong SS, Palsson BØ (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat Genet 36:1056–1058
Fong SS, Marciniak JY, Palsson BØ (2003) Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. J Bacteriol 185:6400–6408
Gagneur J, Klamt S (2004) Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinform 5:175
Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245
Hoffmann S, Hoppe A, Holzhütter HG (2006) Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). Genome Inf 17:195–207
Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem 271:2905–2922
Huang S (2000) Complexity: the practical problems of post-genomic biology. Nat Biotechnol 18:471–472
Kargi F, Weissman JG (1987) Kinetic parameter estimation in microbial desulfurization of coal. Biotechnol Bioeng 30:1063–1066
Kitano H (2010) Violations of robustness trade-offs. Mol Syst Biol 6:384
Klamt S, Stelling J (2003) Two approaches for metabolic pathway analysis? Trends Biotechnol 21:64–69
Klamt S, Gagneur J, Von Kamp A (2005) Algorithmic approaches for computing elementary modes in large biochemical reaction networks. Syst Biol (Stevenage) 152:249–255
Klamt S, Grammel H, Straube R, Ghosh R, Gilles ED (2008) Modeling the electron transport chain of purple non-sulfur bacteria. Mol Syst Biol 4:156
Kooijman SALM (2000) Dynamic energy and mass budgets in biological systems. Cambridge University Press, Cambridge
Llaneras F, Picó J (2010) Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. J Biomed Biotechnol 753904

Maclean RC (2008) The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. Heredity 100:471–477
Majewski RA, Domach MM (1990) Simple constrained-optimization view of acetate overflow in *Escherichia coli*. Biotechnol Bioeng 35:732–738
Makino W, Cotner JB, Sterner RW, Elser JJ (2003) Are bacteria more like plants or animals? Growth rate and resource dependence of bacterial C:N:P stoichiometry. Funct Ecol 17:121–130
Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B (2010) Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab Eng 12:150–160
Miller TE, Burns JH, Munguia P, Walters EL, Kneitel JM, Richards PM, Mouquet N, Buckley HL (2005) A critical review of twenty years' use of the resource-ratio theory. Am Nat 165:439–448
Miller LD, Mosher JJ, Venkateswaran A, Yang ZK, Palumbo AV, Phelps TJ, Podar M, Schadt CW, Keller M (2010) Establishment and metabolic analysis of a model microbial community for understanding trophic and electron accepting interactions of subsurface anaerobic environments. BMC Microbiol 10:149
Molenaar D, Van Berlo R, De Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol 5:323
Neijssel OM, Teixeira De Mattos MJ, Tempest DW (1996) Growth yield and energy distribution. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella*: Cellular and molecular biology American Society for Microbiology. Washington DC, pp 1683–1692
Nielsen J, Villadsen J (1992) Modeling of microbial kinetics. Chem Eng Sci 47:4225–4270
Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BØ (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405
Papp B, Teusink B, Notebaart RA (2009) A critical view of metabolic network adaptations. HFSP J 3:24–35
Perrin N, Sibly RM (1993) Dynamic models of energy allocation and investment. Annu Rev Ecol Syst 24:379–410
Pfeiffer T, Bonhoeffer S (2004) Evolution of cross-feeding in microbial populations. Am Nat 163:E126–E135
Poolman MG, Venkatesh KV, Pidcock MK, Fell DA (2004) A method for the determination of flux in elementary modes, and its application to lactobacillus rhamnosus. Biotechnol Bioeng 88:601–612
Reed JL, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 185:2692–2699
Schilling CH, Schuster S, Palsson BØ, Heinrich R (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Prog 15:296–303
Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32:D431–D433
Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119
Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2:165–182
Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332
Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252:497–504
Schuster S, De Figueiredo LF, Schroeter A, Kaleta C (2011) Combining metabolic pathway analysis with evolutionary game theory: explaining the occurrence of low-yield pathways by an analytic optimization approach. Biosystems 105:147–153
Smallbone K, Simeonidis E, Swainston N, Mendes P (2010) Towards a genome-scale kinetic model of cellular metabolism. BMC Syst Biol 4:6
Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193
Sterner RW, Elser JJ (2002) Ecological stoichiometry: the biology of elements from molecules to the biosphere. Princeton University Press, Princeton

Steuer R, Gross T, Selbig J, Blasius B (2006) Structural kinetic modeling of metabolic networks. Proc Natl Acad Sci U S A 103:11868–11873

Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA (2007) Metabolic modeling of a mutualistic microbial community. Mol Syst Biol 3:92

Straight JV, Ramkrishna D (1994) Modeling of bacterial growth under multiply-limiting conditions: experiments under carbon- or/and nitrogen-limiting conditions. Biotechnol Prog 10:-588–605

Taffs R, Aston JE, Brileya K, Jay Z, Klatt CG, Mcglynn S, Mallette N, Montross S, Gerlach R, Inskeep WP, Ward DM, Carlson RP (2009) In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. BMC Syst Biol 3:114

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS, Quantifying E (2010) *Coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329:533–538

Teixeira De Mattos MJ, Neijssel OM (1997) Bioenergetic consequences of microbial adaptation to low-nutrient environments. J Biotechnol 59:117–126

Teusink B, Passarge J, Reijenga CA, Esgalhado E, Van Der Weijden CC, Schepper M, Walsh MC, Bakker BM, Van Dam K, Westerhoff HV, Snoep JL (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur J Biochem 267:5313–5329

Tilman D (1980) Resources: a graphical-mechanistic approach to competition and predation. Am Nat 116:362–393

Trinh CT, Carlson R, Wlaschin A, Srienc F (2006) Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab Eng 8:628–638

Trinh CT, Unrean P, Srienc F (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74:3634–3643

Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826

Van Der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M, De Leeuw JW, Damste JSS, Ward DM (2005) Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone national park. Appl Environ Microbiol 71:3978–3986

Varma A, Palsson BØ (1993) Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. J Theor Biol 165:503–522

Varma A, Palsson BØ (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60:3724–3731

Varma A, Boesch BW, Palsson BØ (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59:2465–2473

Varner JD (2000) Large-scale prediction of phenotype: concept. Biotechnol Bioeng 69:664–678

Vazquez A, Beg QK, De Menezes MA, Ernst J, Bar-Joseph Z, Barabási AL, Boros LG, Oltvai ZN (2008a) Impact of the solvent capacity constraint on E. coli metabolism. BMC Syst Biol 2:7

Vazquez A, De Menezes MA, Barabási AL, Oltvai ZN (2008b) Impact of limited solvent capacity on metabolic rate, enzyme activities, and metabolite concentrations of S. cerevisiae glycolysis. PLoS Comput Biol 4:6

Wagner C, Urbanczik R (2005) The geometry of the flux cone of a metabolic network. Biophys J 89:3837–3845

Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Mol Syst Biol 7:515

Westerhoff HV, Hellingwerf KJ, Van Dam K (1983) Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. Proc Natl Acad Sci U S A 80:305–309

Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol 6:407

Wlaschin AP, Trinh CT, Carlson R, Srienc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. Metab Eng 8:338–352

Xu B, Jahic M, Enfors SO (1999) Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. Biotechnol Prog 15:81–90

Zhao Q, Kurata H (2009) Maximum entropy decomposition of flux distribution at steady state to elementary modes. J Biosci Bioeng 107:84–89

Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R, Lovley DR (2011) Genome-scale dynamic modeling of the competition between rhodoferax and geobacter in anoxic subsurface environments. ISME J 5:305–316

Zinn M, Witholt B, Egli T (2004) Dual nutrient limited growth: models, experimental observations, and applications. J Biotechnol 113:263–279

Zomorrodi AR, Maranas CD (2012) OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 8:e1002363