Advances in flux balance analysis

Kenneth J Kauffman, Purusharth Prakash and Jeremy S Edwards*

Biology is going through a paradigm shift from reductionist to holistic, systems-based approaches. The complete genome sequence for a number of organisms is available and the analysis of genome sequence data is proving very useful. Thus, genome sequencing projects and bioinformatic analyses are leading to a complete 'parts catalog' of the molecular components in many organisms. The next challenge will be to reconstruct and simulate overall cellular functions based on the extensive reductionist information. Recent advances have been made in the area of flux balance analysis, a mathematical modeling approach often utilized by metabolic engineers to quantitatively simulate microbial metabolism.

Addresses
University of Delaware, Department of Chemical Engineering, Newark, DE 19716, USA
*e-mail: edwards@che.udel.edu

Current Opinion in Biotechnology 2003, **14**:491–496

This review comes from a themed issue on
Biochemical engineering
Edited by Jeremy S Edwards and Kenneth J Kauffman

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2003.08.001

Abbreviation
FBA flux balance analysis

Introduction

Biology is going through a period of fundamental change. The complete genome sequence for several organisms is available, and this number is growing rapidly [1]. Furthermore, the analysis of genome sequence data is proving very useful; for example, 40 to 80% of the open reading frames identified in the fully sequenced microbial genomes have a reproducible, putative function assignment [2]. Thus, the genome sequencing projects and bioinformatic analysis are leading to a complete 'parts catalog' of the molecular components in many organisms. The next challenge will be to reconstruct and simulate overall cellular functions based on the extensive reductionist information.

The traditional engineering approach to analysis and design for metabolic engineering is to have a mathematical or computer model (e.g. a dynamic simulator) of metabolism that is based on fundamental physicochemical laws and principles. The metabolic engineer hopes that such models can be used to systematically 'design' a new (and improved) living cell. The methods of recombinant DNA technology should then be applied to achieve the desired changes in the genotype of the cell of interest. However, a review in the field has concluded that 'despite the recent surge of interest in metabolic engineering, a great disparity still exists between the power of available molecular biological techniques and the ability to rationally analyze biochemical networks' [3]. This conclusion is not surprising, for cellular networks have evolved over millions of years. As a result, the cell has many interconnected pathways that demonstrate complex regulation. Mischaracterization of the interactions and regulation leads to inadequate dynamic models; however, there are alternative approaches that build on insights derived from bioinformatics and genomics.

This review describes flux balance analysis (FBA), a mathematical modeling approach often utilized by metabolic engineers to quantitatively simulate microbial metabolism. Furthermore, simultaneously genomics and bioinformatics are producing a detailed parts catalog of the molecular components found within a cell. This review brings metabolic engineering, bioinformatics and genomics together and demonstrates how genomic information can be used to build quantitative simulators of cellular functions.

Building flux balance analysis models

Constraint-based modeling

Constraint-based modeling uses physiochemical constraints such as mass balance, energy balance, and flux limitations to describe the potential behavior of an organism [4,5,6•,7–12]. The analysis assumes that under any given environmental condition, the organism will reach a steady state that satisfies the physiochemical constraints. As the constraints on a cellular system are never completely known, multiple steady-state solutions are possible. To identify a physiologically meaningful steady state, an optimization is carried out to find the optimal value of a specified objective function with respect to the constraints identified [13••]. FBA is a constraint-based modeling approach in which the stoichiometry of the underlying biochemical network constrains the solution. A brief history of FBA is shown in Table 1.

Theory

FBA assumes that metabolic networks will reach a steady state constrained by the stoichiometry. The stoichiometric constraints lead to an underdetermined system; however, a bounded solution space of all feasible fluxes can be identified [14]. This solution space can be further restricted by specifying maximum and minimum fluxes through any particular reaction and by specifying other

Table 1

Significant milestones in the development of flux balance analysis.

Year	History of flux balance analysis	References
1984	Papoutsakis used linear programming to calculate maximal theoretical yields	[36]
1986	Fell and Small used linear programming to study lipogenesis	[37]
1990	Majewski and Domach studied acetate overflow during aerobic growth	[38]
1992	Savinell and Palsson performed detailed analysis and development of FBA theory	[39,40]
1993	Varma and Palsson used FBA to describe *E. coli* properties	[9,10,12]
1997	Pramanik and Keasling studied growth rate dependence on biomass concentration	[41,42]
2000	Edwards and Palsson carried out a gene deletion, phase plane, robustness study of *E. coli*	[4,31]
	Lee *et al.* identification of alternative optima	[20]
	Schilling *et al.* integrated FBA with extreme pathway analysis	[14]
2001	Burgard and Maranas examined performance limits of *E. coli* and minimal reaction sets	[27]
	Covert, Schilling and Palsson added regulatory constraints to FBA models	[23]
2002	Papin *et al.* studied network redundancy in *Haemophilus influenzae* (alternate optima)	[22]
	Ibarra, Edwards and Palsson looked at adaptive evolution of *E. coli*	[19••]
	Mahadevan, Edwards and Doyle studied dynamic FBA	[25••]
	Beard, Liang and Qian considered the addition of energy balance constraints to FBA	[26•]

physiochemical constraints. Thus, obtaining these constraints gives us the performance capability of the metabolic network, and the constraints can be refined by adding experimental data [4,5,6•,15–17].

Once the solution space describing the capability of the organism is defined, the network's behavior can be studied by optimizing the steady-state behavior with respect to some objective function [13••]. The simulation results can then be experimentally verified and used to further strengthen the model. Ultimately, the iterative model refinement procedure can result in predictive models of cellular metabolism.

FBA model construction

To better understand FBA, the steps are explained in detail and illustrated through an example (Figure 1).

Step I: system definition

The development of a flux balance model requires the definition of all the metabolic reactions and metabolites (Figure 1a). All reactions should be included in the model. The pathways will be regulated, however, and depending on the environmental conditions only a subset of the reactions will be utilized at any given time. In FBA, the regulation of the reactions or pathways is neglected and mathematical modeling is used to predict the pathway flux without explicit consideration of the regulation.

An ideal starting point for the metabolic reconstruction is the annotated genome sequence. The product of each gene is annotated by homology searches so as to identify all the metabolic enzymes. Next, the reactions catalyzed by each of these enzymes should be described. This requires characterizing the reactants and products for each enzymatic reaction. Often at this stage, one or a few enzymes from a known pathway will be present, but several of the other enzymes will be missing. This is one of the places where the iterative nature of the model development is required. The other reactions in a pathway can be added to close the mass balance based on physiological or biochemical data.

In addition to characterizing all enzymatic reactions, all transport mechanisms must be considered. This includes reactions that diffuse through the membrane, diffuse through pores in the membrane or that are actively transported across the membrane.

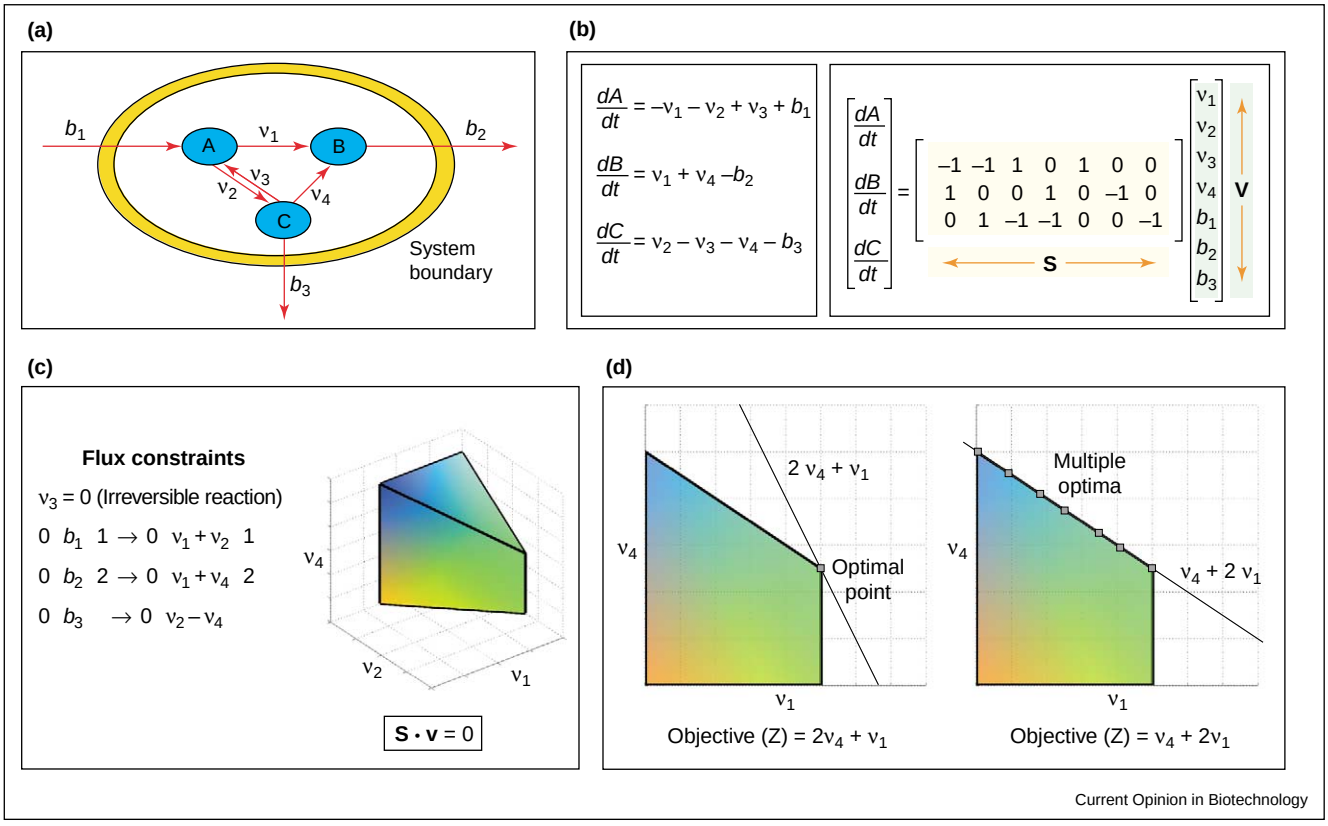
Step II: mass balance

Once all of the reaction and transport mechanisms are identified, a dynamic mass balance is derived for all the metabolites in the metabolic network (Figure 1b). The mass balance is defined in terms of the flux through each reaction and the stoichiometry of that reaction. This gives rise to a set of coupled ordinary differential equations. The differential equations can be represented using a matrix notation, where '**S**' is the stoichiometric matrix and '**V**' is the matrix of the fluxes. The goal of FBA is to identify the metabolic fluxes in the steady-state operation of the metabolic network. As there are more reactions (hence fluxes) than there are metabolites, the steady-state solution for the metabolic fluxes is underdetermined. Thus, additional constraints are needed to uniquely determine the steady-state flux distribution.

Step III: defining measurable fluxes

One way to obtain the additional constraints for the metabolic network, and hence to calculate a value for all fluxes in the network, is to measure fluxes in the metabolic network (the minimum number of measurements is equal to the dimension of the null space). Commonly, the exact flux values are not defined, but rather ranges of allowable flux values are incorporated as additional constraints (Figure 1c).

Figure 1



Methodology for flux balance analysis. **(a)** A model system comprising three metabolites (A, B and C) with three reactions (internal fluxes, v_i, including one reversible reaction) and three exchange fluxes (b_i). **(b)** Mass balance equations accounting for all reactions and transport mechanisms are written for each species. These equations are then rewritten in matrix form. At steady state, this reduced to $\mathbf{S}\cdot\mathbf{V} = 0$. **(c)** The fluxes of the system are constrained on the basis of thermodynamics and experimental insights. This creates a flux cone [14,22] corresponding to the metabolic capacity of the organism. **(d)** Optimization of the system with different objective functions (Z). Case I gives a single optimal point, whereas case II gives multiple optimal points lying along an edge.

Step IV: optimization

Biological metabolic networks always have more fluxes (reactions) in the system than metabolites, but optimization can be used as an alternative to measuring the internal fluxes in the metabolic network (which is a very difficult task). To identify a flux distribution without making additional measurements, it can be assumed that the metabolic network is optimized with respect to a certain objective [13••]. This allows the underdetermined system to be formulated as an optimization problem. If the objective function is linear, the optimization problem is a linear programming problem. Simulations to calculate the internal fluxes of an underdetermined network with the objective function defined as the growth flux have been shown to be consistent with experimental data [13••,18,19••].

For a given objective function (Z) an optimal set of fluxes can be obtained subject to the mass balance ($\mathbf{S}\cdot\mathbf{V} = 0$) and linear inequality ($\alpha_i<V_i<\beta_j$) constraints. It may happen that under certain conditions the system optima lie on an edge instead of a point (Figure 1d). This situation arises when the limiting constraint of the system exactly parallels the objective function. In such cases, the system exhibits multiple optimal solutions along this edge. To identify these alternate optimal solutions a mixed integer linear programming (MILP) approach can be used [20,21]. An analysis of the alternate optimal solutions can be used to find redundancies in the metabolic network [22].

In general, the solution obtained by FBA is only as good as the constraints used to identify it. While specifying known flux limitations constrains the flux space significantly, it still allows for infeasible predictions. Further, characterizing the effect of mutants or shifts in the steady states is difficult to do with traditional FBA [13••].

Second generation flux balance analysis models

In the past three years, FBA models have begun to evolve to incorporate additional biological knowledge. This

evolution can be broken down into three thrusts: incorporation of regulatory constraints, explicit incorporation of thermodynamics, and exploration of alternative classes of objective functions.

Incorporation of regulatory constraints

Regulatory constraints were first imposed as Boolean logic operators by Covert and coworkers [23]. The regulatory constraints represent temporary flux constraints that arise due to a specific environment rather than physiochemical constraints that represent fundamental restrictions on what is possible, regardless of time and space. Using the Boolean logic approach, regulatory constraints were evaluated based on the initial condition of the cellular system. A standard FBA was then carried out, optimizing for growth rate. The calculated solution to the fluxes was then used to re-evaluate the regulatory constraints. This process was repeated over the time course of interest.

Covert and coworkers found that the Boolean approach was sufficient to eliminate a large number of the pathways as infeasible [24••]. This was because many of the extreme pathways required two or more inconsistent regulatory events to occur simultaneously. Thus, the entire extreme pathway could be eliminated from consideration. The analysis further demonstrated the high degree of redundancy in the underlying metabolic network arising from multiple similar extreme pathways available to the cellular system [24••], consistent with previous observations [22].

Mahadevan and coworkers considered transcriptional and translational regulation in a more quantitative manner [25••]. The authors compared two formulations of dynamic FBA for predicting diauxic shift in *Escherichia coli*: the dynamic optimization approach (nonlinear programming problem) and static optimization approach (linear programming approach). The dynamic optimization approach can be used with either an instantaneous objective function or using an end-point objective function — the instantaneous objective function was found to be more consistent with observed behaviors. The static optimization approach is only valid with instantaneous objective functions. Mahadevan and coworkers [25••] concluded that the static optimization approach was computationally simpler to implement provided all of the constraints were linear, whereas the dynamic optimization approach was more flexible and should be quite suitable for the incorporation of experimental data.

The two methods discussed above for incorporating regulatory constraints each have potential problems and limitations; however, the limitations can be reduced by combining the approaches to further gain insight into metabolic regulation.

Explicit incorporation of thermodynamics

In traditional FBA, the thermodynamic constraints are only accounted for in the reversibility/irreversibility of a given reaction. However, the reversibility is dependent on the intracellular conditions, which may change as the environmental conditions change. Beard and coworkers [26•] explored the impact of a full energy balance analysis on the predictions of a flux balance analysis. The analysis requires the solution of a nonlinear optimization problem that provides estimates of the growth rate and the intracellular metabolic fluxes. The nonlinearities arise from the introduction of the free energy changes into the constraints. The resulting feasible solution space is a subset of the space predicted by a traditional FBA; however, due to the nonlinear optimization problem, the approach does not ensure an optimal solution. For *E. coli* metabolism, Beard and coworkers found that combination of the energy balance analysis with FBA gave the same optimal growth rate, but the observed fluxes were substantially different [26•]. The energy balance analysis was also able to explain why certain genes that FBA identified as nonessential were in fact essential — major changes were required in the observed fluxes to compensate for these knockouts [26•]. This explanation is consistent with the observations of Segre and coworkers [13••], which were derived from a different perspective.

Exploration of alternative classes of objective functions

The predictions of FBA are highly dependent on the objective function used for the analysis. Common objective functions include maximization of biomass [4,25••,27,28,29••], maximization of ATP [7] or reducing power, and maximization of the rate of synthesis of a particular product [8].

FBA predictions with maximum growth rate as the objective function are consistent with experimental data approximately 60% of the time for *Helicobacter pylori* [15], approximately 86% of the time for *E. coli* [4], and approximately 91% of the time for *E. coli* when transcriptional regulation was accounted for [23,30]. Alternatively, minimization of metabolic adjustment (MOMA) can be used to improve the prediction efficiency of FBA for studying *E. coli* mutants [13••]. Both MOMA and the transcriptional regulation attempt to account for the burden a cell must adsorb to shift from one operating region to another [13••,23,30].

Maximization of biomass production allows for a wide range of predictions that are consistent with experimental observations for microbial systems [4,15,18,19••,25••,30,31,32••]. However, under some conditions, the behavior of cellular systems is incompatible with maximization of biomass [13••,19••,27,32]. Ibarra and coworkers [19••] demonstrated that *E. coli* will evolve towards maximization of biomass; however, for other situations evaluation

of several objective functions may be necessary to find the most consistent objective for the analysis of the metabolic system of interest [32••]. In other cases, regardless of the objective function used, certain fluxes within a cellular network can be uniquely solved for [33•,34].

Objective functions can be used to explore the capabilities and limitations of a biochemical network. For example, robustness can be explored by varying the maximum flux through a particular pathway and observing the resulting optimization with respect to growth rate. Through such an approach, Edwards and Palsson [31] demonstrated that *E. coli* is robust to changes in individual enzyme or pathway activities. Van Dien and Lidstrom [29••] demonstrated both computationally and experimentally that *Methylobacterium extorquens* AM1 had several redundant pathways that could compensate for each other. Alternatively, the effect of various knockouts and additions on maximum theoretical yields can be explored by maximizing a hypothetical degradation flux on a metabolite of interest. However, any knockouts identified this way should first be screened to ensure viability of the resulting construct.

Conclusions

'Cells obey the laws of [physics and] chemistry' [35], which include conservation of mass, energy and redox potential. Along with other limitations (such as mass transfer), these conservation requirements constrain the cellular behavior, providing bounds on cellular capabilities. Mathematical modeling in which the boundaries of cellular capabilities are defined is proving to be a useful research direction to analyze biological properties; however, many challenges remain.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. TIGR: **Microbial database**. edn September 1, 1998. Edited by: The Institute for Genomic Research; 1998.

2. Iliopoulos I, Tskoka S, Andrade MA, Enright AJ, Carroll M, Poullet P, Promponas V, Liakopoulos T, Palaios G, Pasquier C *et al.*: **Evaluation of annotation strategies using an entire genome sequence**. *Bioinformatics* 2003, **19**:717-726.

3. Stephanopoulos G: **Metabolic engineering**. *Curr Opin Biotechnol* 1994, **5**:196-200.

4. Edwards JS, Palsson BO: **The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities**. *Proc Natl Acad Sci USA* 2000, **97**:5528-5533.

5. Edwards JS, Palsson BO: **Systems properties of the *Haemophilus influenzae* Rd metabolic genotype**. *J Biol Chem* 1999, **274**:17410-17416.

6. • Forster J, Famili I, Fu P, Palsson BO, Nielsen J: **Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network**. *Genome Res* 2003, **13**:244-253.
The largest FBA model to date, consisting of 1175 metabolic reactions and 584 metabolites.

7. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO: **Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints**. *Am J Physiol Regul Integr Comp Physiol* 2001, **280**:R695-R704.

8. Varma A, Boesch BW, Palsson BO: **Biochemical production capabilities of *Escherichia coli***. *Biotechnol Bioeng* 1993, **42**:59-73.

9. Varma A, Palsson BO: **Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors**. *J Theor Biol* 1993, **165**:477-502.

10. Varma A, Palsson BO: **Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns**. *J Theor Biol* 1993, **165**:503-522.

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

12. Varma A, Boesch BW, Palsson BO: **Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates**. *Appl Environ Microbiol* 1993, **59**:2465-2473.

13. •• Segre D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks**. *Proc Natl Acad Sci USA* 2002, **99**:15112-15117.
Defined minimization of metabolic adjustment to more accurately predict knockouts that would be lethal to cellular systems.

14. Schilling CH, Edwards JS, Letscher D, Palsson BO: **Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems**. *Biotechnol Bioeng* 2000, **71**:286-306.

15. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO: **Genome-scale metabolic model of *Helicobacter pylori* 26695**. *J Bacteriol* 2002, **184**:4582-4593.

16. Schilling CH, Edwards JS, Palsson BO: **Toward metabolic phenomics: analysis of genomic data using flux balances**. *Biotechnol Prog* 1999, **15**:288-295.

17. Edwards JS, Ramakrishna R, Schilling CH, Palsson BO: **Metabolic flux balance analysis**. In *Metabolic Engineering*. Edited by Papoutsakis ET: Marcel Dekker; 1999:13-57.

18. Edwards JS, Ibarra RU, Palsson BO: ***In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data**. *Nat Biotechnol* 2001, **19**:125-130.

19. •• Ibarra RU, Edwards JS, Palsson BO: ***Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth**. *Nature* 2002, **420**:186-189.
This work demonstrated that, while the path of evolution cannot be predicted, the final endpoint of evolution for *E. coli* can be predicted using FBA.

20. Lee S, Phalakornkule C, Domach MM, Grossmann IE: **Recursive MILP model for finding all alternate optima in LP models for metabolic networks**. *Comp Chem Eng* 2000, **24**:711-716.

21. Phalakornkule C, Lee S, Zhu T, Koepsel R, Ataai MM, Grossmann IE, Domach MM: **A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering**. *Metab Eng* 2001, **3**:124-137.

22. Papin JA, Price ND, Edwards JS, Palsson BO: **The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy**. *J Theor Biol* 2002, **215**:67-82.

23. Covert MW, Schilling CH, Palsson B: **Regulation of gene expression in flux balance models of metabolism**. *J Theor Biol* 2001, **213**:73-88.

24. •• Covert MW, Palsson BO: **Constraints-based models: regulation of gene expression reduces the steady-state solution space**. *J Theor Biol* 2003, **221**:309-325.

25. •• Mahadevan R, Edwards JS, Doyle FJ: **Dynamic flux balance analysis of diauxic growth in *Escherichia coli***. *Biophys J* 2002, **83**:1331-1340.
Combined, these two manuscripts provide approaches for capturing the dynamic shifts that occur in gene regulation over time. Although substantially different, the two approaches compliment each other.

26. Beard DA, Liang SC, Qian H: **Energy balance for analysis of**
• **complex metabolic networks**. *Biophys J* 2002, **83**:79-86.
Concentration-dependent chemical potentials are used to specify additional nonlinear constraints on the FBA optimization problem. The approach also allows for determination of intracellular concentrations; however, there is no guarantee that the solution is the global optimum.

27. Burgard AP, Maranas CD: **Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions**. *Biotechnol Bioeng* 2001, **74**:364-375.

28. Edwards JS, Ramakrishna R, Palsson BO: **Characterizing the metabolic phenotype: a phenotype phase plane analysis**. *Biotechnol Bioeng* 2002, **77**:27-36.

29. Van Dien SJ, Lidstrom ME: **Stoichiometric model for evaluating**
•• **the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C-3 and C-4 metabolism**. *Biotechnol Bioeng* 2002, **78**:296-312.
Computational prediction and experimental validation of network redundancy.

30. Covert MW, Palsson BO: **Transcriptional regulation in constraints-based metabolic models of *Escherichia coli***. *J Biol Chem* 2002, **277**:28058-28064.

31. Edwards JS, Palsson BO: **Robustness analysis of the *Escherichia coli* metabolic network**. *Biotechnol Prog* 2000, **16**:927-939.

32. Burgard AP, Maranas CD: **Optimization-based framework for**
•• **inferring and testing hypothesized metabolic objective functions**. *Biotechnol Bioeng* 2003, **82**:670-677.
Optimization is used to determine if a given objective function is consistent with experimentally observed fluxes in *E. coli*.

33. Klamt S, Schuster S: **Calculating as many fluxes as possible in**
• **underdetermined metabolic networks**. *Mol Biol Rep* 2002, **29**:243-248.
Although it is not possible to uniquely determine all of the fluxes in an underdetermined system, some of the fluxes can be uniquely specified.

34. Klamt S, Schuster S, Gilles ED: **Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria**. *Biotechnol Bioeng* 2002, **77**:734-751.

35. Watson JD: *Molecular Biology of the Cell*, edn 1. New York: Garland Publishing Inc.; 1972.

36. Papoutsakis ET: **Equations and calculations for fermentations of butyric acid bacteria**. *Biotechnol Bioeng* 1984, **26**:174-187.

37. Fell DA, Small JA: **Fat synthesis in adipose tissue. An examination of stoichiometric constraints**. *J Biochem* 1986, **238**:781-786.

38. Majewski RA, Domach MM: **Simple constrained optimization view of acetate overflow in *E. coli***. *Biotechnol Bioeng* 1990, **35**:731-738.

39. Savinell JM, Palsson BO: **Optimal selection of metabolic fluxes for *in vivo* measurement. I. Development of mathematical methods**. *J Theor Biol* 1992, **155**:201-214.

40. Savinell JM, Palsson BO: **Optimal selection of metabolic fluxes for *in vivo* measurement. II. Application to *Escherichia coli* and hybridoma cell metabolism**. *J Theor Biol* 1992, **155**:215-242.

41. Pramanik J, Keasling JD: **Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements**. *Biotechnol Bioeng* 1997, **56**:398-421.

42. Pramanik J, Keasling JD: **Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model**. *Biotechnol Bioeng* 1998, **60**:230-238.