BRIEFINGS IN BIOINFORMATICS. VOL 7. NO 2. 140–150
Advance Access publication April 26, 2006
doi:10.1093/bib/bbl007

Flux balance analysis in the era of metabolomics

Jong Min Lee, Erwin P. Gianchandani and Jason A. Papin

Submitted: 27th December 2005; Received (in revised form): 30th March 2006

Abstract

Flux balance analysis (FBA) has emerged as an effective means to analyse biological networks in a quantitative manner. Much progress has been made on the extension of FBA to incorporate a priori biological knowledge, provide more practical descriptions of observed cell behaviours, and predict the outcome of network perturbations. Metabolomics is independently advancing as a set of high-throughput data acquisition tools providing dynamic profiles of metabolites in an unbiased manner. These data sets are neither yet sufficiently comprehensive nor accurate enough for generating large-scale kinetic models. Thus, there is a pressing need to develop quantitative techniques that can make use of the emerging data and embrace the associated uncertainties. This article reviews recent advances in FBA to meet this need and discusses the utility of FBA as a complement to metabolomics and the expected synergy as a result of combining these two techniques.

Keywords: *metabolism; metabolome; genome scale analysis; network reconstruction; objective function; optimization*

INTRODUCTION

Metabolomics refers to an analysis technology that aims to identify and quantify *all* metabolites ('metabolome') present in a biological sample from an organism grown under defined conditions [1, 2]. It quantifies metabolites in a non-targeted, non-biased and comprehensive manner using various analytical techniques including nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). These tools are often combined with different separation techniques including gas chromatography (GC), liquid chromatography (LC) and capillary electrophoresis (CE) [3]. The resulting analytical spectra are compared with known samples using statistical techniques such as pattern recognition [4]. Though metabolomics has only recently emerged, dynamic profiles of metabolites at a large scale for a multitude of biological systems are being generated rapidly and have seen successful application in various microbial [5], plant [6, 7] and human [8] systems.

Thus, metabolomics is increasingly considered to be a promising technology for clinical applications. For example, recent studies suggest that an analysis of the entire metabolic profile in mitochondria is required for characterizing tumourigenesis, while previous studies had primarily focused on identifying particular biomarkers derived from a small number of genes [9, 10]. In addition, a complete metabolite profiling strategy may provide a quantitative estimate of disease risk (e.g. diabetes, heart disease and cancer) and suggest a targeted treatment strategy [11].

Despite its importance and potential, metabolomic modelling and analysis is at a beginning stage. Though the current analysis techniques are becoming faster and cheaper, comprehensive and precise identification of all involved metabolites is still

Corresponding author. Jason A. Papin, Department of Biomedical Engineering, University of Virginia, PO Box 800759, Charlottesville, VA 22908, USA. Tel: +1–434-924-8195; E-mail: papin@virginia.edu

Jong Min Lee is a postdoctoral research associate in the Department of Biomedical Engineering at the University of Virginia. His research interests include dynamic analysis of integrated cellular networks and machine learning approaches to the understanding of signal transduction.

Erwin P. Gianchandani is a graduate student in the Department of Biomedical Engineering at the University of Virginia. His research interests include applying systems analyses to the integration of metabolic, regulatory and signalling networks for the characterization of whole-cell properties.

Jason A. Papin is an Assistant Professor of Biomedical Engineering at the University of Virginia. His group has ongoing projects on polymicrobial interactions in human disease, signalling networks in cancer and the integration of signalling, regulatory and metabolic networks in embryogenesis.

considered to be challenging. There is a significant discrepancy between theoretically expected values (e.g. as calculated from a mass balance on the participating species) and the measured quantities of compounds. This is mainly attributable to the fact that many metabolites are found at very low concentrations, which necessitates additional technology development [3, 4, 12]. Much effort is still required to develop this technology beyond the scope of predicting a subset of metabolic reactions [13]. Thus, modelling approaches based on the concentration profiles of metabolites are limited in that the underlying network structure and associated kinetic constants are not known with complete accuracy [14]. Furthermore, quantitative approaches for experimentally measuring flux distributions, known as metabolic flux analysis (MFA) [15], have just begun to be extended to large-scale systems in a high-throughput fashion. For example, the flux distribution of 137 null mutants of *Bacillus subtilis* was recently characterized [16].

A computational framework that complements such 'incomplete' data sets effectively is required to quantify the underlying characteristics of a biological system. A variety of techniques for metabolic network analysis have been developed, including kinetic modelling, which considers concentrations of compounds [17]; metabolic control analysis (MCA), which evaluates the effects of small variations of enzyme or metabolite concentrations [13]; and most recently, energy balance analysis (EBA), which interrogates a network while enforcing the laws of thermodynamics [18]. This article presents one such technique called flux balance analysis (FBA) as a modelling and analysis tool for metabolomics. We review recent advances and current challenges in FBA and discuss potential benefits in applying FBA to metabolomic data.

FUNDAMENTALS OF FBA

All expressed phenotypes of a given biological system must satisfy basic constraints that are imposed on the functions of all cells [19, 20]. These constraints are physicochemical, topological and environmental. Briefly, physicochemical constraints are physical laws like conservation of mass and energy; topological constraints represent spatial restrictions on metabolites within cellular compartments; and environmental constraints include nutrient availability, pH and temperature that vary over time and space. These constraints can be described mathematically, and the set of possible solutions to these mathematical descriptions provides the range of valid states of the biological system. These constraints narrow the spectrum of possible phenotypes and provide an approach for more specifically characterizing cellular network function. Constraints-based modelling has been successful in metabolism, regulation and signalling [20, 21].

FBA is specifically concerned with deriving a feasible set of steady-state fluxes that optimizes a stated cellular objective, e.g. maximizing biomass production within a metabolic network, subject to a set of constraints of conservation of mass [20]. Once this set of steady-state fluxes is identified, optimization techniques may be utilized to evaluate the performance of the biological system at various perturbations, such as different cellular objectives or variability in the imposed constraints [22]. The resultant sets of fluxes may be compared with each other and with experimental data, and, ultimately, the collection of possible fluxes may yield predictive models of large-scale biochemical networks exposed to different conditions. (For a brief history of FBA, see [23].)

BASIC FORMULATION

The first step in an FBA method is the reconstruction of a biochemical network. The reconstruction process begins with an annotated genome sequence, wherein the identified genes correspond to proteins that catalyse reactions within the network [24]. Proteins missing from the network following the initial reconstruction can be identified subsequently through additional experimentation as part of an iterative model refinement process.

An example of a reaction network is presented in Figure 1. This reaction network is used to illustrate the subsequent analysis process. The biochemical network reconstruction is represented in a matrix **S**. **S** is an $m \times n$ matrix of stoichiometric coefficients that captures the underlying reactions of the biochemical network. The rows of **S** correspond to the compounds, and the columns of **S** correspond to reactions. The elements of the matrix are the associated stoichiometric coefficients. For example, the illustrative network has three compounds and seven reactions and is captured by an **S** matrix of size three rows by seven columns. Negative elements of the matrix represent the consumption of compounds and positive values denote production. For example, the first column describes a reaction wherein 1 mole of A is consumed (−1) and 1 mole of B is produced (+1).

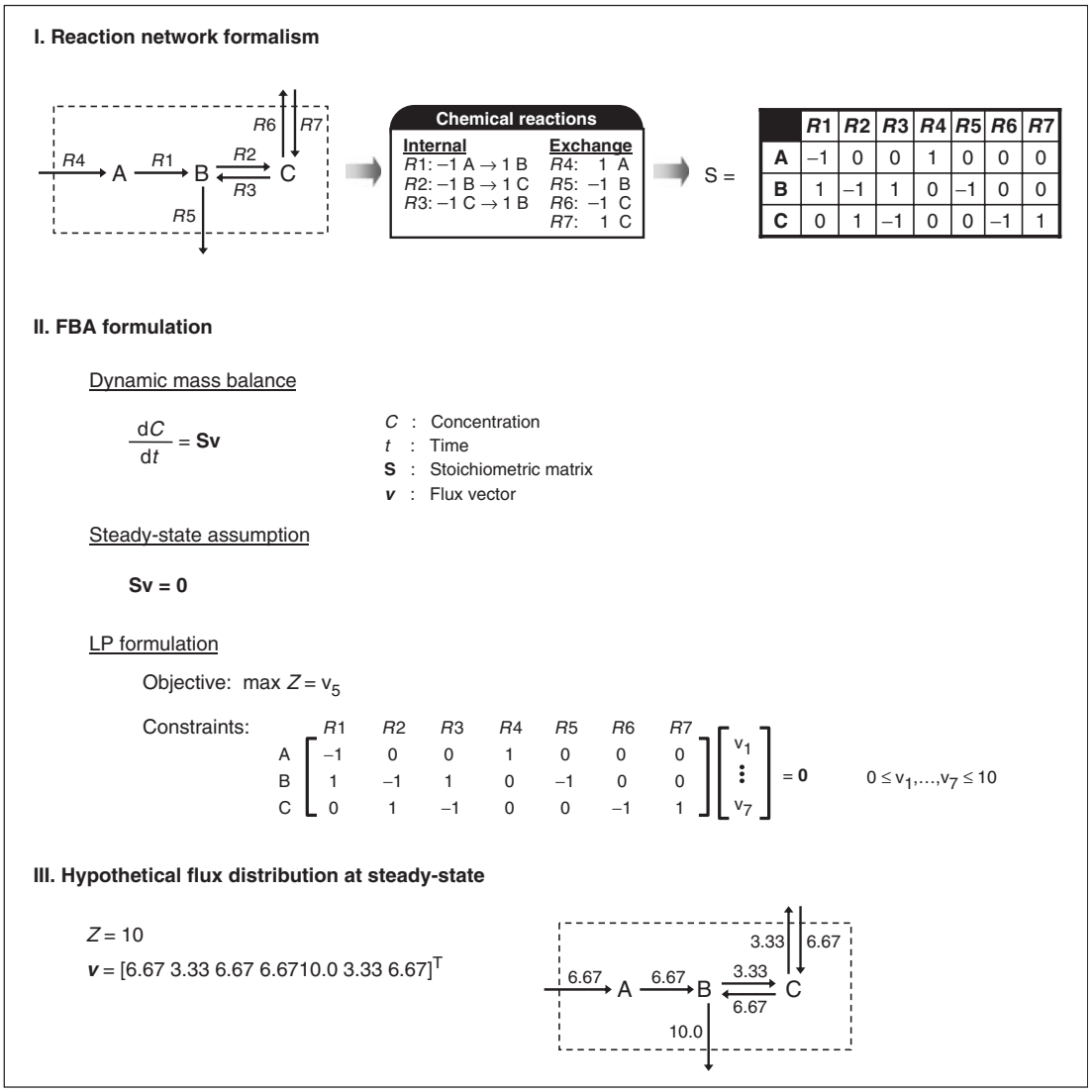

Figure 1: Illustration of a reaction network and flux balance analysis. The reaction network is represented in a stoichiometric matrix **S** and the steady-state assumption is translated into a set of algebraic constraints on a distribution of fluxes (reaction rates). Other constraints can also be imposed such as upper and lower bounds on each flux. Since the number of fluxes is generally larger than the number of compounds, optimization is employed to determine the optimal distribution of fluxes given an objective function (hypothesis). Here, we maximize the flux v_5, which is associated with biomass production. The optimal value of v_5 was found to be 10.0 with the given set of constraints.

After the network is reconstructed, mass balance can be defined in terms of the flux through each reaction and the stoichiometry of that reaction in the form of

$$\frac{dx}{dt} = \mathbf{S} \cdot \mathbf{v} \tag{1}$$

where $\mathbf{v}$ denotes the vector of fluxes with elements corresponding to the fluxes in given reactions (columns) in $\mathbf{S}$.

At steady-state, the change in the amount of a compound x over time t across all reactions within the system becomes zero. This assumption is relevant for most intracellular reactions since they are typically much faster than the rate of change in the resultant cellular phenotypes such as cell growth and process dynamics (e.g. [25]). The steady-state assumption therefore eliminates the time derivative in Equation (1), yielding

$$\mathbf{S} \cdot \mathbf{v} = 0 \tag{2}$$

It is this mass balance that represents the principal constraint in FBA. Additional constraints such as

reversibility of a reaction and energy requirement for cell maintenance can be conveniently incorporated into FBA [25]. This physicochemical constraint by itself represents a set of linear equations wherein the number of equations (one per reactant) is far less than the number of unknown variables (reaction fluxes). Consequently, this set of linear equations is under-determined [23]. FBA, therefore, typically involves optimizing the set of fluxes such that a particular cellular objective is achieved. Objective functions used in practice can take on a linear form

$$Z = \mathbf{c} \cdot \mathbf{v} \tag{3}$$

where $\mathbf{c}$ denotes the vector that defines the coefficients, or weights, for each of the fluxes in $\mathbf{v}$ [18]. This general representation of Z, wherein the elements of $\mathbf{c}$ can be manipulated, enables the formulation of a number of diverse objectives. The optimization strategy employed by FBA attempts to find a solution $\mathbf{v}$ that optimizes Z, while still lying in the bounded solution space defined by the set of physicochemical constraints. Assuming the objective function is a linear equation, FBA constitutes a linear programming (LP) problem that can be easily implemented in a computational environment, even for large-scale systems [24].

Objective functions can take on many forms, including physiologically meaningful objectives as well as design objectives for the interrogation or exploitation of a given system. Common objective functions include maximizing biomass or cell growth, maximizing ATP production or maximizing the rate of synthesis of a particular product [23]. Other objective functions include minimizing ATP production in order to determine conditions of optimal metabolic energy efficiency; and minimizing nutrient uptake in order to evaluate the conditions under which a cell will perform its metabolic functions while consuming the minimum amount of nutrients, etc. In general, maximizing biomass production yields *in silico* predictions that are consistent with experimental observations [23] (for examples of the study of growth as the cellular objective in *Escherichia coli* and *Helicobacter pylori*, see [23, 26–28]). Furthermore, the biochemical production capabilities of mutant strains of *E. coli* have been characterized by considering maximal metabolite production [22].

RECENT EXTENSIONS TO THE BASIC FBA

A central hypothesis of FBA is that biological networks behave optimally, driven by an objective. This assumption has attracted substantial attention and prompted the development of advanced formulations and complementary strategies of FBA in an effort to derive additional biologically meaningful results. In the rest of this section, we review the recent developments of FBA, which are classified by the major goals of the novel strategies.

Incorporation of additional constraints

Whereas conventional FBA finds a solution in the constrained flux space defined by mass conservation and pre-specified reversibility of the given reactions, several strategies accommodating additional constraints have been suggested to account for gene regulation and free energy.

Gene regulation orchestrates the expression of each gene. Boolean logical operators were used to account for the effects of transcriptional regulation on cellular metabolism [27, 29, 30]. A set of Boolean statements are associated with the regulatory rules for a given gene. If the rules are satisfied, the gene is 'on' and the corresponding protein and its associated reaction participate in the network. The Boolean rules are derived from experimental work characterizing regulatory processes, including e.g. microarray data. This formulation was implemented in a genome-scale model of *E. coli* to account for 149 regulatory genes and 45 regulated reactions [27].

Reaction thermodynamics were incorporated in FBA by imposing additional non-linear constraints describing energy balance with the chemical potential [18]. This formulation eliminated the thermodynamically infeasible optimal solutions of *E. coli* metabolism calculated by FBA, which refined predictions regarding gene essentiality. The energy balance formulation was extended further to exclude thermodynamically infeasible cycles from a given reaction network [31]. In this approach, the multiplication between fluxes and associated chemical potential leads to computational complexity for large-scale systems. In addition, the estimation of thermodynamic parameters should be performed judiciously [32].

Variant forms of objective functions for engineered cells

The common choices for an objective function in FBA of prokaryotic metabolism include biomass production [25, 26], energy [33] and by-product production [34]. The rationale behind this practice is that the organism has evolved to maximize the performance under conditions to which it was previously adapted.

Artificially modified cells with gene knockouts may not be accurately described with the objectives used for wild-type cells. In order to better understand the flux states of mutants, a FBA formulation termed MOMA (minimization of metabolic adjustment) was proposed [22]. MOMA solves a quadratic programme (Figure 2) to find the flux state of a mutant with the minimum Euclidean distance to the optimal flux point calculated from FBA of its wild-type counterpart. Effectively, MOMA calculated the feasible flux distribution 'closest' to the optimal flux distribution prior to the gene knockout. This approach was used successfully to predict viability and quantitative flux distributions in *E. coli* after simulated gene knockouts, particularly for cases in which FBA previously failed. It has also been used as a practical tool to engineer strains for metabolite production purposes. Gene knockout simulations based on MOMA experimentally proved capable of refining candidate mutations for increasing the production level of lycopene, an effective anti-oxidant compound, in *E. coli* [35]. It is particularly relevant for characterizing initial transient behaviour after gene manipulation.

ROOM (regulatory on/off minimization) is a recent development that identifies the metabolic flux state of mutants by minimizing the number of 'significant' flux changes from the wild-type flux distribution [36]. The 'significance' is defined by a set of threshold values that a user specifies considering manageable computation time. Effectively, this formulation assigns a cost to the expression of a gene and finds the feasible flux distribution with a minimal number of Boolean flux changes from the flux distribution of the wild-type strain. The resulting optimization problem involves mixed integer linear programming (MILP) (Figure 2). ROOM accurately predicted the post-adaptation states (flux distribution after a significant period of time) for *E. coli* metabolism.

One common aspect between MOMA and ROOM is the calculation of a flux state for a mutant based on a minimal adjustment metric. As these approaches show, the resultant flux state depends on the definition of a distance metric. In principle, other definitions of distance could be explored, but it is still an open question as to which one works better for particular situations.

Solution to the inverse problem

FBA is mainly concerned with the prediction of steady-state fluxes given a set of parameters (e.g. **c** in Equation (3) and the availability of particular gene functions) defining the objective function and stoichiometric model. The inverse problem of this FBA prediction consists of using actual flux observations or desired flux distributions to infer the associated parameters. Such an inverse problem may be difficult to solve because it involves the exploration of a large parameter space, a task that is computationally demanding.

An optimization-based framework called ObjFind proved to be effective for inferring the most plausible objective function given observed experimental data [37]. It is based on a bi-level optimization problem, where the inner optimization is a conventional FBA problem with undetermined parameters and the outer optimization is for evaluating consistency with observed fluxes formulated by a quadratic programme. For computational simplification, a duality condition is introduced and the inner LP optimization problem is converted to a set of non-linear constraints, leading to single-level non-linear programming (NLP) (Figure 2). The non-linearity arises from multiplication between coefficients of the objective and associated fluxes. This approach was applied to *E. coli* metabolism under different culture conditions and the best match to experimentally observed fluxes indicated an objective function associated with maximizing biomass.

The bi-level formulation for inverse optimization problems was extended further to identify gene deletion strategies for the overproduction of a desired metabolite (OptKnock) [38]. The decision variables are binary variables indicating the on/off status of fluxes, and the resulting optimization problem is a single-level MILP. OptKnock identified non-intuitive strategies for production of succinate, lactate and 1,3-propanediol in *E. coli* metabolism.

Dynamic optimization

FBA was used to generate dynamic profiles of cell growth and substrate and by-product

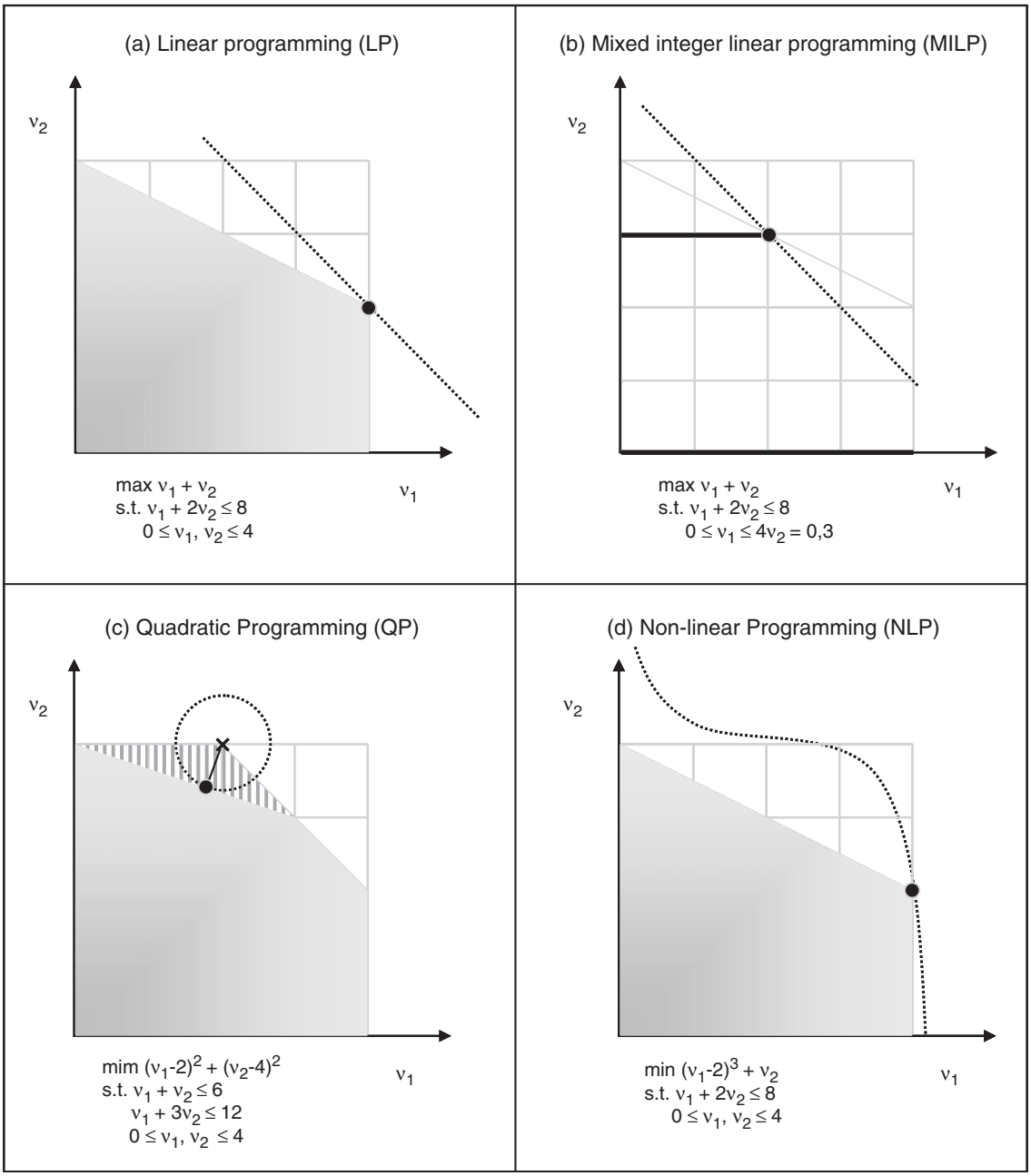

Figure 2: Optimization techniques for flux balance analysis. The formulation of FBA and its solution algorithm are dependent on the objective function and constraints. v_1 and v_2 are fluxes and the shaded area and solid lines represent the feasible solution space defined by flux capacities, stoichiometric relationships, and design specification (e.g. gene deletions). The dotted lines and circular dots represent objective functions and optimal solutions, respectively. (a) Linear programming is a standard technique for conventional FBA. It consists of a linear objective function and linear constraints. Flux capacities, mass balances, and thermodynamic constraints define the feasible space, and maximization of biomass growth is the typical objective function as found in [25]. (b) 'Mixed Integer' denotes that integer variables are involved in a linear programming problem. This can arise from binary variable formulation in considering gene deletion (e.g. ROOM [36] and OptKnock [38]). (c) Quadratic programming (QP) concerns optimization of a quadratic objective function subject to linear constraints. This technique is generally used for finding the closest point to a specified one in the feasible region as in MOMA [22]. For example, the X mark can be viewed as the optimal solution for a wild-type cell and the circular dot is the optimal solution for a mutant. (d) Optimization problems involving non-linear objectives or constraints (except for QP) are referred to as non-linear programming (NLP). It is generally difficult to solve for global optimal solution because of its non-convexity. An example of solving NLP in FBA is identification of a relevant objective function where bi-level optimization formulation is converted to a non-linear optimization problem for identifying objective functions that provide the best fit to a set of experimental flux data [37].

concentrations [25, 27, 30]. Typically, this approach involves discretizing the entire time course of interest and solving FBA at the beginning of each time interval, followed by integration of the cell growth and the extracellular concentrations over the interval. This approach is based on the quasi-steadystate assumption that fluxes are constant over the time interval.

Recently, a dynamic optimization approach was developed to account for flux changes in future time steps [39]. The motivation for this modification is that objectives that span multiple time steps can be incorporated (e.g. maximal biomass at the end of the fermentation). This formulation parametrizes fluxes, metabolite levels and biomass concentration as a function of time using orthogonal collocation [40] to approximate temporal differential terms as algebraic expressions of fluxes. The resulting formulation becomes an NLP problem. Application to diauxic growth of *E. coli* showed that while the static optimization formulation—a special case of dynamic optimization with an instantaneous objective function—is scalable to a large network and more consistent with experimental observations when the intracellular condition is at a steady state. This dynamic optimization approach may be suitable for the analysis of transient behaviours in response to system perturbations/fluctuations [39].

Elucidating structural properties

Updates to FBA have been used to identify topological features of a network such as metabolite connectivity, reaction flux interdependency and pathway redundancy. One such approach, called flux coupling analysis, characterizes interdependency between reaction fluxes by sequentially maximizing all the fluxes normalized to a particular flux [41]. This approach can quantify responses of metabolic networks to the perturbation of particular fluxes. Such an optimization-based approach is scalable to genome-scale networks, for which the use of other techniques may be prohibitive [42].

Graph theory was combined with FBA to characterize redundancy in metabolic pathways [43]. This pathway redundancy renders the network robust against the failure of components (e.g. genes and enzymes) that disrupt some of the pathways. In the approach, FBA provides the initial candidate sequence of metabolic reactions, and the directed-graph representation of the candidate reactions is examined in a combinatorial manner to identify redundant reaction pathways. Application of this technique to *E. coli* metabolism with 48 reactions identified multiple metabolic pathways under different culture conditions. For example, eight multiple reaction pathways were found under the objective of maximizing acetate production.

FBA solutions can elucidate functional organization of biological networks. Enzymes in the yeast metabolic network were classified according to their indispensability under different growth conditions, as quantified by a fitness score [44]. This analysis defined the fitness effect of gene knockouts as a ratio of maximal growth rate of a mutant to that of a wild-type calculated by MOMA. Epistatic interactions in yeast metabolism were also studied by computing maximal growth rates of knockouts [45]. An epistatic interaction parameter based on the fitness score could characterize interactions between all possible gene pairs and identify functional modules in yeast metabolism.

Predictive capability

Prediction of gene deletion strategies for overproduction of a desired metabolite has become practicable with the recent extension of FBA called OptKnock [38]. It solves for the optimal flux distribution that simultaneously optimizes two objective functions, biomass growth and secretion of a target metabolite using a bi-level optimization technique.

This genome-scale mutant design strategy was experimentally validated recently with lactic acid production in *E. coli* [46, 47]. Implementation of three computed gene deletion strategies showed increases in lactic acid secretion ranging from 25 to 73%. These secretion rates were directly coupled to biomass growth rate such that an increase in growth rate results in an increase in the lactic acid secretion, which validated the simultaneous optimization strategy of OptKnock.

INTERFACING METABOLOMICS WITH FBA

FBA has appeared as an effective means to analyse large-scale biological networks and make experimentally testable predictions regarding cellular behaviours. The central challenge of metabolomics is its lack of *comprehensiveness* and *accuracy* in measurements. For example, GC/MS technology is difficult to apply to metabolites like cofactors,

organic disphosphates, or metabolites larger than trisaccharides, because GC can only be used for volatile compounds [48, 49]. Despite ongoing developments of numerous separation and analytic techniques, selectivity and sensitivity of measurements become major concerns for large network systems where simultaneous analysis of hundreds of metabolites is required [48, 50]. The combination of metabolomic data with FBA holds promise for addressing some of these challenges.

The limitations of metabolomics can effectively be complemented by an FBA-based formalism (Figure 3). The use of genome-scale reconstructions allows for the incorporation of partial metabolic information. Measured fluxes or flux ratios can be imposed as equality constraints. In addition, the dynamic profile of concentrations can be used conveniently for estimation of flux information, i.e. $v = \mathrm{d}C/\mathrm{d}t$. The uncertainties associated with concentrations can be assessed by reflecting their confidence bounds [51, 52] in inequality constraints. Various mathematical analyses coupled with FBA such as sensitivity analysis [53,54] and objective function inference tool [37] can guide experiments further by identifying important metabolites. Undetected or important metabolites that require a different analytical method can thus be identified and bottlenecks in the analytical procedures can be alleviated effectively.

REMAINING CHALLENGES AND SYNERGY BETWEEN FBA AND METABOLOMICS

Objective function

The identification of a physiologically relevant objective function is a prerequisite for the successful application of FBA. Whereas microbial cells and other unicellular organisms have been well-described with the objective of maximizing growth rate, metabolic objectives in higher organisms including tissues and organs are not simple to postulate [55].

Metabolomic data can assist in identifying relevant objective functions with the bi-level optimization framework discussed in the previous section [37, 56]. The MFA technique, which involves the uptake of an isotopomer (e.g. ^{13}C-labelled substrate) and the subsequent spectroscopy analysis, has been successfully applied to quantify intracellular flux distributions of various organisms including *E. coli* and *Penicillium chrysogenum* [15, 57–60]. These experimental measurements can be used for deducing a cellular objective function, and thus predicting cell behaviours under various conditions.

A stumbling block for embracing the bi-level optimization technique to infer objective functions is the large experimental effort that is required even for relatively small networks. Furthermore, it is not obvious that a universal objective function exists for describing all applicable cell behaviours. The incorporation of additional constraints was recently shown to alleviate measurement requirements [55]. However, it is evident that advances in metabolomics will accelerate hypotheses validation and the application of FBA to less-studied organisms.

Application to drug design

The identification of gene targets for generating desired phenotype changes or for discovering potential drug targets is a promising area that can be approached with FBA and metabolomic data. For example, drug targets for the tuberculosis-causing agent *Mycobacterium tuberculosis* were identified by systematically deleting genes *in silico* and evaluating the essentiality of associated proteins using FBA. The validity of these hypotheses was experimentally validated [61]. This strategy reduces the potential list of experimental tests required for drug design. Moreover, flexibility in dealing with constraints and scalability of FBA allowed for the identification of a set of core metabolic reactions as likely antimicrobial drug targets [62].

As FBA has proven to be effective in understanding the role of particular genes in altering metabolism, genome-scale metabolic reconstruction of medically relevant organisms is beginning to emerge [63]. However, suggested inhibitors are not guaranteed to block or enhance a desired activity or function of metabolism (for example, the inhibitor may not be able to enter a cell) [63]. Because metabolomics can characterize involved metabolites in an unbiased manner, integrating FBA with metabolomics can facilitate identifying elusive drug targets and validating suitability of suggested inhibitors.

CONCLUSIONS

Though several challenges remain, FBA has significant potential for embracing increasingly available yet non-comprehensive metabolomic data. Practical applications such as drug target designs are emerging. There have been many innovative developments to

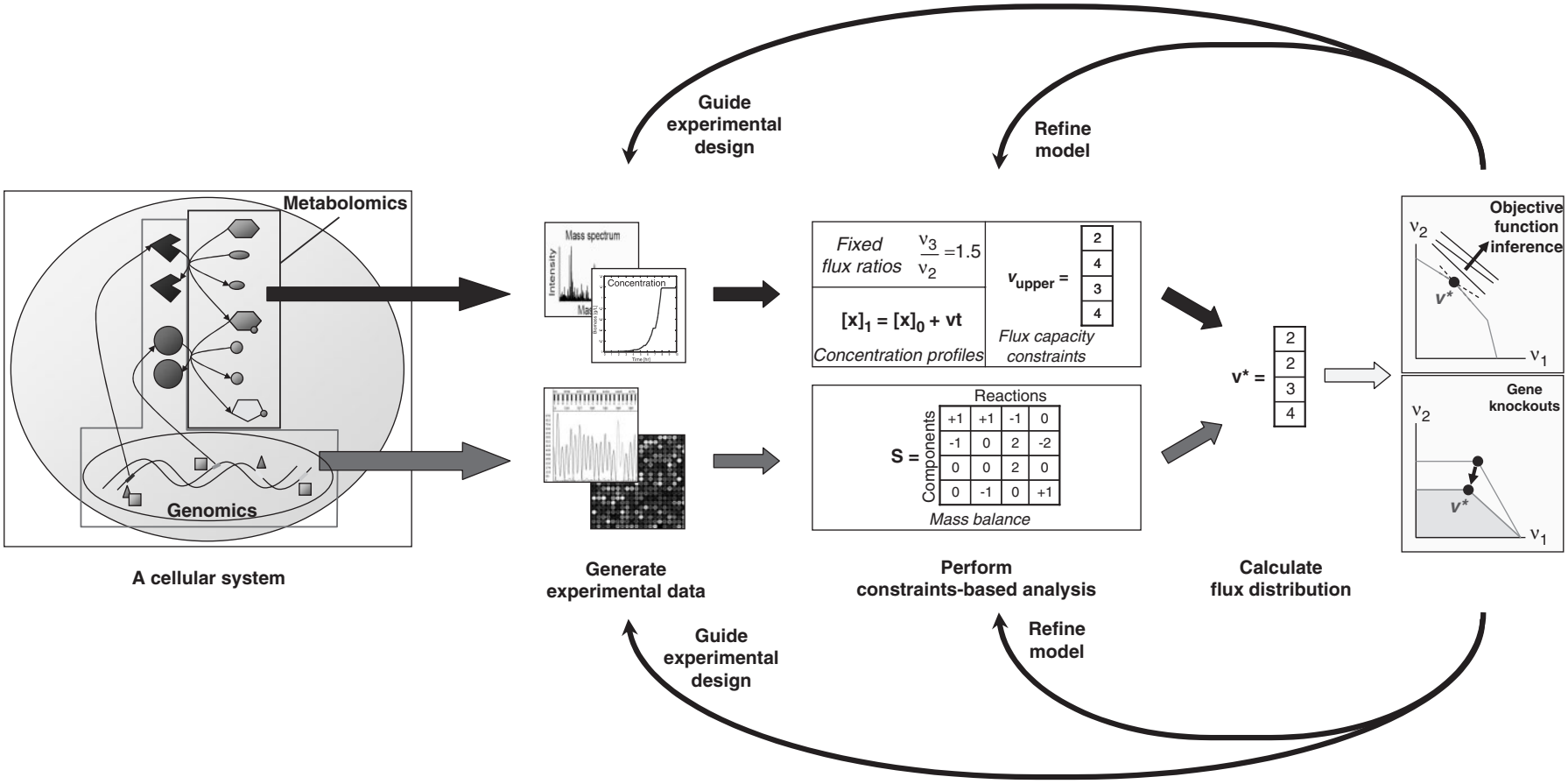

Figure 3: The combination of metabolomics with FBA. The intrinsic limitations of metabolomics, including specifically its lack of comprehensiveness and accuracy in measurements, can be effectively complemented by an FBA-based formalism. Metabolomics data, including the results of mass spectrometry as well as metabolite concentrations over time and other kinetic parameters, can complement FBA. The resultant flux vectors can be analysed for network properties, including objective function inference (i.e. identifying the objective function of a biochemical network, such as biomass production in central metabolism) and sensitivity analysis. Ultimately, the results of these analyses can drive further experimentation, as part of an iterative model refinement process that aims to generate new hypotheses and identify emergent properties of biological systems.

improve the predictive capabilities of FBA. Hence, the recent advances in metabolomics hold promise for synergistic outcomes through incorporating such data into the FBA framework.

Key Points

- Flux balance analysis (FBA) is an effective tool for the analysis of metabolic networks.
- Though the development of analytic techniques has facilitated the generation of dynamic profiles of metabolites, such data sets are not accurate enough for generating large-scale kinetic models.
- FBA can complement the uncertainty and incompleteness of metabolomic data, and thereby provide a better characterization of cellular phenotypes.
- Recent advances in FBA include the prediction of flux distributions of engineered cells, investigation of a cellular objective and the design of a mutant strain with desired properties.

Acknowledgement
We thank Dr Daniel Segrè for his helpful comments on the manuscript.

References

1. German JB, Watkins SM, Fay LB. Metabolomics in practice: emerging knowledge to guide future dietetic advice toward individualized health. *J Am Diet Assoc* 2005; **105**:1425–32.
2. Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 2005;**10**:1551–58.
3. Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* 2003;**54**:669–89.
4. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 2002;**48**:155–71.
5. Smedsgaard J, Nielsen J. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J Exp Bot* 2005;**56**:273–86.
6. Fiehn O, Kopka J, Dormann P, *et al*. Metabolite profiling for plant functional genomics. *Nat Biotechnol* 2000;**18**:1157–61.
7. Roessner U, Wagner C, Kopka J, *et al*. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J* 2000;**23**: 131–42.
8. Jansen JJ, Hoefsloot HC, Boelens HF, *et al*. Analysis of longitudinal metabolomics data. *Bioinformatics* 2004;**20**: 2438–46.
9. Salas A, Yao YG, Macaulay V, *et al*. A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med* 2005;**2**:e296.
10. Zanssen S, Schon EA. Mitochondrial DNA mutations in cancer. *PLoS Med* 2005;**2**:e401.
11. German JB, Hammock BD, Watkins SM. Metabolomics: building on a centruy of biochemistry to guide human health. *Metabolomics* 2005;**1**:3–9.
12. Mendes P, Camacho D, de la Fuente A. Modelling and simulation for metabolomics data analysis. *Biochem Soc Trans* 2005;**33**:1427–9.
13. Kell D, Mendes P. Snapshots of systems: metabolic control analysis and biotechnology in the postgenomic era. In: Cornish-Bowden AJ, Cardenas ML, (eds). *Technical and Medical Implications of Metabolic Control Analysis*. Amsterdam: Kluwer, 2000;2–25.
14. Mendes P, Sha W, Ye K. Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics* 2003;**19**:122–9.
15. Christensen B, Nielsen J. Metabolic network analysis. A powerful tool in metabolic engineering. *Adv Biochem Eng Biotechnol* 2000;**66**:209–31.
16. Fischer E, Sauer U. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat Genet* 2005;**37**:636–40.
17. Teusink B, Passarge J, Reijenga CA, *et al*. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* 2000;**267**:5313–29.
18. Beard DA, Liang SD, Qian H. Energy balance for analysis of complex metabolic networks. *Biophys J* 2002;**83**:79–86.
19. Covert MW, Famili I, Palsson BO. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol Bioeng* 2003;**84**: 763–72.
20. Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2004;**2**:886–97.
21. Papin JA, Palsson BO. The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. *Biophys J* 2004;**87**:37–46.
22. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 2002;**99**:15112–7.
23. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol* 2003;**14**:491–6.
24. Forster J, Famili I, Fu P, *et al*. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 2003;**13**:244–53.
25. 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**:3724–31.
26. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 2000;**97**:5528–33.
27. Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002;**277**:28058–64.
28. Schilling CH, Covert MW, Famili I, *et al*. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol* 2002; **184**:4582–93.
29. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* 2003;**221**:309–25.

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

31. Beard DA, Babson E, Curtis E, Qian H. Thermodynamic constraints for biochemical networks. *J Theor Biol* 2004;**228**: 327–33.

32. Maskow T, von Stockar U. How reliable are thermodynamic feasibility statements of biochemical pathways? *Biotechnol Bioeng* 2005;**92**:223–30.

33. 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**:R695–704.

34. Varma A, Boesch BW, Palsson BO. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol* 1993;**59**: 2465–73.

35. Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* 2005;**7**:155–64.

36. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* 2005;**102**:7695–700.

37. Burgard AP, Maranas CD. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 2003;**82**:670–77.

38. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;**84**:647–57.

39. Mahadevan R, Edwards JS, Doyle FJ,3rd. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002;**83**:1331–40.

40. Cuthrell JE, Biegler LT. On the optimization of differential-algebraic process systems. *Aiche Journal* 1987;**33**:1257–70.

41. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 2004;**14**:301–12.

42. Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep* 2002;**29**:233–6.

43. Lee DY, Fan LT, Park S, *et al*. Complementary identification of multiple flux distributions and multiple metabolic pathways. *Metab Eng* 2005;**7**:182–200.

44. Papp B, Pal C, Hurst LD. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 2004;**429**:661–64.

45. Segre D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. *Nat Genet* 2005;**37**:77–83.

46. Fong SS, Burgard AP, Herring CD, *et al*. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;**91**:643–48.

47. Fong SS, Joyce AR, Palsson BO. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 2005;**15**:1365–72.

48. Weckwerth W, Fiehn O. Can we discover novel pathways using metabolomic analysis? *Curr Opin Biotechnol* 2002;**13**: 156–60.

49. Kopka J, Fernie A, Weckwerth W, *et al*. Metabolite profiling in plant biology: platforms and destinations. *Genome Biol* 2004;**5**:109.

50. Fischer E, Sauer U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur J Biochem* 2003;**270**:880–91.

51. Landaw EM, DiStefano JJ,3rd. Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *Am J Physiol* 1984;**246**: R665–77.

52. Petersen B, Gernaey K, Vanrolleghem PA. Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Sci Technol* 2001;**43**: 347–55.

53. Pramanik J, Keasling JD. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering* 1997; **56**:398–21.

54. Pramanik J, Keasling JD. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol Bioeng* 1998;**60**:230–8.

55. Nolan RP, Fenley AP, Lee K. Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis. *Metab Eng* 2006;**8**:30–45.

56. Pharkya P, Burgard AP, Maranas CD. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol Bioeng* 2003;**84**: 887–99.

57. de Graaf AA. Metabolic flux analysis of Corynebacterium glutamicum. In: Schügerl K, Bellgardt KH, (eds). *Bioreaction Engineering: Modeling and Control*. Berlin-Heidelberg-New York: Springer, 2000.

58. Klapa MI, Stephanopoulos G. Metabolic flux analysis. In: Schügerl K, Bellgardt KH, (eds). *Bioreaction Engineering: Modeling and Control*. Berlin-Heidelberg-New York: Springer, 2000;106–22.

59. Szyperski T. 13C-NMR, MS and metabolic flux balancing in biotechnology research. *Q Rev Biophys* 1998; **31**:41–106.

60. Wiechert W. 13C metabolic flux analysis. *Metab Eng* 2001; **3**:195–206.

61. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* 2005;**1**:e46.

62. Almaas E, Oltvai ZN, Barabasi A-L. The activity reaction core and plasticity of metabolic networks. *PLoS Comput Biol* 2005;**1**:e68.

63. Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol* 2005;**5**:8.