Flux balance analysis: A geometric perspective

Smallbone, Kieran and Simeonidis, Evangelos

2009

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Flux balance analysis: A geometric perspective

Kieran Smallbone [a,b,*], Evangelos Simeonidis [a,c]

[a] Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, 131 Princess Street, Manchester M1 7DN, UK
[b] School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[c] School of Chemical Engineering and Analytical Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK

ARTICLE INFO

ABSTRACT

Advances in the field of bioinformatics have led to reconstruction of genome-scale networks for a number of key organisms. The application of physicochemical constraints to these stoichiometric networks allows researchers, through methods such as flux balance analysis, to highlight key sets of reactions necessary to achieve particular objectives. The key benefits of constraint-based analysis lie in the minimal knowledge required to infer systemic properties. However, network degeneracy leads to a large number of flux distributions that satisfy any objective; moreover, these distributions may be dominated by biologically irrelevant internal cycles. By examining the geometry underlying the problem, we define two methods for finding a unique solution within the space of all possible flux distributions; such a solution contains no internal cycles, and is representative of the space as a whole. The first method draws on typical geometric knowledge, but cannot be applied to large networks because of the high computational complexity of the problem. Thus a second method, an iteration of linear programs which scales easily to the genome scale, is defined. The algorithm is run on four recent genome-scale models, and unique flux solutions are found. The algorithm set out here will allow researchers in flux balance analysis to exchange typical solutions to their models in a reproducible format. Moreover, having found a single solution, statistical analyses such as correlations may be performed.

1. Introduction

Recent advances in genome sequencing techniques and bioinformatic analyses have led to an explosion of systems-wide biological data. In turn the reconstruction of genome-scale networks for micro-organisms has become possible. Whilst the first stoichiometric model of *E. coli* was limited to the central metabolic pathways (Varma and Palsson, 1993), the most recent reported model is much more comprehensive, consisting of 2077 reactions and 1039 metabolites (Feist et al., 2007). Reaction networks for *S. cerevisiae* have been similarly expanded through incorporation of more genes and their corresponding metabolites —a recent consensus model consists of 1761 reactions and 1168 metabolites (Herrgård et al., 2008). Genome-scale stoichiometric models for other micro-organisms (Kim et al., 2008) and even *H. sapiens* (Duarte et al., 2007) have been developed.

The ability to analyse, interpret and ultimately predict cellular behaviour has been a long sought-after goal. The genome sequencing projects are defining the molecular components

within the cell, and describing the integrated function of these molecular components will be a challenging task (Edwards and Palsson, 2000). Ideally, one would like to use kinetic modelling to characterize fully the mechanics of each enzymatic reaction, in terms of how changes in metabolite concentrations affect local reaction rates. However, a considerable amount of data is required to parameterize even a small mechanistic model; the determination of such parameters is costly and time-consuming, and moreover many may be difficult or impossible to determine experimentally. Instead, genome-scale metabolic modelling has relied on constraint-based analysis (Beard et al., 2002; Covert et al., 2003; Kim et al., 2008; Price et al., 2004), which uses physicochemical constraints such as mass balance, energy balance, and flux limitations to describe the potential behaviour of an organism. In particular, flux balance analysis (FBA) (Kauffman et al., 2003) highlights the most effective and efficient paths through the network in order to achieve a particular objective function, such as the maximization of biomass or ATP production. The key benefit of FBA and similar techniques lies in the minimal amount of biological knowledge and data required to make quantitative inferences about network behaviour (Bonarius et al., 1997).

In general there is degeneracy in stoichiometric networks, leading to an infinite number of flux distributions satisfying the

* Corresponding author at: Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, 131 Princess Street, Manchester M1 7DN, UK. Tel.: +44 161 306 5146; fax: +44 161 306 5201.

E-mail address: kieran.smallbone@manchester.ac.uk (K. Smallbone).

given optimality criteria. It is a great focus of the FBA community to reduce the size of this optimal flux space, through imposing tighter limits on each flux based, for instance, on measurements of intracellular fluxes with nuclear magnetic resonance, or other additional constraints (Kim et al., 2008). Despite use of these techniques, the resultant solution remains a *space* of fluxes, rather than a *unique* flux. In a recent paper, Price et al. (2004) do not view this as a problem, stating

> The mathematical notion of equivalent optimal states is coincident with the biological notion of silent phenotypes. This property distinguishes *in silico* modelling in biology from that in the physicochemical sciences where a single and unique solution is sought.

While it is true that equivalent solutions may be important in representing biological reality, at the same time we also believe that the ability to state a well-defined, single solution that is representative of the space of all possible fluxes would be of great benefit to the modelling community. From a practical perspective, researchers often do quote a single flux that results from their analysis; since this is chosen randomly from a large space of possible fluxes, their results are irreproducible and entirely dependent on the software or algorithm used to solve the linear programming (LP) problem. From a scientific perspective, the ability to extract a representative solution from the space would allow us to perform typical analyses, for example correlating flux with associated protein levels.

In this paper we show that, by examining the geometry underpinning the problem, we may find a well-defined, unique flux through application of a well-known and fundamental mathematical theorem. Unfortunately, this method proves impracticable for large, genome-scale models. Thus, we define a second method that gives similar results to the first, and moreover is computationally feasible. The algorithm is applied to a range of existing genome-scale models.

2. Methods

The problems encountered when performing constraint-based analysis, and the methods we propose to overcome these problems are best described through referral to a simple example. Consider the small metabolic network presented in Fig. 1, whose fluxes we wish to estimate. This may be addressed through appealing to FBA. This method allows us to identify the optimal path through the network in order to achieve a particular objective; quantitative predictions will then hold true if the cell optimizes its growth under the conditions considered. When applying LP to predict flux distributions it is assumed that the cell has found an "optimal solution" for survival through natural selection, where we equate survival with growth (Edwards and Palsson, 2000).

For our example network, FBA would involve maximizing output flux (v_5) subject to a limited nutrient consumption rate ($v_1 \leqslant 1$, say). The maximal solution to the problem here is clear
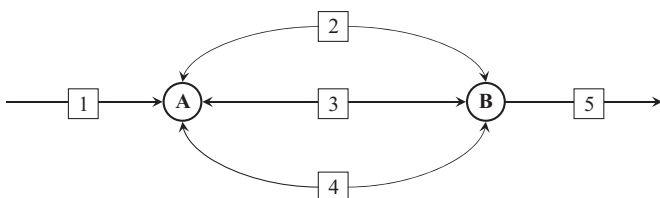

Fig. 1. A small metabolic network that admits multiple solutions: the output flux may be delivered through any of reactions 2, 3 or 4.

($v_5 = 1$), but such "back of the envelope" calculations would not be possible in a genome-scale network with thousands of reactions and species. Rather, we must formulate the problem in matrix form

$$\text{maximize} \quad Z = f^T v$$
$$\text{subject to} \quad N v = 0,$$
$$v^{\min} \leqslant v \leqslant v^{\max},$$

where

$$f = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad N^T = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ -1 & 1 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad v^{\min} = \begin{pmatrix} 0 \\ -\infty \\ -\infty \\ -\infty \\ 0 \end{pmatrix}, \quad v^{\max} = \begin{pmatrix} 1 \\ \infty \\ \infty \\ \infty \\ \infty \end{pmatrix}.$$

The FBA problem is now well-defined, in that it leads to a unique, finite objective value $Z = Z^\star = 1$; however, even for this small network there is degeneracy, leading to an infinite number of flux distributions v with the same optimal value. Returning to Fig. 1, we can see that this occurs because the flux may be delivered through any of reactions 2, 3 or 4, or any combination thereof. Indeed the actual non-unique flux solution found will be entirely dependent on the underlying algorithm of one's software. For example, one possible solution is given by $v^\star = (1, 1, 1000, -1000, 1)^T$. Such an answer is unappealing—the result is dominated by an internal cycle between fluxes 3 and 4; whilst cells are known to demonstrate such profligacy with regard to flux (Westerhoff et al., 1983), it is unlikely to occur to such an extent. The problem we shall address through the remainder of this paper is how to extract a solution from the space of flux distributions that is both unique (and hence reproducible) and sensible, from a biological perspective.

2.1. Method one

The set of all possible solutions to the FBA problem is given by the equation and constraints

$$A v = b, \quad v^{\min} \leqslant v \leqslant v^{\max}, \tag{1}$$

where

$$A = \begin{pmatrix} N \\ f^T \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ Z^\star \end{pmatrix}.$$

From a geometric perspective, Eq. (1) defines a *polyhedron*; appealing further to geometric understanding (and using the fact that $b \neq 0$), the representation theorem (Minkowski, 1910) tells us that polyhedra such as these may be naturally decomposed as the sum of a *convex hull* and a *pointed cone* as shown in Fig. 2.

$$v \in \underbrace{\left\{ \sum_i \lambda_i x_i : \lambda_i \geqslant 0, \sum_i \lambda_i = 1 \right\}}_{\text{convex hull}} + \underbrace{\left\{ \sum_j \mu_j y_j : \mu_j \geqslant 0 \right\}}_{\text{pointed cone}}.$$

The hull is uniquely defined by its *vertices* x_i while the cone is uniquely (up to multiplication) defined by its *extremal rays* y_j. These vertices and rays may be found through application of Chernikova's algorithm (Chernikova, 1965; Le Verge, 1992).
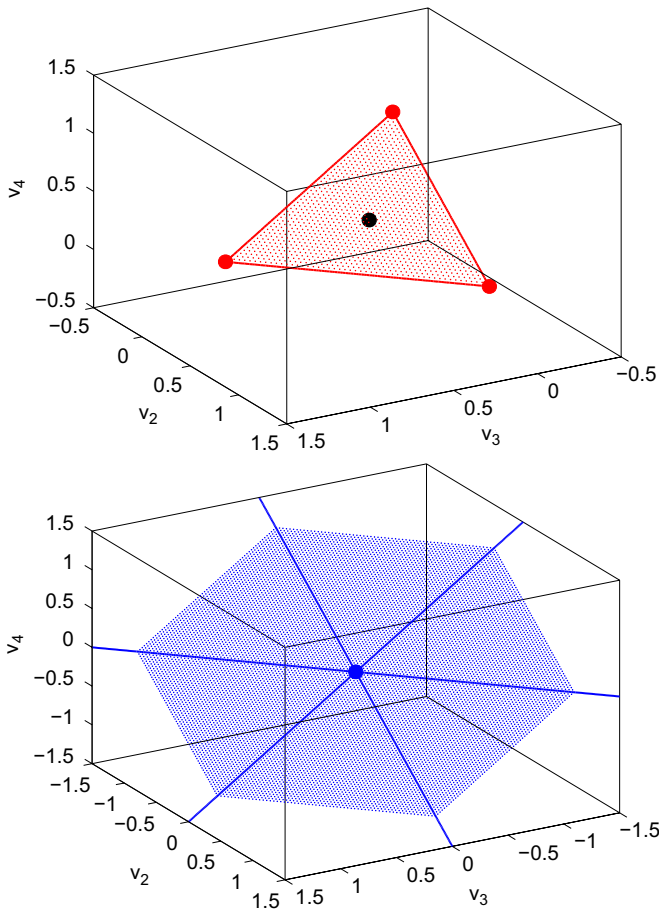
Fig. 2. The set of all optimal solutions of a FBA problem may be decomposed as the sum of the convex hull of its vertices (red, top) and the pointed cone of its extreme rays (blue, bottom). A natural choice of solution is given by the centre of the hull (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For our example we have three vertices and six rays given by

$$x_i = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad y_j = \begin{pmatrix} 0 \\ \pm 1 \\ \mp 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \pm 1 \\ 0 \\ \mp 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \pm 1 \\ \mp 1 \\ 0 \end{pmatrix}.$$

It is clear from this result that the extremal rays define the internal cycles in the network. Rather, the interesting potential solutions are to be found within the hull. Since the vertices of the hull are uniquely defined, so the *centre* of the solution hull is uniquely defined as

$$v^\star = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{pmatrix} 1 \\ 1/3 \\ 1/3 \\ 1/3 \\ 1 \end{pmatrix}.$$

This is indeed a natural solution to our FBA problem. No flux is wasted through internal cycles. Moreover, in the absence of experimental information as to which of reactions 2, 3 or 4 are used preferentially, the result unbiassedly assumes an equal flux through each.

Choosing the centre of the convex hull seems an ideal method for fixing a unique FBA solution. Unfortunately, Chernikova's algorithm is exponential in the number of reactions (Le Verge,

1992), and hence is computationally infeasible for genome-scale systems; indeed, the related problem of determining elementary flux modes has only proved possible in systems with tens, rather than thousands, of species and reactions (Klamt and Stelling, 2002). Thus we turn to an alternative, but related, method for dealing with such large systems.

2.2. Method two

The algorithm we propose to define a unique flux balance solution may be visualized as follows: whilst it is not computationally possible to define the vertices of a convex hull for a genome-scale network, the problem of finding the smallest box that may contain that hull may be posed as a (soluble) set of LP problems.

We first define the method, then explain its workings by means of application to the example of Fig. 1. Starting with Eq. (1), the algorithm may be stated in *pseudocode* as:

$n := 1$
$\mu_1 := (0, \ldots, 0)$
while $\Delta > \varepsilon$ **do**
   $Z_n :=$ minimize $\|v - \mu_n\|_1$ subject to
      $Av = b, \quad v^{\min} \leqslant v \leqslant v^{\max}, \quad \|v - \mu_i\|_1 = Z_i, \quad i = 1, \ldots, n-1$
   for each j **do**
      $v_j^{\max} \hookleftarrow$ maximize v_j subject to
         $Av = b, \quad v^{\min} \leqslant v \leqslant v^{\max}, \quad \|v - \mu_i\|_1 = Z_i, \quad i = 1, \ldots, n$
      $v_j^{\min} \hookleftarrow$ minimize v_j subject to
         $Av = b, \quad v^{\min} \leqslant v \leqslant v^{\max}, \quad \|v - \mu_i\|_1 = Z_i, \quad i = 1, \ldots, n$
$n \hookleftarrow n + 1$
$\mu_n := \frac{1}{2}(v^{\max} - v^{\min})$
$\Delta \hookleftarrow \max_i |v_i^{\max} - v_i^{\min}|$

where $\|v\|_1 = \sum_i |v_i|$ denotes the taxicab—norm and ε is some specified tolerance.

Let us examine the workings of the algorithm through its application to our example.

2.2.1. Iteration 1

For the first iteration we make the sensible assumption that the cell will minimize the total flux ($\sum_i |v_i|$) required to produce its objective $Z = Z^\star$, which by decomposing v into its positive and negative parts may be posed as a LP problem. This idea for flux minimization was first suggested by Holzhütter (2004); assuming that flux is proportional to enzyme concentration, this is equivalent in some sense to the cell minimizing the amount of enzyme required. The step further ensures that internal cycles that can produce fluxes $v_i = \infty$ are removed, a result also found in previous methods (Price et al., 2006; Qian and Beard, 2005). Indeed, returning to Fig. 1, the solution space to this new problem is now given exactly by the convex hull, and the cone is empty. Having found the minimal total flux through the network $Z_1 = 3$, we now find the bounds on each flux subject to this additional constraint (a new take on flux variability analysis, Mahadevan and Schilling, 2003). Naturally, $v_i = 1$, for $i = 1, 5$, whilst $0 \leqslant v_i \leqslant 1$ for $i = 2, 3, 4$. These limits are set as the new upper and lower bounds for the problem. The centre is set as the mean of these new bounds, namely $\mu_2 = (1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1)^T$, as this is most representative of all solutions.

2.2.2. Iteration 2

In the first method, we were able to choose the centre of the convex hull ($v^\star$) as a solution. Here we have placed a box around the hull (defining new bounds); unfortunately the centre of the box (μ_2) is not necessarily a solution to the original problem. Thus we instead minimize the distance between the flux and the centre, subject to the new constraint Z_1 found in the previous

Table 1

The results of the algorithm when applied to a selection of genome-scale models, with tolerance $\varepsilon = 10^{-6}$.

Model	Reactions	Species	Iteration 0		Iteration 1		Iteration n	
			Fixed	At zero	Fixed	At zero	n	At zero
Example	5	2	2	0	2	0	22	0
H. plyori iIT341 (Thiele et al., 2005)	554	562	0	0	328	130	3	272
M. tuberculosis iNJ661 (Jamshidi and Palsson, 2007)	1028	911	863	530	996	599	3	599
S. cerevisiae iND750 (Duarte et al., 2004)	1266	1177	1086	890	1240	978	2	978
E. coli iAF1260 (Feist et al., 2007)	2382	1972	2251	1897	2358	1967	13	1967

iteration. In turn this leads to new bounds $0 \leqslant \nu_i \leqslant \frac{1}{2}$, for $i = 2, 3, 4$ and corresponding centre $\mu_3 = (1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1)^T$.

2.2.3. Iteration n

Each iteration of the algorithm adds an additional constraint, as the flux is drawn towards the centre of the bounds. Hence, with each iteration, the box surrounding the allowable flux space shrinks in size. After a finite number of iterations the bounds will converge to a single solution, within a given tolerance; this process takes 22 iterations for our simple example, with tolerance $\varepsilon = 10^{-6}$. The final solution $\nu^\star = (1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1)^T$ coincides with the solution found by the first method.

3. Results and discussion

We run the algorithm presented in method two on genome-scale metabolic models for four organisms to test its efficacy. The four models—with known biomass representations and nutrient limitations—were imported to MATLAB® using libSBML (Bornstein et al., 2008) and solved with GLPK (GNU Linear Programming Kit) (Makhorin, 2001).

The results may be found in Table 1. The maximal biomass yield was found and flux variability analysis performed to find the number of non-fixed fluxes (presented as "iteration 0"). The algorithm was then run until convergence at our specified tolerance $\varepsilon = 10^{-6}$. The first point of note is that, in most cases, the algorithm converges very quickly to a solution; three of the four networks converge in less than three iterations. Indeed, the motif of our small example network proves to be a worst-case scenario.

Let us consider the application of the algorithm to the *S. cerevisiae* network iND750 (Duarte et al., 2004), consisting of 1266 reactions. The assumption that the network maximizes its biomass production forces 1086 of the reactions to have a fixed value. Two iterations further and we have defined the unique *central* point; 978 of the reactions are fixed at zero, leaving only $1266 - 978 = 288$ taking a non-zero value. These reactions may be considered those critical for cellular growth and greatly simplify the problem of investigating metabolic function. Similar benefits are found when the method is applied to other models.

Our flux solution for the *S. cerevisiae* network is compared to the flux given in the original paper and is presented in Fig. 3. Whilst the vast majority of fluxes are identical, this is due to the assumption of maximal biomass production; indeed 108 of the $1266 - 1086 = 180$ fluxes not forced to have a specific value are found to be different. Moreover, some of these differences are large—of the same order as the cell's glucose consumption rate $(1 \, \mathrm{mmol} \, \mathrm{g}^{-1} \, \mathrm{h}^{-1})$. These differences illustrate the wide solution space typically found in FBA.

Where available, each reaction in the *S. cerevisiae* network is associated with a protein or proteins. Thus we may compare our flux solution with the protein copy number as given by
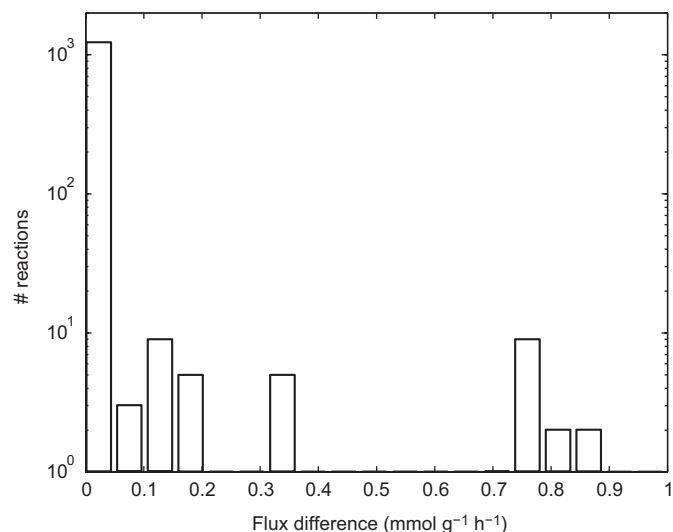

Fig. 3. Our flux solution for the *S. cerevisiae* network iND750 (Duarte et al., 2004) is compared to the flux presented in the original paper. 108 of a possible 180 fluxes were found to differ. Note that glucose consumption is defined as $1 \, \mathrm{mmol} \, \mathrm{g}^{-1} \, \mathrm{h}^{-1}$.

Ghaemmaghami et al. (2003). Where a reaction is associated with multiple proteins, we sum their copy numbers. The correlation between flux and protein levels is small (0.13), but significant ($p < 10^{-3}$), and compares slightly favourably to the flux given in the original paper (Duarte et al., 2004, correlation 0.12). We admit this correlation is not evidence of the correctness of our flux solution—such can only be shown with direct measurement—but do believe it to show the flux to be reasonable.

4. Conclusion

The algorithm presented in this paper allows researchers from the flux balance community to choose a unique and well-defined flux from the space of all possible solutions. In turn, any results produced with the use of this algorithm will be fully reproducible and allow the exchange of typical solutions to models in a reproducible format.

From a biological perspective, the exact flux utilized by the cell will be dependent on a wide range of stimuli, and thus impossible to predict from network structure alone. Nonetheless, the solution chosen by our algorithm is sensible in a number of respects. The algorithm picks out the minimal solution required to satisfy the given objective; assuming that flux correlates with enzyme levels, this is equivalent to the cell minimizing the amount of enzyme required to satisfy this objective. Moreover, any fluxes representing thermodynamically infeasible internal cycles are removed. From the remaining set of solutions, our chosen flux is in

a sense "central", and hence can be considered unbiassedly representative of all possible solutions.

Whilst a similar result may be obtained through application of Chernikova's algorithm, our method easily scales to networks with thousands of reactions and reactants. Thus, as network size increases and objectives gain in complexity, the method will continue to find application within the community.

Acknowledgements

References

Beard, D.A., Liang, S.C., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. Biophys. J. 83, 79–86.

Bonarius, H.P.J., Schmid, G., Trampe, J., 1997. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends Biotechnol. 15, 308–314.

Bornstein, B.J., Keating, S.M., Jouraku, A., Hucka, M., 2008. LibSBML: an API library for SBML. Bioinformatics 24, 880–881.

Chernikova, N.V., 1965. An algorithm for finding a general formula for the non-negative solutions of linear inequalities. USSR Comput. Math. Math. 5, 228–233.

Covert, M.W., Famili, I., Palsson, B.Ø., 2003. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? Biotechnol. Bioeng. 84, 763–772.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.Ø., 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. USA 104, 1777–1782.

Duarte, N.C., Herrgård, M.J., Palsson, B.Ø., 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.

Edwards, J.S., Palsson, B.Ø., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. USA 97, 5528–5533.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.

Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kırdar, B., Penttilä, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S.G., Mendes, P., Nielsen, J., Kell, D.B., 2008. A consensus yeast metabolic network obtained from a community approach to systems biology. Nat. Biotech. 26, 1155–1160.

Holzhütter, H., 2004. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur. J. Biochem. 271, 2905–2922.

Jamshidi, N., Palsson, B.Ø., 2007. Investigating the metabolic capabilities of mycobacterium tuberculosis H37Rv using the in silico strain inj661 and proposing alternative drug targets. BMC Syst. Biol. 1, 26.

Kauffman, K.J., Prakash, P., Edwards, J.S., 2003. Advances in flux balance analysis. Curr. Opin. Biotechnol. 14, 491–496.

Kim, H.U., Kim, T.Y., Lee, S.Y., 2008. Metabolic flux analysis and metabolic engineering of microorganisms. Mol. Biosyst. 4, 113–120.

Klamt, S., Stelling, J., 2002. Combinatorial complexity of pathway analysis in metabolic networks. Mol. Biol. Rep. 29, 233–236.

Le Verge, H., 1992. A note on Chernikova's algorithm. Technical Report 635, IRISA.

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276.

Makhorin, A., 2001. GNU linear programming kit, Moscow Aviation Institute, Moscow, Russia.

Minkowski, H., 1910. Geometrie der zahlen. B.G. Teubner, Leipzig.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2, 886–897.

Price, N.D., Thiele, I., Palsson, B.Ø., 2006. Candidate states of Helicobacter pylori's genome-scale metabolic network upon application of "loop law" thermodynamic constraints. Biophys. J. 90, 3919–3928.

Qian, H., Beard, D.A., 2005. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. Biophys. Chem. 114, 213–220.

Thiele, I., Vo, T.D., Price, N.D., Palsson, B.Ø., 2005. An expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico senome-scale characterization of single- and double-deletion mutants. J. Bacteriol. 187, 5818–5830.

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

Westerhoff, H.V., Helliingwerf, K.J., van Dam, K., 1983. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. Proc. Natl. Acad. Sci. USA 80, 305–309.