

# Finding elementary flux modes in metabolic networks based on flux balance analysis and flux coupling analysis: application to the analysis of *Escherichia coli* metabolism

Shayan Tabe-Bordbar · Sayed-Amir Marashi

Received: 29 April 2013 / Accepted: 19 July 2013 / Published online: 28 September 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Elementary modes (EMs) are steady-state metabolic flux vectors with minimal set of active reactions. Each EM corresponds to a metabolic pathway. Therefore, studying EMs is helpful for analyzing the production of biotechnologically important metabolites. However, memory requirements for computing EMs may hamper their applicability as, in most genome-scale metabolic models, no EM can be computed due to running out of memory. In this study, we present a method for computing randomly sampled EMs. In this approach, a network reduction algorithm is used for EM computation, which is based on flux balance-based methods. We show that this approach can be used to recover the EMs in the medium- and genome-scale metabolic network models, while the EMs are sampled in an unbiased way. The applicability of such results is shown by computing “estimated” control-effective flux values in *Escherichia coli* metabolic network.

**Keywords** Constraint-based modeling · Elementary flux modes · Flux balancing analysis · Flux coupling analysis · Metabolic networks · Metabolic pathways · Random sampling

## Introduction

Genome-scale metabolic models are now widely used in biotechnology and bioengineering (Zomorodi et al. 2012). Constraint-based analysis of these models provide a mathematical framework for understanding biotransformations in metabolism and possibly rational design of strains (Lewis et al. 2012; Price et al. 2003).

Elementary flux modes (EMs) are defined as the minimal sets of reactions in a metabolic network that can be active in steady-state conditions (Schuster et al. 2000). Therefore, each EM can be interpreted as a metabolic “pathway”. EMs are applicable to various fields of metabolic network studies, e.g., analysis of the capabilities of cellular metabolism (Çakir et al. 2007; Schilling and Palsson 2000; Schuster et al. 1999; Stelling et al. 2002), determination of minimal medium requirements (Schilling and Palsson 2000), measuring the flexibility of metabolic networks (Çakir et al. 2007; Papin et al. 2002), identification of pathways with optimal yield (Hädicke and Klamt 2010; Trinh et al. 2008) and determination the importance of a specific reaction (Wlaschin et al. 2006).

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s10529-013-1328-x) contains supplementary material, which is available to authorized users.

---

S. Tabe-Bordbar · S.-A. Marashi (✉)  
Department of Biotechnology, College of Science,  
University of Tehran, Tehran, Iran  
e-mail: marashi@ut.ac.ir

S. Tabe-Bordbar  
e-mail: shayan.bordbar@ut.ac.ir

Current approaches for calculating EMs generally use an algorithm called the double description method (Fukuda and Prodon 1996; Gagneur and Klamt 2004). In the double description method, a set of linear constraints are given as the input. In case of metabolic networks, these constraints are the stoichiometric and the reversibility constraints (Marashi 2011). The linear constraints, in general, characterize a polytope in space. The output of the algorithm is the set of extreme rays of the polytope. For metabolic networks, the set of computed extreme rays is equivalent to the set of elementary modes.

Despite recent developments in EM calculation (Terzer and Stelling 2008), these methods are unable to calculate EMs in genome-scale metabolic networks (due to running out of memory). This shortcoming is a consequence of the combinatorial explosion during the enumeration of the full set of EMs in a genome-scale network. Many efforts have been made to overcome this issue including enumerating EMs with a specific objective or constraint (de Figueiredo et al. 2009), enumerating all achievable pathways through chosen reactions that satisfy the steady-state flux of the whole network (Kaleta et al. 2009), breaking the network down in modules (Schuster et al. 2002; Schwartz et al. 2007) and enumerating the extreme rays of the projected flux cone onto a lower-dimensional subspace (Marashi et al. 2012). A method for randomly computing a subset of the EMs has been reported (Machado et al. 2012).

In spite of the mentioned endeavors, enumeration of the full set of the EMs in genome-scale networks remains unsolved. In this study, we introduce a novel random sampling approach to calculate EMs that is comparable to the core reductive algorithm (Pál et al. 2006; Yizhak et al. 2011). Briefly, in order to compute a random elementary mode, including a certain objective reaction, in each iteration we delete a subset of reactions and then, using flux balance analysis (FBA), we check if the objective reaction can carry a non-zero flux in steady-state conditions. This step is repeated until no further reaction can be deleted. The remaining set of reactions represents a flux distribution with a minimal number of non-zero elements and, therefore, it characterizes an elementary mode. We also applied flux coupling analysis (FCA) to remove reactions rationally. The ultimate goal of this study is to present a random sampling of EMs in genome-scale networks (where most of the current methods are unable of computing EMs).

## Methods

### Metabolic network models and computing EMs

As a medium-scale metabolic network, we used a simplified core metabolic network model of *Escherichia coli*, which includes 60 (unblocked) reactions and 56 metabolites. The SBML file of this model is presented in Supplementary Data 1. In order to test the validity of our results for larger models, we used the genome-scale metabolic network of *E. coli* which includes 1,075 reactions and 761 metabolites (Reed et al. 2003). Where applicable, “efmtool” software was used to compute the full set of EMs in metabolic networks (Terzer and Stelling 2008).

### Reductive algorithm for computing EMs

We developed an algorithm for computing EMs based on computing minimal sets of reactions, possibly including an objective reaction, which can be active in steady-state conditions. Our algorithm consists of three main parts:

- (1) Preprocessing, including removing blocked reactions and merging fully coupled reactions;
- (2) Randomly deleting  $k$  reactions; and
- (3) Examining the effect of the previous deletion on the network by FBA. If flux through the objective reaction drops to zero, the deleted group of reactions will be restored to the network; otherwise the deletion is accepted.

Steps 2 and 3 are iterated until no further reaction can be deleted. (It should be noted that if no fixed objective reaction is given in the input, then one reaction is chosen randomly as the objective reaction.)

In the constraint-based modeling of metabolic networks (Marashi 2011), it is often assumed that the network functions in steady-state conditions.

In the first step, i.e. the preprocessing step, we applied linear programming (LP) to detect reactions carrying no flux in steady-state, i.e. the blocked reactions (Burgard et al. 2004). Additionally, metabolic networks often include reactions which always have a fixed flux ratio, these are called fully-coupled reactions (Burgard et al. 2004; Larhlimi et al. 2012). We also merge each set of fully-coupled reactions into one “cumulative reaction”. For this purpose, we apply the function “removeTrivialFC” of the F2C2 tool

(Larhlimi et al. 2012). The preprocessing step helps to reduce the problem size, and consequently, lessens the calculation time.

In the second step, we remove reactions by a  $k$ -elimination scheme. Overall, the multiple deletion strategy gives us the advantage to solve fewer LPs. The number of reactions to be deleted in each iteration can be estimated by finding the largest integer number  $k$ , such that:

$$(1 - m/a)^k > 0.5 \quad (1)$$

where  $k$  is the number of reactions to be deleted in each iteration,  $m$  is the mean length of EMs and  $a$  is the total number of reactions that are currently present in the network.

Inequality (1) sets the deletion acceptance probability of a group of  $k$  reactions, greater than the probability of their rejection in third step. Applying inequality (1) after each deletion updates the value of  $k$  during gradual decrease in total number of reactions. In practice, we remove  $k + n$  reactions in each iteration, where  $n \geq 0$  is kept fixed during the algorithm. This scheme will be referred to as the  $(k + n)$ -elimination scheme. Therefore,  $k$ -elimination scheme is a special case of  $(k + n)$ -elimination scheme, where  $n = 0$ . In order to choose the best value, we ran the program several times and tested different values of  $n$ .

The set of reactions to be deleted can be chosen rationally. We utilize FCA for this purpose. F2C2 tool (Larhlimi et al. 2012) is applied to determine flux coupling relations between reactions. As an example, suppose that Fig. 1 represents the flux coupling graph of some metabolic network (Burgard et al. 2004). In this graph, there is an arrow from reaction  $i$  to reaction  $j$  if zero flux through  $i$  results in zero flux through  $j$ , i.e., reactions  $j$  and  $i$  are directionally coupled (Larhlimi and Bockmayr 2006). Reactions at the top of this graph have deletion priority over other reactions. Once one reaction is removed, all its descendant reactions become blocked. Therefore, in order to reduce the running time of the algorithm, only other reactions should be checked for removal. For example, in Fig. 1, once we remove reaction 2, in the next step we may remove one of the reactions 1, 3, 6 and 7.

For the third step of the algorithm, FBA in the COBRA toolbox v 2.0.3 (Schellenberger et al. 2011) was used to calculate the maximum flux through a

(random) objective reaction. If by deletion of  $k$  reactions in the previous step our objective reaction becomes inactive, then those  $k$  reactions are restored to the network, in order to perform the deletion and flux checking for each of the  $k$  reactions. Otherwise, the deletion is accepted. All the computations are performed on a 64-bit Windows system with Intel Core i7 2.00 GHz processor.

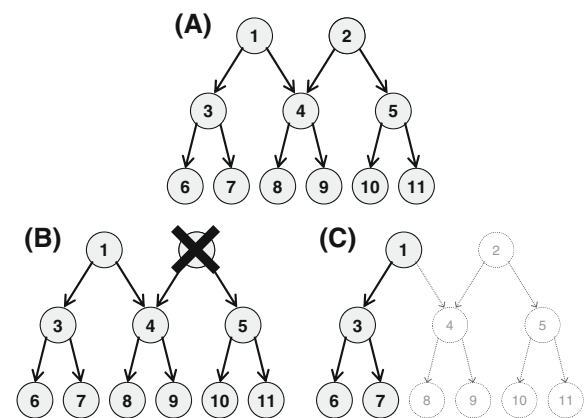
### Control effective flux (CEF) calculations

We used Stelling's algorithm to calculate CEFs (Stelling et al. 2002). ATP synthesis (by ATP synthase) and bio mass production reaction were used as our objective reactions.

## Results

### Randomly sampled EMs

In the first step, we applied our algorithm to the simplified *E. coli* core model. In this network, "efmtool" (Terzer and Stelling 2008) finds 1,164



**Fig. 1** **a** Flux coupling graph of an imaginary metabolic network. Each reaction is shown as a node of this graph. An *arrow* from reaction  $i$  to reaction  $j$  means that zero flux through  $i$  results in zero flux through  $j$ . **b** For rational deletion of reactions from the network, flux coupling graph is used. For example, suppose that reactions 2 is deleted in the first step; **c** Then, in the flux coupling graph, all reactions below a deleted reaction will be deleted automatically. In this example, since reaction 2 is deleted in the previous step, reactions 4, 5, 8, 9, 10 and 11 will be blocked. Therefore, the next reaction to be deleted should be chosen only from the remaining set of reactions, i.e., 1, 3, 6 and 7

EMs. We used our algorithm to compute EMs in this network. The number of unique EMs as a function of computation time is shown in Fig. 2.

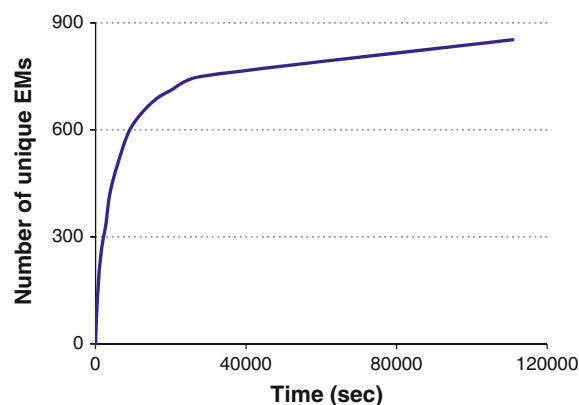
More than half of the EMs are found in the first 2.5 h of running the program, while 750 EMs (i.e., about two-third of all EMs) can be computed in  $\sim 9$  h 30 min. In this figure, as we go further in time, the EM finding rate slows down. Since our algorithm randomly finds an EM in each iteration, it is expected that at the beginning only unique EMs are detected, while the probability of computing a previously-found EM increases in the next iterations.

To examine whether the set of sampled EMs is an unbiased representative of the full set of EMs, we compared the lengths of the 750 calculated EMs with the length distribution of the full set. Figure 3 demonstrates that the two distributions are qualitatively similar, showing the same bimodal distributions.

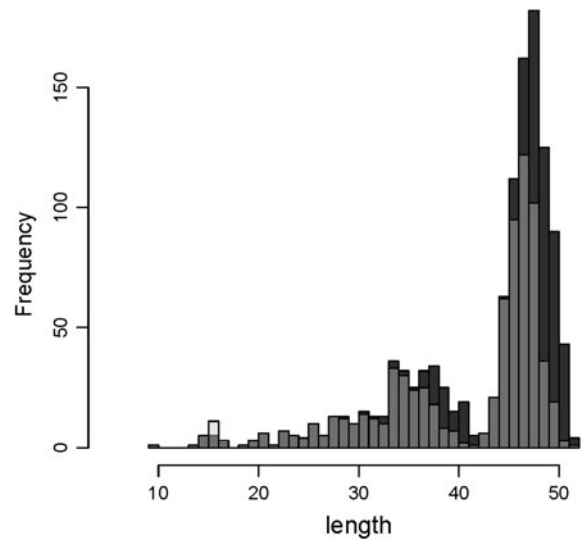
#### “Estimated” CEF values

One of the important applications of EMs is in the calculation of CEFs. The CEF value of a reaction is a measure of its efficiency in the metabolic network (Stelling et al. 2002). The CEF concept has been previously applied in biotechnology and bioengineering studies (Çakir et al. 2004, 2007; Zhang et al. 2010). Therefore, we decided to compare “actual” and “estimated” CEFs to examine the relevance of our results.

Figure 4 shows the difference between actual and estimated CEF values. One can observe that the difference values are generally very close to zero, which is reflected in the sharp peak in the center of the



**Fig. 2** Number of unique EMs found by our algorithm during time



**Fig. 3** Distribution of EM lengths. The full set of EMs (dark gray) and a set of 726 sampled EMs computed by our algorithm (light gray) are shown

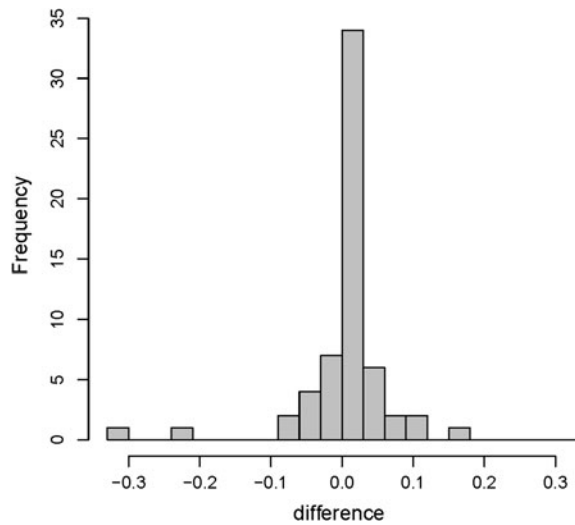
histogram (slightly greater than zero). The average was not found to be significantly different from zero ( $p$  value = 0.74; one-sample  $t$  test). The CEF values of 57 out of 60 reactions lie between  $-0.088$  and  $+0.112$ . Hence, we suggest that estimated CEF can be considered as a good approximation of actual CEF.

#### Computing EMs in genome-scale metabolic network of *E. coli*

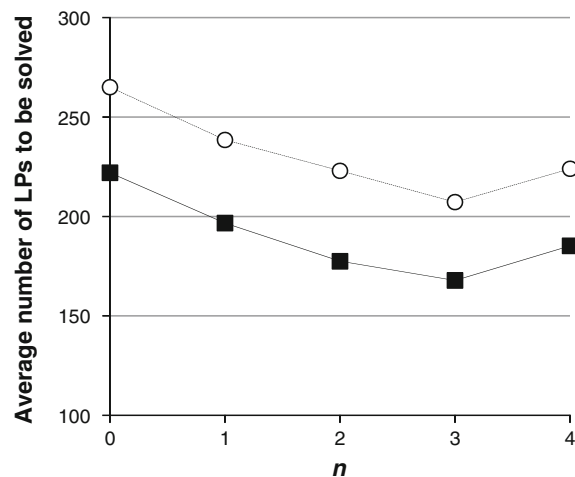
We applied our algorithm to sample EMs of the *E. coli* genome-scale metabolic network (iJR904) consisting of 1,075 reactions. At the beginning we removed blocked reactions, and additionally, the fully coupled reactions were merged. The resulting network included 582 reactions.

Instead of solving 582 LPs to compute each EM, we used the  $k$ -elimination scheme, resulting in an average of 265 LPs for computing each EM (which takes  $\sim 9.8$  s on average). We also tested the  $(k + n)$ -elimination scheme, for fixed  $n$  values ranging from 0 to 4. The results are presented in Fig. 5. The optimal value of  $n$  was empirically found to be 3, which results in  $\sim 207$  LPs to be solved for finding each EM, which takes  $\sim 8$  s on average.

We used FCA to show the effect of rational reaction elimination on calculation time of our algorithm. As illustrated in Fig. 5, utilizing FCA significantly reduces



**Fig. 4** Distribution of differences between the “actual” and “estimated” CEF values. The estimated CEF values are based on the first 150 sampled EMs (averaged over 50 runs). The average of this distribution is 0.003 and its standard deviation is 0.067



**Fig. 5** Average number of LPs for computing an EM in the genome-scale metabolic network of *E. coli* as a function of  $n$ . The  $(k + n)$ -elimination scheme is used for  $n = 0, 1, \dots, 4$ . The values are shown for computation with (filled square) and without (open circle) FCA-based rational reaction elimination

the number of LPs to be solved. In the  $(k + n)$ -elimination scheme, the minimum number of LPs required for computing each EM occurs in  $n = 3$ , where only 168 LPs should be solved on average to find an EM. With these conditions, each EM can be computed in  $\sim 6.7$  s. This is equivalent to 31 % decrease in computation time of each EM (compared

to the case of  $k$ -elimination scheme without applying rational elimination).

## Discussion

Each EM is a minimal set of reactions which can work together in steady-state conditions. In the present paper, in order to compute EMs we utilized a method inspired by the natural reductive evolution of metabolic networks. This approach is different from the previously reported method for random EM computation (Machado et al. 2012), which is based on double-description method. Moreover, we used flux coupling relations for rational selection of reactions to be deleted, which resulted in the speed up of EM computation procedure.

The most important benefit of our algorithm over current EM finding algorithms is the ability to calculate (at least a subset of) EMs in genome-scale metabolic networks. This can be considered as a great advantage, because current EM-finding algorithms cannot compute any EM in genome-scale metabolic models due to running out of memory (Marashi et al. 2012). Moreover, our algorithm can be run easily in parallel, due to its random sampling basis. Therefore, computers can independently compute EMs, which can be checked for redundancies afterwards.

Comparing the lengths of calculated EMs with the length distribution of full set of EMs indicates that sampling EMs is performed in an unbiased way, which can be seen as an improvement to the previous approaches for random sampling of EMs (Machado et al. 2012), which is noticeably biased towards computing short EMs. The unbiased sampling in our algorithm is the prerequisite for applicability of sampled EMs. We show that for certain applications, e.g. for computing CEFs, only a subset of (well-sampled) EMs may produce acceptable results.

One drawback of our current approach is that our implementation is slower than other software based on double-description method. Therefore, we believe that this strategy is still open for further optimizations.

## References

- Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 14:301–312

- Çakir T, Kirdar B, Ülgen KÖ (2004) Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. *Biotechnol Bioeng* 86: 251–260
- Çakir T, Kirdar B, Önsan ZI, Ülgen KÖ, Nielsen J (2007) Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. *BMC Syst Biol* 1:18
- de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 25:3158–3165
- Fukuda K, Prodon A (1996) Double description method revisited. In: Deza M et al (eds) *Combinatorics and computer science*, vol 1120. Springer, Berlin, Heidelberg, pp 91–111
- Gagneur J, Klamt S (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* 5:175
- Hädicke O, Klamt S (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J Biotechnol* 147:88–101
- Kaleta C, De Figueiredo LF, Schuster S (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res* 19:1872–1883
- Larhlmi A, Bockmayr A (2006) A new approach to flux coupling analysis of metabolic networks. *Lect Notes Comput Sci* 4216:205–215
- Larhlmi A, David L, Selbig J, Bockmayr A (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* 13:57
- Lewis NE, Nagarajan H, Palsson BØ (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Rev Microbiol* 10:291–305
- Machado D, Soons Z, Patil KR, Ferreira EC, Rocha I (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics* 28:i515–i521
- Marashi S-A (2011), *Constraint-based analysis of substructures of metabolic networks*, Ph.D. Thesis, Freie Universität Berlin, Berlin, Germany
- Marashi S-A, David L, Bockmayr A (2012) Analysis of metabolic subnetworks by flux cone projection. *Algorithms Mol Biol* 7:17
- Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440:667–670
- Papin JA, Price ND, Edwards JS, Palsson BØ (2002) The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J Theor Biol* 215:67–82
- Price ND, Papin JA, Schilling CH, Palsson BØ (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* 21:162–169
- Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (*iJR904* GSM/GPR). *Genome Biol* 4:R54
- Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6:1290–1307
- Schilling CH, Palsson BØ (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J Theor Biol* 203:249–283
- Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol* 17:53–60
- Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol* 18:326–332
- Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* 18:351–361
- Schwartz JM, Gaugain C, Nacher JC, de Daruvar A, Kanehisa M (2007) Observing metabolic functions at the genome scale. *Genome Biol* 8:R123
- Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420:190–193
- Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* 24:2229–2235
- Trinh CT, Unrean P, Sreenc F (2008) Minimal *Eschenchia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol* 74:3634–3643
- Wlaschin AP, Trinh CT, Carlson R, Sreenc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. *Metab Eng* 8:338–352
- Yizhak K, Tuller T, Papp B, Ruppin E (2011) Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol Syst Biol* 7:479
- Zhang Q, Wang W, Xiao H, Xiu Z (2010) Effect of oxygen level on efficiencies of metabolic fluxes in *Klebsiella pneumoniae*, 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), Chengdu, China, article number 5517846
- Zomorodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in metabolic networks. *Metab Eng* 14:672–686