

COMPARING FLUX BALANCE ANALYSIS TO NETWORK EXPANSION: PRODUCIBILITY, SUSTAINABILITY AND THE SCOPE OF COMPOUNDS

KAI KRUSE¹

kai.kruse@student.hu-berlin.de

OLIVER EBENHÖH^{2,3}

ebenhoe@mpimp-golm.mpg.de

¹*Theoretical Biophysics, Institute for Biology, Humboldt University, Berlin, Germany*

²*Max-Planck-Institute for Molecular Plant Physiology, Potsdam-Golm, Germany*

³*Institute for Biochemistry and Biology, University of Potsdam, Germany*

The producibility of metabolites from available resources is investigated systematically using flux balance analysis (FBA) and network expansion. Calculations are performed for the genome-scale metabolic networks of *Escherichia coli* and *Methanosarcina barkeri*. Strict biological interpretation of the results obtained with FBA leads to the concept of *sustainability*, which reduces the set of producible metabolites by assuming a growing and dividing cell. A systematic comparison showed that applying network expansion in many cases results in exactly the set of all sustainable metabolites. The purely heuristic approach of allowing for certain cofactors to facilitate reactions during the process of network expansion dramatically helps to improve agreement of the results from the two different approaches. In conclusion, we state that network expansion, due to its enormous advantages in computational speed, is a valuable alternative to determining producible metabolites with FBA.

Keywords: Flux balance analysis (FBA); network expansion; producibility; sustainability; scope; metabolism

1. Introduction

Considering the enormous speed at which sequencing projects are currently proceeding (at present, 769 fully sequenced genomes have been published and 2781 more are ongoing [11]), we have now access to the complete inventories of genes for a large number of organisms across all domains of life. In principle, from this information complete, genome-scale metabolic networks can be inferred by sequence comparison to genes or proteins which have been previously characterized. Genome-scale models are extremely useful for a wide variety of theoretical and computational analyses. Once the complete set of reactions is known, the powerful framework of flux balance analysis [10, 14] can be applied to predict optimal flux distributions maximizing the production of biomass or other, potentially exploitable, metabolites [8]. Further, it is possible to assess the effect of gene knock-outs and comparison of the computational predictions to experimentally measured fluxes can potentially point at erroneous or incomplete structures of the genome-scale network model [3].

Another strategy to analyze genome-scale networks is given by the method of network expansion [2, 7], which particularly aims at relating structural to functional features of large-scale metabolic networks. In this approach, networks of increasing size are constructed starting from an initial set of substrates (the seed) by step-wise adding all those reactions from the analyzed metabolic network which use as substrates only compounds provided by the seed or as products of reactions incorporated in earlier steps. The set of metabolites contained in the final network is called the scope of the seed and comprises all those metabolites which the network is capable of producing when only the seed compounds are initially available.

It can, however, be argued that a scope does not realistically describe the true biosynthetic capacity of an organism, because the idealized situation that exclusively some external compounds and no further internal metabolite is present does never occur under normal circumstances. To account for this fact, we have introduced for several practical applications [1, 6] a modified version of the expansion process which takes into account that it is unrealistic to assume that some key metabolites, the so-called cofactors, have to be synthesized *de novo* from the available nutrients. This purely heuristic approach is based on common biological knowledge and the fact that some cofactor pairs are participating in a considerable number of biochemical reactions. For example, a highly frequent pattern is the transfer of a phosphate group from ATP to some acceptor molecule resulting in the formation of ADP. Similarly, NAD^+ may accept electron pairs to yield the reduced form NADH, thus mediating redox reactions.

In this work, we systematically investigate the producibility of metabolites from available resources and compare the results from the two approaches. FBA provides a strict mathematical framework and can be used to assess whether a given metabolic network is capable of carrying a flux such that a particular metabolite can be produced. In contrast, the method of network expansion, especially in its modified form to allow for cofactor functionality, relies on heuristics to assess which metabolites may be synthesized from a given combination of nutrients. By comparing corresponding results obtained by the two approaches, we can show that in many cases the resulting sets are coinciding, which has practical consequences for an efficient, large-scale functional analysis of metabolic networks.

2. Theory

A metabolic network is commonly described by the stoichiometric matrix \mathbf{N} . A matrix with r rows and m columns describes a network in which r reactions connect m metabolites. An entry n_{ij} denotes the stoichiometric coefficient of metabolite i in reaction j , which is negative if the metabolite is consumed by the reaction, positive if it is produced and zero if there is no net production or consumption.

The metabolic state of a cell can be described by a vector $v \in \mathbb{R}^r$ containing the rates of all r biochemical reactions. These rates describe the velocities with which chemical conversions are performed and determine the temporal change of

the concentrations by

$$\frac{dC}{dt} = \mathbf{N}v, \quad (1)$$

where $C \in \mathbb{R}^m$ is the vector of the concentrations. Under physiological conditions, the fluxes v_j undergo certain limitations. Due to thermodynamic constraints, some reactions may only proceed in one direction, resulting in the constraint $v_j \geq 0$. An upper bound for the reaction rates may result from limited amount of free enzyme. In the following, we will only consider the former type of constraint since the latter will not play a role for our principle considerations.

For practical reasons, we will treat reversible reactions as two irreversible reactions proceeding in opposite directions. This is achieved by introducing an additional column to the matrix \mathbf{N} in which the signs of all stoichiometric coefficients are reversed. This leads to an increased number of reactions which all obey the same sign constraint $v_j \geq 0$.

Intuitively, a metabolite k is producible from a combination of nutrient metabolites if there exists a flux distribution such that only the nutrient metabolites are consumed, the metabolite k is produced and all other metabolites are at least not consumed. In this consideration, it is assumed that “side-products”, which are additionally produced, pose no problem to the organism and can be degraded or exported by other means. A mathematical description of producibility in the context of flux balance analysis has been given in [9]. If the available nutrients, or the seed, is denoted by $U \subset \{1 \dots m\}$, a metabolite k is producible if there exists a flux vector $v = (v_j)$ with $v_j \geq 0$ such that

$$[\mathbf{N}v]_k > 0 \text{ and } [\mathbf{N}v]_i \geq 0 \text{ for } i \notin U. \quad (2)$$

For the components $i \in U$, there is no restriction since these compounds may be imported from the environment. Condition (2) can be tested by phrasing it as a linear programming problem. Following the terminology introduced in [9], we call metabolites fulfilling this condition *producible* from the nutrients U . The entirety of all metabolites that are producible from U is denoted $P(U)$.

By definition, a network may carry a stationary flux leading to an increase in concentration of the producible metabolites while only the nutrients are consumed. However, this interpretation holds only as long as it is assumed that the cell is in a stationary, non-growing state. If a growing and reproducing organism is considered, stricter conditions for the producibility of metabolites must be imposed. In particular, all metabolites not contained in the set $P(U)$ are not producible and therefore their amount may not continuously increase. If we assume a persistent increase in cellular volume, the concentrations of such metabolites necessarily decrease and eventually reach zero and, as a consequence, are not available as substrates for other reactions. We take into account these considerations by repeating the calculation of all producible metabolites with the additional constraint that all those reactions are forbidden which use as substrate any metabolite that is not contained in $P(U)$. More precisely, all those metabolites are identified for which flux vectors $v = (v_j)$

exist with $v_j \geq 0$ and $v_l = 0$ if reaction l uses a substrate not contained in $P(U)$. A reaction l fulfils this condition if the set $\{i \notin P(U) | n_{il} < 0\}$ is non-empty. If this additional restriction results in a reduction of the set of producible compounds, the calculation is repeated with even stricter conditions. This process is iterated until the set of producible metabolites remains unchanged. The final set of metabolites is denoted by $S(U)$ and a metabolite within this set is termed *sustainable* since it has the property that it can be produced from the nutrients U even if the cell is constantly growing.

Sustainable metabolites are determined by repeatedly decreasing the set of producible metabolites until only those remain which can be produced from available nutrients without requiring the presence of any non-sustainable intermediates. In contrast, in the method of network expansion the scope of the seed U is determined by stepwise expanding a set of metabolites. Starting with the set U , all those reactions are identified that use exclusively substrates contained in the set and their products are included in the expanding set. Expansion stops if no further products are included and the final set is called the scope of the nutrients U , denoted $\Sigma(U)$. From the construction of the scope it is evident that every metabolite contained in the scope is also sustainable in the above defined sense. Therefore,

$$\Sigma(U) \subset S(U) \subset P(U). \quad (3)$$

The concepts of producibility, sustainability and scope can be viewed as different definitions of which metabolites can be synthesized by a given network with increasingly stricter conditions.

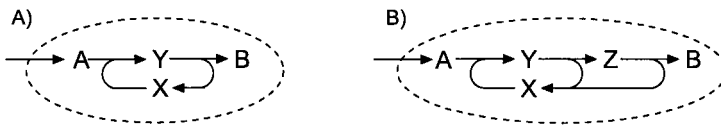


Fig. 1. Toy networks illustrating the differences of the presented concepts of producibility. A) A simple network producing biomass B from nutrient A in two consecutive reactions. Intermediates X and Y are essential for the production of biomass but neither is producible. This example demonstrates that a compound may be producible but not sustainable. B) A simple network producing biomass B from nutrient A in three consecutive reactions. All intermediates X , Y and Z and biomass B are sustainable. Since in an expansion starting from A , metabolite X is not available, the scope of A contains only A itself, $\Sigma(A) = \{A\}$. This example demonstrates that a compound may be sustainable but not included in the scope of the nutrients.

The difference between producible and sustainable metabolites is characterized in the toy network depicted in Fig. 1A. Here, clearly a steady state flux distribution exists such that metabolite B may be produced while only consuming nutrient A without a net consumption of X or Y . However, since the sum of X and Y is strictly balanced, it is not possible to produce either of these intermediates while consuming only nutrient A . Therefore, these metabolites are not producible. Imposing the constraint that no reactions may proceed which use one of those compounds as

substrates no longer allows for a production of B . Therefore, B is producible but not sustainable.

In the network depicted in Fig. 1B, metabolite B is sustainable from the nutrient A . The difference to the network from Fig. 1A is that from each consumed molecule A one excess molecule X may be produced and therefore the concentrations of all intermediates X , Y and Z may increase simultaneously. Consequently, all intermediates and the product B are sustainable. Fig. 1B also demonstrates why metabolites which are sustainable on U are not necessarily contained in the scope of U . Clearly, since intermediate X is required along the synthesis route to B , the expansion stops with A . In fact, the problem of the biological interpretation of a scope results from the fact that network expansion cannot account for such cyclic dependencies in which the presence of a metabolite is necessary for its own production. A key metabolite exhibiting this kind of dependency is ATP. In early steps during the synthesis of adenine nucleotides thermodynamically unfeasible reactions requiring the consumption of ATP are involved.

Despite the completely different approaches underlying the definitions of sustainability and scopes, we have found that they are often identical.

3. Results

We systematically compare sets of producible and sustainable metabolites with the corresponding scopes for the two genome-scale metabolic networks of *Escherichia coli* [13] and *Methanosarcina barkeri* [4]. These well-characterized organisms have been fully sequenced and their genome-scale networks have been manually curated and are therefore considered as representative examples.

The network of *E. coli* contains 932 reactions connecting 761 metabolites. For 143 metabolites exchange fluxes are defined, meaning that they may pass through the surface of the cell and are available as nutrients if abundant in the environment. If all these metabolites are assumed to be present, the set of sustainable metabolites amounts to 628. In contrast, the scope of the 143 external metabolites results in a set of only 312 compounds. If, however, it is assumed that ATP is available as a cofactor, the scope of the nutrients contains exactly all 628 sustainable metabolites. Similarly, for *M. barkeri*, a network with 619 reactions connecting 628 metabolites of which 70 are external, there are 475 metabolites which are sustainable on the set of all 70 nutrients. The scope of the nutrients comprises only 138 metabolites. In this case, the addition of the cofactor functionalities of ATP and NADH and NADPH yields a scope which is identical to the set of sustainable metabolites.

Inspired by these findings, we perform a systematic comparison by first considering the idealized situation in which exactly one metabolite is initially available. Further, we assume that water is largely abundant and therefore include water to every seed without explicitly mentioning it. In Fig. 2 the number of metabolites in the scope of a single metabolite (and water) is compared to the number of metabolites which are producible from this metabolite (and water) alone according

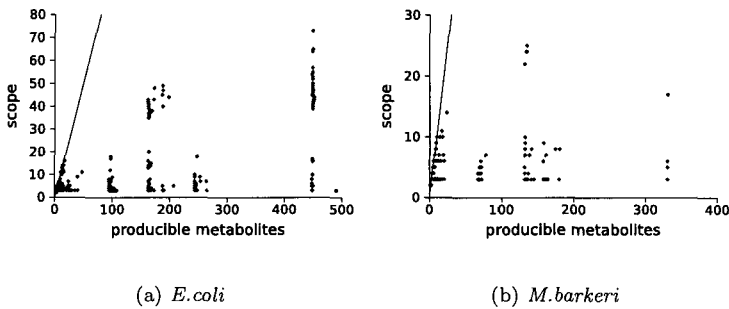


Fig. 2. Comparison between scope size and numbers of producible metabolites. Each dot represents one metabolite. Dots on the straight line represent metabolites for which the scope size equals the number of producible metabolites.

to condition (2). In approximately 44% of all cases, the scope is identical to the set of producible metabolites. However, identity is only observed for small sets with the majority being those cases in which the scope is identical to the seed. In most cases the set of producible metabolites is considerably larger than the size of the corresponding scope. This is not surprising, considering that the criteria for obtaining producible metabolites are weaker than for metabolites in the scope.

Interestingly, the size distributions of both sets are clearly structured. For the scopes, this property has been extensively investigated in [2, 7] and the results have been used to derive a hierarchical ordering of metabolism [5, 12]. Apparently, there exists a similar ordering of sets of producible metabolites.

Fig. 3 shows the direct comparison of scope sizes and numbers of sustainable metabolites. In the *E. coli* network, these sets are identical in 97% of all cases and in *M. barkeri* identity is observed in almost 99%. Those metabolites for which the corresponding sets differ are labelled by the abbreviations used in [4, 13]. Remarkably, many metabolites in the *E. coli* network exhibiting differences in the sets of sustainable metabolites and those in the scope are related to important cofactors. In particular, many adenine nucleotide phosphates and nicotineamide dinucleotide phosphates belong to this class. In both networks, many sugar phosphates also show a considerable difference in the corresponding sets.

Because cofactors apparently take on a role as key metabolites in both networks, a detailed investigation of their influence on scope size and contents is performed. We specifically consider the following four cofactor functionalities: 1) transfer of a phosphate group from ATP to an acceptor, yielding ADP, 2) simultaneous hydrolysis of two phosphate groups from ATP yielding AMP, 3) reduction of NAD^+ to yield NADH, thereby oxidizing another compound, 4) the analogous process but involving $\text{NADP}^+/\text{NADPH}$. Apparently, the introduction of a cofactor functionality can only increase the scope. We have systematically compared the scopes resulting for all 16 combinations of cofactor functionalities with the sets of sustainable metabolites.

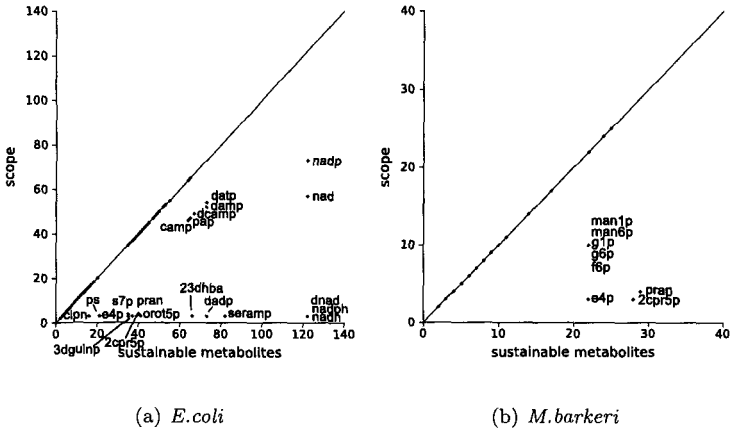


Fig. 3. Comparison between scope size and number of sustainable metabolites. Metabolites are represented as dots. Metabolites for which the scope size is not identical to the number of sustainable metabolites (located below the diagonal) are labeled. For clarity, two metabolites were omitted in figure (a): *acg5p* (sustainable metabolites 279, scope 3) and *glu5p* (264, 3).

In Fig. 4, the results for the *E. coli* network for the four cofactor combinations ATP/ADP, ATP/ADP and ATP/AMP, NADH and NADPH, and all cofactors are shown.

Interestingly, introduction of the redox cofactors NADH and NADPH lead to a stronger increase in scope size as the introduction of the phosphate transfer cofactors ATP/ADP and ATP/AMP. The latter case, in which both ATP related cofactor functionalities are introduced, is of particular importance. Here, the scopes of many central metabolites including NAD^+ , NADP^+ and deoxyadeninephosphates are identical to the corresponding sets of sustainable metabolites. There exist, however, other metabolites whose scope is always considerably lower than the set of sustainable metabolites, which holds true for both investigated networks.

A thorough investigation of the participating reactions preventing the expansion of the scope leads to the identification of metabolites, whose addition directly to the seed resulted in identity of scope and sustainable metabolites. In Table 1, the

Table 1. Selection of metabolites that have to be added to the seed in order to obtain the same result for scope and sustainable metabolites.

network	addition to seed	affected metabolites
<i>E. coli</i>	Proton (H^+)	ps, 3dglp, orot5p
	ATP	dnad, nadh, nadph
	D-Ribulose 5-phosphate	e4p, s7p
<i>M. barkeri</i>	D-Ribulose 5-phosphate	man1p, man6p, g1p, f6p, g6p, e4p
both	Proton (H^+)	pran, 2cpr5p

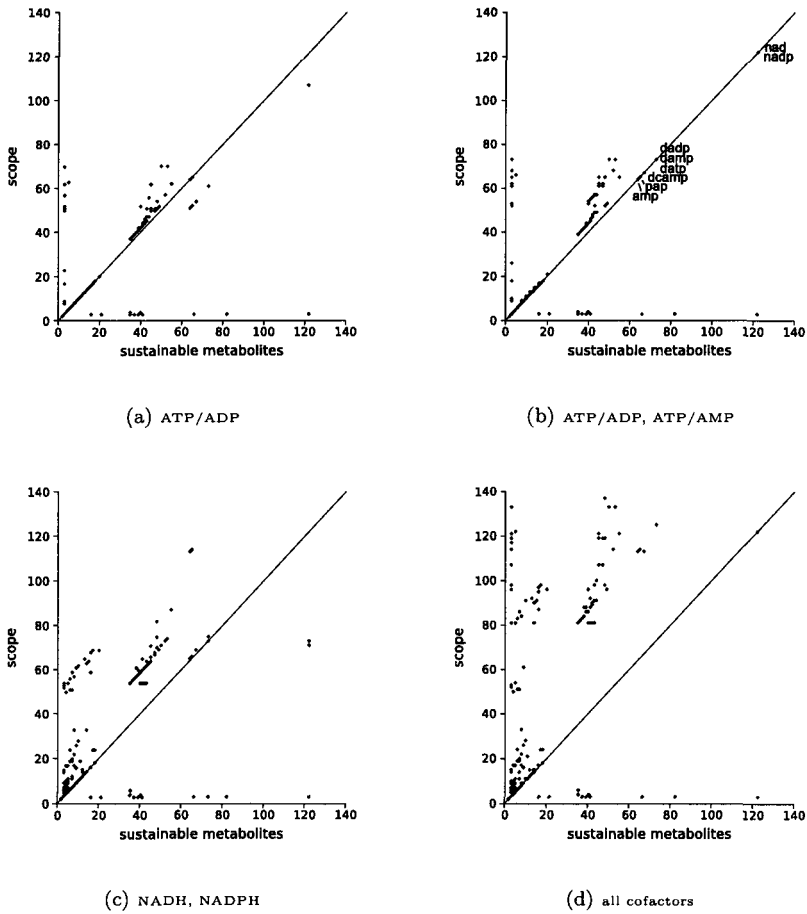


Fig. 4. Comparison of scope size with different cofactor functionalities with the numbers of sustainable metabolites for the *E. coli* network. In (b) metabolites have been labeled whose scope rose to the size of the sustainable metabolites by adding both cofactor functionalities of ATP.

most predominant examples are presented.

To study whether the finding that the inclusion of cofactor functionality improves the agreement of scopes with sets of sustainable metabolites is of a general nature, we performed a Monte Carlo simulation. For this, we randomly generated 1000 seeds with sizes varying between 10 and 100. For both networks, the scopes for all possible combinations of cofactors as well as the corresponding sets of sustainable metabolites have been determined. In Fig. 5 the degree of agreement of the sets is plotted versus the seed size. Interestingly, the behaviour differs for both networks. Whereas in both cases the agreement increases with increasing seed size

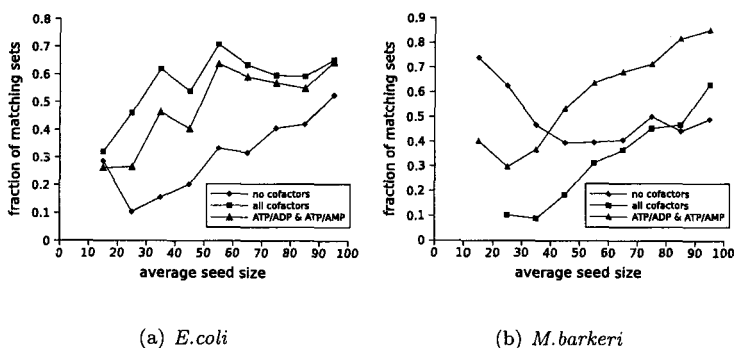


Fig. 5. Degree of identity of scopes and sustainable metabolites for both investigated networks as a function of seed size. Black line represents network expansion without cofactors, the green line with the two ATP-related cofactors and the red line considering all four cofactor functionalities.

when cofactors are included, this is not true for scopes without cofactors. In the case of *E. coli*, the best agreement is obtained by considering all cofactor functionalities simultaneously (for seed sizes larger than 10). In contrast, in *M. barkeri* inclusion of both ATP related cofactor functionalities for large seed sizes (> 40) yields the highest degree of identity.

4. Discussion

We have introduced several mathematical descriptions defining the producibility of metabolites from available nutrients. Simple producibility is given when a steady state flux through the metabolic network may exist such that the concentration of a metabolite increases while exclusively consuming the nutrients. By considering a cell under persistent growth, we arrive at the concept of sustainability, which defines metabolites whose concentrations may be increased even if all intermediates are simultaneously diluted. The method of network expansion provides the concept of a scope of nutrients, describing what a network may produce if exclusively the nutrients are present and all intermediates possess zero concentration.

We have systematically compared sets of producible and sustainable metabolites with the scopes obtained from single initial compounds and found that the scope is often identical to the set of sustainable compounds. We could further show that including cofactor functionalities, which are derived from heuristic arguments, can significantly increase the number of identical cases. More importantly, Monte Carlo simulations for larger sets of nutrients showed a tendency towards greater accordance of scope and sustainability with an increasing number of nutrients.

For some metabolites, the introduction of cofactor functionality was not sufficient to produce a scope identical to the set of sustainable metabolites. It is to be expected that this also holds true for combinations of seed compounds. In some

cases, the addition of protons to the nutrients was sufficient to enlarge the scope to the sustainable metabolites. Since protons in most cases do not influence the size of a scope, it seems reasonable to generally include them in the seed. This is in particular plausible since we always considered water to be abundant and in aqueous solutions protons are always present. An interesting observation was made for some metabolites occurring in the pentose phosphate pathway. Erythrose-4-phosphate (E4P), for example, exhibits a very small scope but in both networks the corresponding sets of sustainable metabolites are significantly larger. This observation can be explained by considering the structure of the pentose phosphate cycle which contains many bimolecular reactions. A subset can easily be assembled allowing for a stationary flux producing, for example, xylulose-5-phosphate and glyceraldehyde-3-phosphate from two molecules of E4P. However, since E4P never appears as a single substrate, it is evident that the scope of E4P only contains E4P itself. This fact has practical consequences for a whole class of other organism-specific networks. Most photosynthetic organisms, such as plants or green algae, can fix CO₂ by means of the Calvin cycle which bears high similarities with the pentose phosphate cycle. To realistically assess the biosynthetic capabilities from nutrient combinations including CO₂, also other compounds of the Calvin cycle, such as ribulose-1,5-bisphosphate should be added. A thorough investigation of genome-scale networks of photoautotrophic organisms is still outstanding.

Although the concept of sustainability is mathematically more rigorous, it has the drawback that it is computationally very intensive. For some calculations of sets of sustainable metabolites, several hundred linear programming problems have to be solved. In contrast, the network expansion algorithm is extremely simple and fast and can easily be applied millions of times on a normal personal computer, rendering it suitable for large-scale applications for example to investigate thousands of nutrient combinations for hundreds of networks. Considering that the agreement of scopes with sets of sustainable metabolites is in most cases extremely accurate, we conclude that the enormous gain in computational speed justifies the inaccuracies that the network expansion method unavoidably displays due to the introduction of heuristic cofactor functionalities.

References

- [1] Ebenhöh, O., Handorf, T., and Kahn, D., Evolutionary changes of metabolic networks and their biosynthetic capacities. *Syst Biol (Stevenage)*, 153(5):354–358, Sep 2006.
- [2] Ebenhöh, O., Handorf, T., and Heinrich, R., Structural analysis of expanding metabolic networks. *Genome Inform*, 15(1):35–45, 2004.
- [3] Edwards, J. S. and Palsson, B. O., Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 1:1, 2000.
- [4] Feist, A.M., Scholten, J.C.M., Palsson, B.O., Brockman, F.J., and Ider, T., Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology*, 2:2006.0004, 2006.
- [5] Handorf, T., Ebenhöh, O., Kahn, D., and Heinrich, R., Hierarchy of metabolic compounds based on their synthesising capacity. *Syst Biol (Stevenage)*, 153(5):359–363,

Sep 2006.

- [6] Handorf, T., Christian, N., Ebenhöh, O., and Kahn, D., An environmental perspective on metabolism. *J Theor Biol*, 252:530–537, Nov 2007.
- [7] Handorf, T., Ebenhöh, O., and Heinrich, R., Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J Mol Evol*, 61(4):498–512, Oct 2005.
- [8] Ibarra, R.U., Edwards, J.S., and Palsson, B.O., *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, Nov 2002.
- [9] Imielinski, M., Belta, C., Rubin, H., and Halasz, A., Systematic analysis of conservation relations in *Escherichia coli* genome-scale metabolic network reveals novel growth media. *Biophysical Journal*, 90:2659–2672, 2006.
- [10] Kauffman, K.J., Prakash, P., and Edwards, J.S., Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5):491–496, Oct 2003.
- [11] Liolios, K., Mavromatis, K., Tavernarakis, N., and Kyrpides, N.C. The genomes on line database (gold) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res*, 36(Database issue):D475–D479, Jan 2008.
- [12] Matthäus, F., Salazar, C., and Ebenhöh, O., Biosynthetic potentials of metabolites and their hierarchical organization. *PLoS Comput Biol*, 4(4):e1000049, Apr 2008.
- [13] Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O., An expanded genome-scale model of *Escherichia coli* K-12 (ijr904 gsm/gpr). *Genome Biology*, 4(9):R54.1–R54.12, 2003.
- [14] Schilling, C. H., Edwards, J. S., Letscher, D., and Palsson, B. O., Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng*, 71(4):286–306, 2000.