METABOLIC SYNERGY: INCREASING BIOSYNTHETIC CAPABILITIES BY NETWORK COOPERATION

NILS CHRISTIAN[1]
nils.christian@physik.fu-berlin.de

THOMAS HANDORF[1]
thomas.handorf@physik.hu-berlin.de

OLIVER EBENHÖH[2]
ebenhoeh@mpimp-golm.mpg.de

[1] *Institute for Biology, Humboldt University Berlin, Germany*
[2] *Max Planck Institute for Molecular Plant Physiology, Potsdam-Golm, Germany*

Cooperation between organisms of different species is a widely observed phenomenon in biology, ranging from large scale systems such as whole ecosystems to more direct interactions like symbiotic relationships. In the present work, we explore inter-species cooperations on the level of metabolic networks.
For our analysis, we extract 447 organism specific metabolic networks from the KEGG database [7] and assess their biosynthetic capabilities by applying the method of network expansion [5]. We simulate the cooperation of two organisms by unifying their metabolic networks and introduce a measure, the gain Γ, quantifying the amount by which the biosynthetic capability of an organism is enhanced due to the cooperation with another species. For all theoretically possible pairs of organisms, this synergetic effect is determined and we systematically analyze its dependency on the dissimilarities of the interacting partners. We describe these dissimilarities by two different distance measures, where one is based on structural, the other on evolutionary differences.
With the presented method, we provide a conceptional framework to study the metabolic effects resulting from an interaction of different species. We outline possible enhancements of our analysis: by defining more realistic interacting networks and applying alternative structural investigation methods, our concept can be used to study specific symbiotic and parasitic relationships and may help to understand the global interplay of metabolic pathways over the boundary of organism specific systems.

Keywords: metabolism; scope; KEGG; symbiosis; systems biology; synergy.

1. Introduction

For a few years, the number of fully sequenced and annotated genomes is increasing with an amazing speed and, considering the number of ongoing sequencing projects, is likely to increase even faster in the near future. Using homology matching methods and a tedious manual curation, for a substantial number of organisms largely complete metabolic networks have been characterized. With the emergence of comprehensive metabolic databases such as KEGG [7] or MetaCyc [9], such networks have become readily accessible. Existing methods to analyze large scale metabolic networks include elementary flux modes [11, 12], the closely related concept of extreme fluxes [10], flux balance analysis [8] as well as graph theoretical approaches [6, 14].

All these methods have in common that they can be performed even without the specific knowledge of the kinetic properties of the enzymes catalyzing the biochemical reactions.

So far, large scale metabolic network analyses focused to a large extent on single organism networks, see e. g. [13, 16]. In their natural habitats, however, all species are in constant interaction with organisms belonging to other species. On a population level, the interaction of different species is mathematically described in the research field of population dynamics, ranging from simple predator prey models to very complex ecosystem models, for an overview see e. g. [3]. However, the interaction does not only take place on the level of populations, but also on the level of single individuals by the exchange of metabolites. Examples are given by a predator that consumes and digests its prey, or, more directly, by an intracellular symbiont living inside a host cell and exchanging intermediates by specific transporters. Inspired by these facts, we have developed a conceptional framework to study such interactions on the metabolic level, and to quantify the benefit for each of the organisms.

We retrieve 447 organism specific networks from the KEGG database and determine their capability to incorporate glucose as a sole carbon source into their metabolism. To quantify this capability, we apply the concept of scopes [5], where a scope characterizes the biosynthetic capability of a network when it is provided with certain external resources. To determine how a cooperation of two organism's metabolic networks may enhance the biosynthetic capabilities of each other, we construct unified networks for all possible pairs of organisms. We introduce a measure, called the gain, quantifying the increase in biosynthetic capabilities, by comparing the performance of the unified network with those of the single organism networks. We investigate how the gain correlates with the dissimilarity of the networks, for which we provide two measures, one based purely on structural properties and the other exclusively on phylogenetic information.

The introduced methodology as well as the results from the systematic interaction analysis provides a basis for the investigation of specific biological examples of parasitic or symbiotic behavior. We expect that the biological significance of such investigations may be considerably enhanced by refining the models for interacting networks as well as by applying other network analysis techniques.

2. Concepts

2.1. *Biosynthetic capabilities*

The metabolic network of a particular organism, denoted O, is defined by a specific set of biochemical reactions. We evaluate the biosynthetic capability of a network using the method of network expansion [5]. Starting from a set of compounds, called the seed and denoted S, a series of expanding networks is constructed in an iterative manner. In each step, those reactions from O are added to the network which use exclusively those metabolites as substrates which occur either in the seed or as products of reactions included in earlier steps. The iteration stops if

no new reactions can be included. The set of compounds within the final network is called the scope of the seed, denoted $\Sigma^{O}(S)$, and by construction comprises all those compounds which can in principle be produced under the condition that only the seed compounds are available. Often, we are concerned not with the exact composition of a scope but rather with the number of metabolites it contains. We denote the scope size by $|\Sigma^{O}(S)|$. The scope is in general a useful measure to relate structural and functional network properties and is used in this work to characterize the biosynthetic capabilities of a metabolic network.

In cellular metabolism, there exist a small number of key metabolites, the cofactors, which occur in many reactions and mostly perform one particular function. For example, the most common usage of ATP is the transferral of a phosphate group to another molecule yielding ADP and, due to the free energy change of the hydrolyzation, drive reactions that would otherwise be thermodynamically unfeasible. Similarly, NADH is involved in a large number of redox reactions in which it functions as an electron donor yielding the oxidized form, NAD^{+}. In the process of network expansion, a reaction involving a cofactor may only be used if the cofactor has already been synthesized from the seed compounds by reactions incorporated into the network in previous steps. Under most physiological conditions, however, a cell maintains a substantial level of such cofactors and therefore it is unrealistic to assume that they have to be manufactured de novo. Throughout this work, we use a modified form of the expansion algorithm, which allows that cofactor functions can be performed even if the cofactors have not yet been synthesized. The inclusion of cofactor functionalities in the algorithm was introduced in [5] and [4].

Some reactions are considered as irreversible because under physiological conditions they can only proceed in one direction. However, in principle every biochemical reaction may be reversed and the actual direction depends strongly on the present state of a cell as well as on the considered cell or tissue type. In our analysis, we have considered all reactions as reversible.

Clearly, the scope strongly depends on the network composition as well as on the available seed compounds. Naturally, the choice of seed compounds is crucial for the biological interpretation of the biosynthetic capability. One important function of metabolism is to incorporate external carbon sources and convert them into organic compounds used by other processes. It is therefore of interest to study how different carbon sources may be incorporated into cellular metabolism. In this work, we focus on glucose, which is a central metabolite in the energy metabolism ubiquitous throughout all domains of life. To assess this capacity, we identify all non carbon containing compounds occurring in at least one network and include them, together with glucose, in the seed.

2.2. *Metabolic Synergy*

In biological environments, no species lives completely isolated. Rather, metabolites are exchanged between different species by a variety of mechanisms.

The aim of this work is to study how metabolic networks may mutually benefit from each other by sharing their metabolic resources. For this, we investigate pairs of organisms and assume the simplest possible metabolic interaction, namely that the two organisms may exchange all intermediate metabolites. Such a scenario is simply described by a metabolic network which is the union of the two single species networks. Let O_1 and O_2 denote the networks of the single organisms, then the unified (interacting) network is written as $O_1 \cup O_2$. For a pair of organisms, we determine the metabolic capabilities of the single organisms, $\Sigma^{O_1}(S)$ and $\Sigma^{O_2}(S)$, as well as the metabolic capability for the unified network, $\Sigma^{O_1 \cup O_2}(S)$. Clearly,

$$\Sigma^{O_1 \cup O_2}(S) \supseteq \Sigma^{O_1}(S) \cup \Sigma^{O_2}(S), \tag{1}$$

where equality signifies that there is no increase in the metabolic capabilities as a result of network interaction. To quantify the positive synergetic effect resulting from sharing the metabolic resources of two networks, we introduce the gain Γ as the increase in size of the scope of the unified network over the union of the scopes of the single networks,

$$\Gamma(O_1, O_2; S) = \left|\Sigma^{O_1 \cup O_2}(S)\right| - \left|\Sigma^{O_1}(S) \cup \Sigma^{O_2}(S)\right|. \tag{2}$$

Thus, the gain equals to the number of metabolites which can be produced from the interacting network, but not from either of the single networks.
While the gain Γ describes the synergetic effect for a pair of organisms, it is also of relevance to study how each of the partners benefits from the interaction. For this, we introduce the quantities

$$\Gamma_{1/2}(O_1, O_2; S) = \left|\Sigma^{O_1 \cup O_2}(S)\right| - \left|\Sigma^{O_{1/2}}(S)\right| \tag{3}$$

and measure the asymmetry of the interaction by the bias

$$\beta(O_1, O_2; S) = \frac{|\Gamma_1(O_1, O_2; S) - \Gamma_2(O_1, O_2; S)|}{\Gamma_1(O_1, O_2; S) + \Gamma_2(O_1, O_2; S)}. \tag{4}$$

This value is only defined if at least one of the organisms benefits from the interaction. It is 0 if both organisms increase their metabolic capabilities by the same amount of metabolites, and 1 if only one of the partners benefits. For simplicity, the arguments of Γ and β will be omitted if they are unambiguous.

2.3. *Distances between networks*

An intriguing question is whether structural properties such as the degree of similarity of the interacting networks determine the principle capacities for a positive synergetic effect. To study this, we relate the dissimilarity of a pair of organisms with the increase of metabolic capability resulting from cooperation. The increase in capability is quantified by the gain, defined in Eq. (2). To quantify the dissimilarity of two organisms, we introduce two distance measures, one based on differences in the underlying network structure and the other on their phylogenetic distance.

Network distance. We measure the structural distance between two metabolic networks by counting those reactions which occur in only one of the networks. The quantity

$$d_N(O_1, O_2) = \left|O_1 \cup O_2\right| - \left|O_1 \cap O_2\right| \tag{5}$$

is the Manhattan distance between the two sets of reactions O_1 and O_2, where $|O|$ is the number of biochemical reactions within a network O. Clearly, two identical networks possess zero distance whereas two completely distinct networks possess a distance equal to the sum of the reactions in both single networks. A noteworthy feature of this distance is that, if one network is completely contained in the other, their distance may still be large.

Evolutionary distance. Using the NCBI taxonomy tree [1], we approximate the evolutionary distance of two species by the number of edges on the shortest path between the two organisms. We define the evolutionary distance

$$d_E(O_1, O_2) = \text{number of edges on shortest path connecting } O_1 \text{ and } O_2. \tag{6}$$

This distance measure can only give an approximation of the true evolutionary distance because it depends on the structure of the underlying phylogenetic tree. Moreover, it weighs every edge in this tree equally, while the number of levels may vary substantially from subtree to subtree, often reflecting the thoroughness with which a particular subtree has been studied rather than a true evolutionary distance.

3. Results

3.1. *Metabolic capacities of single organisms*

In the following, we study the ability of organisms to incorporate glucose in their metabolism. For this, we define a seed containing glucose and inorganic material, as described in Sec. 2.1. Because this particular seed contains 92 chemical species, this is also the minimal number of compounds contained in each scope.

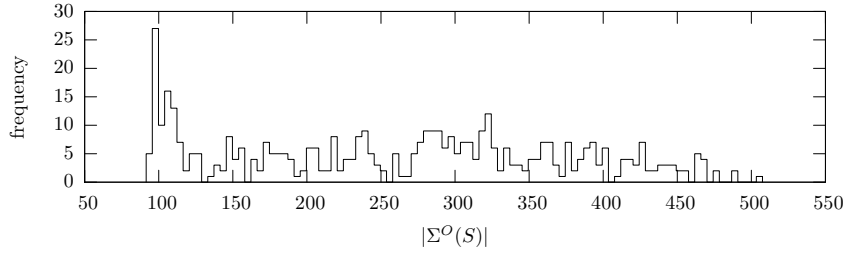

Fig. 1. Histogram of the scope sizes for all considered organisms. The smallest observed scope size is 92, which corresponds to the number of seed metabolites.

Fig. 1 shows a histogram of the metabolic capabilities for all investigated 447 organism specific metabolic networks. A substantial fraction (90) of the considered

organisms display a scope size of less than 130, meaning they are capable to produce less than 40 new metabolites. Three of these organisms (*N. equitans* and the two *Phytoplasmae OY* and *AYWP*) are not capable to synthesize any new metabolites. The largest scope size is 508, exhibited by the beta proteobacterium *Burkholderia sp. 383*, meaning it can synthesize 416 new carbon containing metabolites.

3.2. *Interaction of metabolic networks*

We systematically investigate how the synthesizing capacities are increased if two organisms share their metabolic reactions. We first study how even the beneficial effects of the interactions are distributed among the two partners. For this, we calculate the bias values β, defined in Eq. (4), for all 99681 possible network pairs. In 97 cases the bias is not defined ($\Gamma_1 = \Gamma_2 = 0$), meaning that neither partner can increase its biosynthetic capability. Fig. 2 shows a histogram of the bias values β. The pairs in which only one of the partners benefits from the interaction ($\beta = 1$)

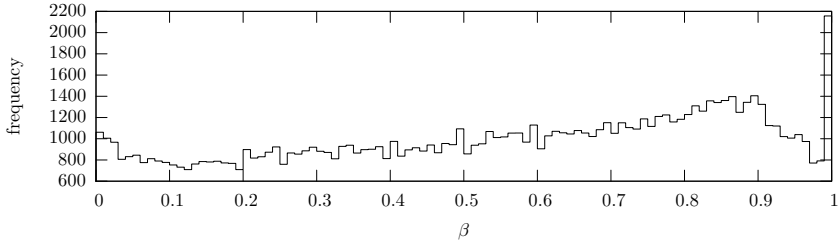

Fig. 2. Histogram of the bias value β as defined in Eq. (4). A value of $\beta = 0$ describes an interaction which is beneficial for both partners, $\beta = 1$ an interaction in which only one partner benefits.

are overrepresented. A closer inspection shows that in most of these cases a small network is almost completely contained in a larger one, thus explaining why the latter cannot increase its capacity due to the interaction. A more detailed analysis will be necessary to determine whether the small sizes reflect the biological reality or whether they result from incomplete annotations. In the majority of pairs (98.5%), both partners benefit from the interaction, where, as a tendency, interactions with a stronger bias are more frequent than those with a weaker bias.

To visualize how the gain Γ, defined in Eq. (2), correlates with the network distance d_N, defined in Eq. (5), we sort the organism pairs by their distance and group them into equidistant bins. The number of organism pairs per bin is plotted in the top panel of Fig. 3, demonstrating that most organisms exhibit a network distance between 400 and 1000. For the gain values in each bin, we determine the 10% quantile, the median, the 90% quantile and the maximum and plot these values versus the network distance in the bottom panel of Fig. 3.

It can be observed that for very small network distances ($d_N < 100$, 0.3% of all organism pairs) the gain is also very small. This is not surprising since similar

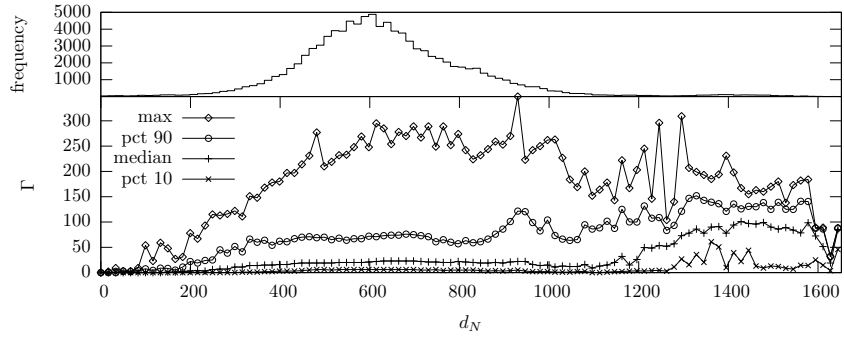

Fig. 3. Top panel: Histogram of the network distances d_N, defined in Eq. (5), for all pairs of networks. Bottom panel: Correlation between the gain Γ, defined in Eq. (2), and the network distance d_N. Plotted are the 10% quantile, the median, the 90% quantile and the maximum gain values as a function of d_N.

networks contain mostly the same pathways. As a consequence, an interaction of such networks is unlikely to yield positive effects.

For network distances $100 < d_N < 400$ (6.6% of pairs), the 90% quantile and the maximum increase strongly with increasing distance. This reflects that a certain degree of dissimilarity is necessary for one network to be able to utilize a subnetwork of the other, in order to increase its capacity.

Remarkably, the quantiles do not change considerably for a large range of network distances ($400 < d_N < 1100$, 91.1% of pairs). This indicates that the difference between two networks alone is not sufficient to predict the increase in biosynthetic capabilities. As long as two networks are not too similar or too distant, particular structural features like the specific occurrence of pathways are apparently more important for the synergetic effect than the degree of dissimilarity of the whole networks. A noticeable characteristic is the increase in the 90% quantile in the range $850 < d_N < 1050$, containing a total of over 8400 network pairs. The most abundant organisms in this region are *C. albicans* and *A. thaliana*, occurring in 386 and 335 pairs, respectively. Only 734 of these 8400 pairs show a gain larger than 100. Surprisingly, in about one third (277) of these pairs, one of the partners is *A. thaliana*, and 239 pairs contain *C. albicans*. The high frequency of pairs containing one of these two organisms together with the fact that pairs involving these organisms tend to produce a high gain explains the observed increase of the 90% quantile.

For even larger network distances ($d_N > 1100$, 2.1% of pairs), the median and 90% quantile significantly increase, whereas the maximum strongly fluctuates. A closer inspection reveals that in 62.8% of the pairs within this distance region, one of the interacting partner is either human (*H. sapiens*), mouse (*M. musculus*), or rat (*R. norvegicus*) and, when considering only those pairs yielding a gain $\Gamma > 50$, the fraction increases to 77.3%. Again, the increase in the quantiles can be explained by a high abundance of organisms which on average yield a high gain.

Fig. 4A depicts the correlation between gain Γ and evolutionary distance d_E, defined

by Eq. (6). It corresponds to Fig. 3 with the difference that pairs exhibiting the same evolutionary distance are grouped into bins. The similar appearance of these two figures can be explained by the high correlation of the two distance measures, which are plotted against each other in Fig. 4B.

A remarkable attribute of Fig. 4A is that the maximum is already very pronounced for the smallest possible distance $d_E = 2$. We found that this pair consists of two strains of the gamma proteobacteria *Shewanella* whose network sizes differ significantly by over 400 reactions, thus resulting in a large network distance d_N, which, as outlined above, is required for a high gain. The large network distance hints at incomplete or faulty annotations of the corresponding genomes because it seems unlikely that organisms may develop such a drastic different metabolic network composition during a relatively short evolutionary time span.

4. Discussion and Outlook

Motivated by the observation that no organism exists in complete isolation, but rather exchanges metabolites with organisms belonging to other species, we have provided a conceptional framework to analyze changes in biosynthetic capabilities that result from a cooperation between metabolic networks. In this work, we have measured biosynthetic capabilities using the concept of scopes. A scope describes the maximal synthesizing capacity of a network when it is provided with a specific set of external resources. We have systematically investigated the pairwise interactions of 447 organisms, for which we have retrieved the networks from the KEGG database. We focused on finding correlations between the maximal amount of increase in biosynthetic capabilities and the dissimilarities of the investigated organisms, both with respect to the network structure and to the phylogenetic distance.

We have found that in some cases there is no measurable increase compared to the functioning of the networks in isolation, but for some network pairs the biosynthetic capability increases dramatically due to the mutual cooperation.

Naturally, the obtained results will critically depend on the quality of the underlying

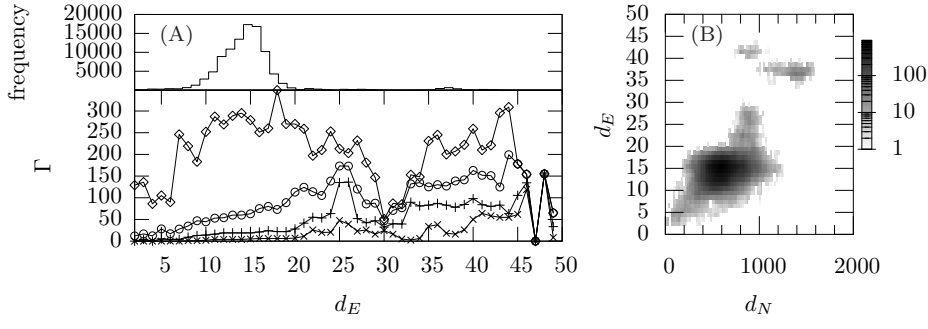

Fig. 4. (A) Same as Fig. 3, but with values plotted against the evolutionary distance d_E, defined by Eq. (6). (B) Correlation of the two distance measures, d_N and d_E.

network data. It is striking that those organisms which, according to our statistical analysis, appear to play an outstanding role, are also among the most thoroughly studied (*H. sapiens*, *R. norvegicus*, *M. musculus* and *A. thaliana*). This observation gives rise to the speculation that the networks of these organisms are far more complete than networks of less intensely studied species, and thus have the potential to infer a stronger synergetic effect in conjunction with other organisms. Erroneous network structures may also arise from reactions in the KEGG database displaying inconsistent stoichiometries. Such reactions were simply ignored in our calculations. In addition, many of the metabolic enzymes included in the database are inferred by sequence analysis. However, homology matches can never provide absolute certainty that an identified gene actually exists in the investigated organism, that it is actively transcribed and translated into the protein, and that the gene product catalyzes the assumed reaction.

Despite the dependence on high quality networks, the here described methodological framework opens a wide field of future investigations. While the systematic approach yields interesting results on a general basis, such as the average increase in biosynthetic capability as a function of network dissimilarity, only the closer investigation of well studied interacting species will provide insight into the specific mechanisms that are responsible for a mutual benefit and therefore into the principles of symbiotic relationships. A particularly interesting field of study will be the symbiosis between plants of the Fabaceae family with Rhizobia, bacteria that possess the ability to fix nitrogen from the air into nitrate or ammonia which is usable by the plant. For an overview of the mechanisms, see e. g. [2].

In real biological systems, the exchange of metabolites between species is limited by the fact that transporters are required to bring the substances into the cells. A more realistic description of interacting networks can be obtained by the inclusion of such transport mechanisms. Whereas this improvement is technically easy to achieve, the actual realization is unfortunately still hindered by the limited knowledge on the transporters and their substrate specificities.

In the presented work, we have applied the method of network expansion to assess biosynthetic capabilities. A major drawback of this method is that it can only account for positive effects of the interaction. However, negative effects may occur in parasite-host interactions where the species are competing for a substrate or the parasite is drawing important intermediates, such as glucose phosphate or ATP, from the host and thereby reduces its biosynthetic production rates. Both, positive and negative effects can be accounted for by invoking other large scale network examination methods such as flux balance analysis [8], allowing for the calculation of optimal flux distributions. An interesting object of study is Wolbachia, which resides inside the cells of several insect species, but is also found in nematodes. While it is clearly parasitic in insects, its relationship to nematodes can rather be described as symbiotic [15]. By comparison of parasites and symbionts, we expect to gain understanding how such mutual interdependencies may have evolved.

5. Acknowledgments

We thank the following organizations for financial support: The International Research Training Group "Genomics and Systems Biology of Molecular Networks" (Christian, N.), the German Research Foundation, in particular the Collaborative Research Center "Theoretical Biology: Robustness, Modularity and Evolutionary Design of Living Systems" (Handorf, T.) and the German Federal Ministry of Education and Research, Systems Biology Research Initiative "GoFORSYS" (Ebenhöh, O.).

References

[1] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., and Wheeler, D.L., GenBank, *Nucleic Acids Res.*, 28(1):15–18, 2000.

[2] Denison, R.F. and Toby, Kiers. E., Why are most rhizobia beneficial to their plant hosts, rather than parasitic?, *Microbes Infect.*, 6(13):1235–1239, 2004.

[3] Edelstein-Keshet, L., *Mathematical Models in Biology*, Society for Industrial and Applied Mathematics, 2005.

[4] Handorf, T. and Ebenhöh, O., Metapath online: A web server implementation of the network expansion algorithm, *Nucleic Acids Res.*, 35(Web Server issue):W613–618, 2007.

[5] Handorf, T., Ebenhöh, O., and Heinrich, R., Expanding metabolic networks: Scopes of compounds, robustness, and evolution, *J. Mol. Evol.*, 61(4):498–512, 2005.

[6] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L., The large-scale organization of metabolic networks, *Nature*, 407(6804):651–654, 2000.

[7] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M., From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res.*, 34(Database issue):354–357, 2006.

[8] Kauffman, K.J., Prakash, P., and Edwards, J.S., Advances in flux balance analysis, *Curr. Opin. Biotechnol*, 14(5):491–496, 2003.

[9] Krieger, C.J., Zhang, P., Mueller, L.A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y., and Karp, P.D., MetaCyc: a multiorganism database of metabolic pathways and enzymes, *Nucleic Acids Res.*, 32(Database issue):D438–442, 2004.

[10] Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., and Palsson, B.O., Metabolic pathways in the post-genome era, *Trends. Biochem. Sci.*, 28(5):250–258, 2003.

[11] Schuster, S., Fell, D.A., and Dandekar, T., A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nat. Biotechnol.*, 18(3):326–332, 2000.

[12] Schuster, S. and Hilgetag, C., On elementary flux modes in biochemical reaction systems at steady state, *J. Biol. Syst.*, 2(2):165–182, 1994.

[13] Thiele, I., Vo, T.D., Price, N.D., and Palsson, B.O., Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants, *J. Bacteriol*, 187(16):5818–5830, 2005.

[14] Wagner, A. and Fell, D.A., The small world inside large metabolic networks, *Proc. Biol. Sci.*, 268(1478):1803–1810, 2001.

[15] Werren, J.H., O'Neill, S.L., and Hoffman, A. Eds., *Influential Passengers:inherited microorganisms and arthropod reproduction*, Oxford University Press, 1997.

[16] Wiback, S.J., Mahadevan, R., and Palsson, B.O., Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum, *Biotechnol Bioeng*, 86(3):317–331, 2004.