UCRL-JRNL-227543

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Optimal flux patterns in cellular metabolic networks

E. Almaas

January 29, 2007

Chaos

Disclaimer

Optimal flux patterns in cellular metabolic networks

Eivind Almaas[*]

Biosciences and Biotechnology Division,

Lawrence Livermore National Laboratory,

Livermore, California 94551, USA

Abstract

The availability of whole-cell level metabolic networks of high quality has made it possible to develop a predictive understanding of bacterial metabolism. Using the optimization framework of flux balance analysis, I investigate metabolic response and activity patterns to variations in the availability of nutrient and chemical factors such as oxygen and ammonia by simulating $30,000$ random cellular environments. The distribution of reaction fluxes is heavy-tailed for the bacteria *H. pylori* and *E. coli*, and the eukaryote *S. cerevisiae*. While the majority of flux balance investigations have relied on implementations of the simplex method, it is necessary to use interior-point optimization algorithms to adequately characterize the full range of *activity* patterns on metabolic networks. The interior-point activity pattern is bimodal for *E. coli* and *S. cerevisiae*, suggesting that most metabolic reaction are either in frequent use or are rarely active. The trimodal activity pattern of *H. pylori* indicates that a group of its metabolic reactions (20%) are active in approximately half of the simulated environments. Constructing the high-flux backbone of the network for every environment, there is a clear trend that the more frequently a reaction is active, the more likely it is a part of the backbone. Finally, I briefly discuss the predicted activity patterns of the central-carbon metabolic pathways for the sample of random environments.

UCRL-JRNL-227543.

Metabolic networks consist of thousands of molecules that are processed and inter-converted by enzymatic reactions. The availability of high-fidelity annotated genomic data has made it possible to generate accurate models of whole-cell metabolism. The combination of these models with an optimization framework has brought us one step closer to a predictive understanding of metabolic function. In recent years, multiple cellular-level metabolic models have become available for bacteria, as well as one model for yeast, which is a eukaryote. Using a linear optimization method called "flux balance analysis," a number of computational studies have predicted cellular metabolic responses to perturbations that range from gene knockouts and metabolic network engineering, to the removal or addition of nutrients. Not only have these predictions compared surprisingly well with experimental data, but they have opened up new avenues of harnessing the abilities of single-celled organisms that have been fine-tuned by evolution. In this paper, I will discuss how the flux-balance approach predicts activity patterns of metabolic networks in response to variations in nutrients and chemical factors in the environment surrounding the cell. As a biological example, I will focus on the central carbon metabolism – the collection of metabolic pathways responsible for generating the 12 precursor metabolites used as basic building blocks by cells.

I. INTRODUCTION

The study of how components of a cell, such as mRNA molecules, proteins, and metabolites, interact and integrate to generate organism-level behavior has received enormous attention the last few years. The recent availability of large-scale empirical datasets for gene expression[1–3] and gene-regulation[4–6], signal-transduction[7], protein interactions[8–13], and metabolic networks[14–17] have made it possible to develop predictive models on the whole-cell level. The grand challenge posed by these datasets is twofold. First, it is necessary to develop methods capable of extracting meaningful information, preferably giving us the ability to answer central questions about how organisms function. Second, these datasets have mostly been analyzed independently of each other, although the structure and dynamics of *e.g.* the signal-transduction network strongly influences cellular metabolic activity.

Complex network analysis[18–20] offers a promising avenue not only for extracting non-trivial information from organism-level genetic, protein, and metabolic systems, but also seem well suited to integrate these varied datasets. The "systems" approach that is implicit in network representations is often complementary to the reductionist approach, which has experienced tremendous successes during the last few centuries. In particular, there exists considerable variation in the nature of both the elementary building blocks and the interactions in these biological systems[21,22].

Network approaches have been applied to a variety of biological systems, including protein-interaction and metabolic networks[18–20]. The surprising discovery was that these systems share many large-scale characteristics, although the nature of their nodes and the interactions between the nodes are very different. The most compelling such similarities are their small-world nature, large clustering, and heavy-tailed connectivity distribution[19]. However, the majority of network analyses have focused on the topological properties of these systems and not taken dynamics into account.

Flux balance analysis (FBA) is a computational method based on linear programming that takes as input a metabolic network in the form of a list of chemical reactions, a set of constraints on these reactions, and an objective function. Assuming that the linear problem is feasible, FBA will return predicted metabolic reaction turn-over rates, or fluxes, corresponding to an optimal utilization of the metabolic network with respect to the chosen goal function. FBA thus provides a tool to extend complex network analysis of metabolic networks from a purely topological approach to also consider dynamic effects. Note that, while the majority of flux balance analyses have been conducted by implementations of the simplex algorithm[23], I will discuss metabolic activity patters resulting from an interior-point (IP) algorithm[24]. For systems with a unique optimal solution, the two approaches will naturally yield the same answer. However for systems where the optimal state is degenerate, such as metabolic networks, the simplex algorithm typically returns an optimal solution with a small number (although not the minimal number) of non-zero reaction fluxes. In contrast, the IP algorithm will return an optimal solution with the largest possible number of non-zero fluxes. This distinction is important for the systematic calculation of metabolic activity patterns, which is the focus of this work.

II. OPTIMIZATION AND METABOLIC FLUXES

A. The flux balance method

The flux-balance method is based on the three basic assumptions that can briefly be stated as: (i) cellular metabolism is in a steady state, (ii) mass is conserved, and (iii) metabolic reaction fluxes optimize a goal function on the network[25]. The first hypothesis is motivated by the dire lack of information on kinetic parameters inside a cell. By assuming a steady state, it is possible to formulate the problem without appealing to time-derivatives, consequently avoiding kinetic parameters altogether. The trade-off is that the problem has now become underdetermined. The third hypothesis breaks the stalemate by allowing the question to be formulated as an optimization problem. The goal function to be maximized is typically chosen to be cellular growth rate, as competition for resources and evolution selects for fast growing organisms in many natural and laboratory situations. However, alternative objective functions such as maximal ATP production, or minimization of flux through a given pathway have also been investigated[25,26].

To formulate the flux-calculation problem more succinctly, the mass conservation constraint is expressed as

$$\frac{d[A_i]}{dt} = \sum_{j=1}^{n} S_{ij}\nu_j, \tag{1}$$

where $[A_i]$ is the concentration of metabolite i, ν_j is the unknown flux rate of reaction j, and S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. For example, if the chemical reaction j was $A + 2\,B \rightarrow C$, the stoichiometric coefficients would be $S_{A,j} = -1$, $S_{B,j} = -2$, and $S_{C,j} = 1$, while the coefficients $S_{X,j} = 0$ for all other metabolites X. Applying the steady-state approximation to Eq. (1), the linear problem is typically underdetermined, as there usually are more metabolic reactions that metabolites, $m > n$.

The FBA approach can be formulated as a standard form linear program

$$\min\{c^T\nu : S\nu = b,\ \nu \geq 0\}, \tag{2}$$

where S is the $m \times n$ stoichiometric matrix, $c, \nu \,\epsilon\, \mathbb{R}^n$, $b \,\epsilon\, \mathbb{R}^m$. The vector c corresponds to the goal function, commonly chosen to correspond to biomass, and b represents flux constraints, including the environmental availability of chemicals and nutrients. I will assume that a strictly positive, feasible solution exists. Further details are available in *e.g.* Refs. 25–27.

There is ample experimental evidence that the FBA methodology is biologically relevant. For instance, it is possible to simulate gene knockouts by computationally removing genes with products (enzymes) that participate in metabolic processes. In the case of the model of *H. pylori* metabolism, the predicted growth phenotypes (being characterized as "viable" or "lethal") agree with experimental results in 60% of the tested cases[28]. The results for the *E. coli*[15] and *S. cerevisiae*[17] metabolic networks are even more compelling, with agreement rates of 86% and 83% respectively. It is also possible to experimentally determine transport fluxes (uptake or production of metabolites and chemicals from the environment) using batch cultivation of the organism[29] or a chemostat. For batch growth of *E. coli*, a direct comparison of predicted and measured uptake flux values of acetate and oxygen as well as cellular growth rates, had an average error of only 5.8%, while using succinate as nutrient source instead of acetate resulted in an average error of 10.7%[29].

Since the goal function of choice for the FBA method is maximal cellular growth rate, it is important to be aware of underlying assumptions in making this choice. Arguably the three most important assumption are that, (i) the organism is functioning optimally, that (ii) it is in the exponential growth phase (the cellular population doubles at regular intervals), and that (iii) the environment is not changing. Carefully conducted experiments on wild-type and mutant *E. coli* cells kept in the exponential growth phase with a stable nutrient source have demonstrated that both wild-type and mutant cellular populations were initially operating in a sub-optimal metabolic state relative to the given nutrient conditions[30,31]. However, after undergoing adaptive evolution over a few hundred generations while keeping the nutrient environment stable, the end-point cellular populations were well characterized by FBA using maximal growth as the goal function[30,31].

B. Metabolic reaction activity

The majority of FBA investigations of metabolic networks have implemented optimization algorithms based on the simplex method[23] (Ref 32 is an exception). The simplex method evaluates the objective function along the vertices of the feasible flux cone in such a way that the objective function never decreases in value. However, the optimal solution is highly degenerate for these metabolic models[33,34]. The optimal flux states generated by the simplex method contain a low number of non-zero reaction fluxes (determining the *minimal* number

of reactions needed for sustaining growth is a mixed-integer linear problem[35]). Consequently, it is possible that metabolic reactions determined to have zero flux in one optimal solution by the simplex method could in fact carry a non-zero flux in an alternative optimal state. An interior-point linear programming algorithm[24], however, will find optimal solutions with the maximal number of reactions active since it traces through the interior of the flux cone, rather than along the flux-cone vertices. To consistently determine all the instances when the flux of a given reaction may be non-zero in the optimal state, it is hence necessary to implement an IP optimization approach.

A metabolic reaction i is defined as being active if it carries a non-zero flux. By changing the environmental conditions (*i.e.* the boundary constraint vector b), the whole-network optimal flux patterns shift to accommodate the new constraints. This flow-change is accomplished on the individual reaction level according to two mechanisms[36]. Reaction flux ν_i either displays flux-plasticity by adjusting its flux magnitude, or it displays structural-plasticity by going from an active (inactive) to an inactive (active) state. Sampling a large number N of possible environments, it is thus possible to estimate the activity fraction of a reaction i as $q_i = N_i/N$, where N_i is the number of environments for which i carried a non-zero flux. In the following, I will investigate the properties of the distribution of activities $P(q)$ and its possible biological implications.

Implementing the FBA approach for the metabolic models of the bacteria *H. pylori*[28] and *E. coli*[15], and the yeast *S. cerevisiae*[17], the calculated metabolic activity in response to $N = 30,000$ different boundary conditions demonstrated that there exists a metabolic core; a set of metabolic reactions that are active in all environments ($q_i = 1$ for the core reactions)[36]. The metabolic core is dominated by reactions that, on inhibition, renders the cell incapable of growth. Additionally, analysis of mRNA expression levels of the metabolic core enzymes show that they are highly synchronized and have significantly elevated half-life times[36]. Consequently, the set of $q = 1$ reactions may serve as a starting point in the search for new antimicrobial approaches[36].

In the following, I will compare the activity patterns obtained from calculating optimal fluxes for the *H. pylori*, *E. coli*, and *S. cerevisiae* metabolic models using both a simplex-based routine[37] and an interior-point predictor-corrector linear programming package[38]. For both methods, $N = 30,000$ randomly selected optimal conditions[36] were sampled and used as basis for the analysis. Briefly, each environmental condition is simulated by assigning

a fixed flux value $\nu \; \epsilon \; [0, \nu_{max}]$ to all the uptake-reaction fluxes, and we only analyze the conditions for which the optimal goal function (maximal cellular growth rate) is non-zero, using $v_{max} = 20$. Consequently, each condition potentially contains all carbon sources (albeit at varying accessibility) the cellular models are capable of extracting from the environment.

The interior-point algorithm generates flux distributions that potentially are qualitatively different from those of the simplex algorithm. Fig. 1 shows the calculated flux distribution for *H. pylori*, *E. coli*, and *S. cerevisiae* using an IP method[38]. The optimal flux distributions are in close agreement with observations from simplex-based calculations[36,39]. However, the activity pattern is distinctly different for the two approaches: From the previous discussion of the general properties of the simplex and the IP methods, it is expected that the IP flux distributions contain a larger number of active reactions. Using the simplex algorithm, the *H. pylori* average activity level $\langle q \rangle = \sum q_i P(q_i)$ for the metabolic system is $\langle q \rangle = 0.61$, while the IP approach yields the significantly higher activation of $\langle q \rangle = 0.74$. Similarly, there is a marked difference between the reaction activation levels for the simplex and the IP results in *E. coli* with 35% vs. 76%, and in yeast with 20% vs. 35% over the sample set.

The average activity level per condition $\langle q \rangle$ gives us insight into the overall network activity and the comparison between the simplex and the IP results sheds light on potential metabolic pathway redundancy. In Fig. 2, I show the distribution $P_S(q)$ of the activity values q for the three metabolic networks using simplex optimization. The "U"-shape of $P_S(q)$, with maxima at $q \approx 0$ and $q \approx 1$ and very low activity levels for intermediate values suggests an approximate binary classification of the reactions, where the majority are either members of the frequently activated or the rarely activated reaction sets. Note that the right-hand most bins in each panel include $q\epsilon(0.9, 1]$, whereas the metabolic core is strictly $q = 1$ reactions.

Fig. 3 shows the distribution of reaction activities $P_{IP}(q)$ when using IP linear programming. Not surprisingly, the distribution is shifted towards larger q-values. However, the $P_{IP}(q)$ distribution is now trimodal for *H. pylori*, and arguably so for *E. coli*, with the third peak appearing at activity values of $q = 0.55$ and $q = 0.75$ for the *H. pylori* and the *E. coli* models respectively. Consequently, the activity patterns generated by the IP optimization uncovers a set of metabolic reactions in the bacterial metabolism that are intermittently activated. The activity pattern for the *S. cerevisiae* model has also shifted towards larger activity values, while the shape of the distribution is still bimodal.

C. Metabolic activity and the high-flux backbone

It was recently demonstrated that the metabolic fluxes not only are highly heterogeneous on the organism level (Fig. 1), but also on the level of single metabolites[39]. This suggested that for most metabolites, it is possible to identify two reactions that dominate the production and consumption of the metabolite, respectively. A simple algorithm can be constructed to automatically uncover the locally maximal flow-paths, consisting of the following steps:[39]

1. For each metabolite, remove all reactions but the one carrying the largest incoming flux (production) ν_{in}^{max} and the one carrying the largest outgoing flux (consumption) ν_{out}^{max}.

2. For metabolites A and B being *educt* and *product*, respectively, in reaction R. Metabolites A and B, and reaction R are part of the HFB if $\nu_R = \nu_{A,out}^{max}$ and $\nu_R = \nu_{B,in}^{max}$.

3. For metabolites A and B being *product* and *educt*, respectively, in reaction R. Metabolites A and B, and reaction R are part of the HFB if $\nu_R = \nu_{A,in}^{max}$ and $\nu_R = \nu_{B,out}^{max}$.

The resulting sub-network corresponds to the cellular high flux backbone (HFB) of the metabolism, whose identity is specific to the given growth condition. As demonstrated in Ref. 39, the HFB is dominated by a giant connected component with a star-like topology which includes almost all metabolites produced under the given growth condition, and only a few pathways are disconnected. Although these pathways are parts of the HFB, their end product serves only as the second most important source (or less) for some other metabolite.

The HFB may serve as a novel network reduction algorithm that uncovers the parts of a metabolic network most actively involved in mass transport. It is highly non-trivial that groups of individual HFB reactions for the most part overlap with the traditional, biochemistry-based partitioning of cellular metabolism: *e.g.* for a simple oxic environment with glucose as the single carbon source, all metabolites of the citric acid cycle of *E. coli* are uncovered. Additionally, a considerable fraction of other important pathways are also uncovered, such as those involved in histidine-, murein- and purine biosynthesis. However, while the detailed nature of the HFB depends on the particulars of the growth condition, the HFB captures the subset of reactions that dominates metabolic activity for this condition. As such, it offers a complementary approach to elementary flux mode and extreme pathway analysis[40].

As the flux of the individual metabolic reactions sensitively depends on the growth conditions, it is necessary to investigate the response of the HFB to changes in the environment conditions. Analysis of the relationship between flux magnitude and HFB participation in Ref. 39 suggests that most reactions in the high flux region undergo noticeable flux changes, while reactions within the intermediate and low flux regions remain practically unaltered. Fig. 4 shows distribution of HFB participation f of the individual reactions over the dataset: For each organism, the HFB was separately generated for every condition, and the frequency with which a reaction i participated in the HFB is calculated as $f_i = M_i/N$. Here, M_i is the number of times reaction i was a part of the HFB, and N is the number of environments sampled. The average overall HFB participation is $\langle f \rangle = 0.46, 0.40$, and 0.28 for *H. pylori*, *E. coli*, and *S. cerevisiae*, respectively. Note that, the shape of the three distributions for $P(f)$ (Fig. 4) is similar to that of the activity $P_S(q)$: The metabolic reactions can be approximately divided into two groups; the first group consists of reactions carrying the locally maximal fluxes almost all the time, while the other group rarely contributes to the HFB.

The results of Fig. 4 suggest that there is a correlation between HFB participation f_i and the reaction activity q_i. Fig. 5 shows $\langle f(q) \rangle$ for each of the metabolic networks. The best fit to the function $y = ax$ is shown as dashed lines, with $a = 0.59, 0.49$, and 0.76, while the Pearson correlation of the pairs (q_i, f_i) is $r = 0.56, 0.50$, and 0.72 for *H. pylori*, *E. coli*, and yeast, respectively. The solid line with unit slope serves as a guide for the case of always HFB membership when active. Consequently, the slopes for the different organisms indicate the average propensity of reactions to be a part of the HFB for that organism.

III. ACTIVITY PATTERNS OF THE CENTRAL CARBON METABOLISM

Cellular metabolic functions and pathways are traditionally described in terms of either being catabolic, which primarily are related to the degradation of compounds into 12 precursor metabolites, and anabolic, the construction of cellular building blocks from the 12 precursors[41]. The central carbon metabolism is the collective description of the enzymatic reactions that synthesize these 12 precursors. Consequently, it may not be surprising that most organisms possess nearly complete copies of this "metabolic engine." The central carbon metabolism is further divided into the following four pathways: (i) Glycolysis (often called the Embden-Meyerhof-Parnas (EMP) pathway), which converts glucose-6-phosphate

(G6P) to pyruvate (PYR); (ii) the Tricarboxylic Acid (TCA) cycle (often also calle the citric acid cycle), which, in the presence of oxygen, converts acetyl-coA (AcCoA) to two carbon dioxid (CO_2) molecules, and in combination with glycolysis can produce up to 38 molecules of ATP (adenosine triphosphate) per glucose. Hence, the TCA cycle generates most of the energy needed for cellular growth; (iii) the Pentose-Phosphate (PP) pathway, which oxidizes G6P to CO_2; (iv) many bacteria also contain the Entner-Doudoroff (ED) pathway, which either may serve as a backup for the EMP pathway, or, as in a variety of Gram-negative bacteria such as the *Pseudomonads*, replaces it altogether.

Fig. 6 shows a simplified schematic of the metabolic reactions in the central carbon metabolism. The four major pathways have been highlighted. The activity fraction q_i is printed next to each of the enzymatic reactions, with the activity of *H. pylori* on top (blue), *E. coli* in the middle (green) and *S. cerevisiae* on the bottom (red). Missing enzymatic capability of an organism is indicated with "N/A".

The majority of the reactions in the EMP pathway are utilized in all the environments for the three organisms. However, for gluconeogenesis (the production of glucose from other carbon compounds such as pyruvate (PYR) or amino acids), which simply put is glycolysis backwards, the fructose bisphosphatase reaction (EC 3.1.3.11) FDP + H2O → F6P + PI is used in only 1% of the sampled yeast conditions, 76% of the *E. coli* conditions and in the vast majority (93%) of the *H. pylori* conditions.

As expected on grounds of its biochemical importance for the generation of a cell's energy currency, ATP, many of the reactions in the TCA cycle have activity levels close to unity. However, the reactions of the glyoxylate shunt (depicted in the center of TCA cycle in Fig. 6) show a varied activity pattern. One of the main purposes of the glyoxylate bypass is to replenish intermediates, such as dicarboxylic and tricarboxylic acid, which are usually provided by the TCA cycle. In Fig. 6, the glyoxylate bypass consists of the two reactions (i) ICIT → GLX + SUCC, and (ii) AcCoa + GLX + H20 → Coa + H + MAL. Although frequently used by *H. pylori* and *E. coli*, only reaction (i) is active in *S. cerevisiae*, with the humble activity level of $q = 0.01$. This difference in use pattern is likely caused by the much enhanced plasticity of the yeast metabolic network, which is related to its significantly larger size[36].

While the Pentose-Phosphate pathway has a highly uniform and almost identical use pattern among the three organisms, the ED pathway use is different in *E. coli* and *H. pylori*.

Unlike the EMP pathway, which has a theoretical net yield of two molecules of ATP per glucose, the ED pathway only results in one ATP. In *E. coli*, the ED pathway is always in use, while the *H. pylori* activity level is 61%. This result is particularly interesting, as it highlights a dramatic difference between the predicted optimal flux solutions obtained from simplex-based algorithms, and those obtained from interior-point methods: The ED pathway is never utilized in simplex solutions for the unperturbed *E. coli* model. Recent experimental flux measurements conducted on the *E. coli* central carbon metabolism using GC-MS with ^{13}C-labeled substrates[42] found that a significant flux passed through the ED pathway for wild-type cells. Although it is currently not known if or when flux distributions resulting from simplex-based methods are biologically more relevant than those from IP-based methods, this observation suggests that, using current technologies, it is possible to devise experiments to address this question.

IV. CONCLUSION

Successful metabolic function is central to the survival and proliferation of all cellular organisms. Microorganisms such as bacteria have developed an astonishing capability to utilize the often scant availability of nutrients in their environment. In this paper, I have presented results on metabolic activity patterns of three organism-level metabolic networks using the flux-balance optimization framework. In contrast to the majority of previous studies of cellular metabolic networks, the linear programming algorithm is based on an interior-point point method and not the simplex algorithm. This has made it possible to systematically investigate individual reaction activity patterns despite the large number of degenerate optimal solutions.

The metabolic flux distribution on the three metabolic networks is heavy-tailed, consistent with previous results based on simplex linear programming and the non-optimal hit-and-run method[39]. The reaction activity pattern suggests that the majority of the reactions are activated in an approximately binary fashion where they are either almost always in use or very rarely carry flux. The activity q_i of a reaction shows an interesting correlation with the reaction's participation f_i in the metabolic high-flux backbone, with a Pearson correlation between q_i and f_i for *S. cerevisiae* of $r = 0.72$. When analyzed in their biochemical context, the calculated reaction-activity patterns offer further insight into the organization

of metabolism.

However, the organism-level metabolic networks are numerically very challenging. Although interior-point optimization methods should return strictly complementary solutions, numerical instabilities result in degenerate solutions[43] where some primal and dual variable pairs are returned as zero or below the accuracy threshold. The effect of this inherent numerical difficulty of the cellular metabolic network models is that the IP method may underestimate the number of possible active reactions. Recent advances in optimization algorithms, in particular the balanced solution approach[44], offer promising avenues for further studies.

Acknowledgments

---

[*] Electronic address: `almaas@llnl.gov`

[1] A. Pandey and M. Mann, Nature **405**, 837 (2000).

[2] H. Caron, B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. Hermus, R. van Asperen, K. Boon, P. Voute, et al., Science **291**, 1289 (2001).

[3] C. Burge, Nat. Genet. **27**, 232 (2001).

[4] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 61 (2002).

[5] N. Guelzim, S. Bottani, P. Bourgine, and F. Kepes, Nat. Genet. **31**, 60 (2002).

[6] N. M. Luscombe, M. M. Babu, H. Y. Yu, M. Snyder, S. A. Teichmann, and M. Gerstein, Nature **431**, 308 (2004).

[7] A. Ma'ayan, S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, et al., Science **309**, 1078 (2005).

[8] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, Nucleic Acids Res. **34**, D535 (2006), URL `http://www.thebiogrid.org`.

[9] S. McCraith, T. Holtzman, B. Moss, and S. Fields, Proc. Natl. Acad. Sci. USA **97**, 4879 (2000).

[10] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, et al., Nature **403**, 623 (2000).

[11] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc Natl Acad Sci USA **98**, 4569 (2001).

[12] A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, n. A. M. Micho, C. M. Cruciat, et al., Nature **415**, 141 (2002).

[13] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, et al., Nature **415**, 180 (2002).

[14] P. D. Karp, C. A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahren, S. Tsoka, N. Darzentas, V. Kunin, , et al., Nucleic Acids Res. **19**, 6083 (2005), URL `http://www.biocyc.org`.

[15] J. S. Edwards and B. O. Palsson, Proc. Natl. Acad. Sci USA **97**, 5528 (2000).

[16] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson, J. Bacteriol. **184**, 4582 (2002).

[17] N. C. Duarte, M. J. Herrgard, and B. O. Palsson, Genome. Res. **14**, 1298 (2004).

[18] S. H. Strogatz, Nature p. 268 (2001).

[19] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[20] M. E. J. Newman, SIAM Review **45**, 167 (2003).

[21] P. W. Anderson, Science **177**, 393 (1972).

[22] H. Kitano, Nature **420**, 206 (2002).

[23] G. B. Dantzig, *Linear Programming and Extensions.* (Princeton University Press, Princeton, New Jersey, 1963).

[24] Y. Ye, *Interior point algorithms : theory and analysis.* (John Wiley & Sons, New York, NY, 1997).

[25] K. J. Kauffman, P. Prakash, and J. S. Edwards, Curr. Opin. Biotechnol. **14**, 491 (2003).

[26] N. D. Price, J. L. Reed, and B. O. Palsson, Nat. Rev. Microbiol. **2**, 886 (2004).

[27] H. P. J. Bonarius, G. Schmid, and J. Tramper, Trends Biotechnol. **15**, 308 (1997).

[28] I. Thiele, T. D. Vo, N. D. Price, and B. O. Palsson, J. Bact. **187**, 5818 (2005).

[29] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, **19**, 125 (2001).

[30] R. Ibarra, J. Edwards, and B. Palsson, Nature **420**, 186 (2002).

[31] S. Fong and B. Palsson, Nature Genet. **36**, 1056 (2004).

[32] S. Bell and B. Palsson, Comput. Chem. Eng. **29**, 481 (2005).

[33] S. Lee, C. Phalakornkule, M. M. Domach, and I. E. Grossmann, Comp. Chem. Eng. **24**, 711 (2000).

[34] J. L. Reed and B. O. Palsson, Genome Res. **14**, 1797 (2004).

[35] A. P. Burgard, S. Vaidyaraman, and C. D. Maranas, Biotechnol. Progr. **17**, 791 (2001).

[36] E. Almaas, Z. N. Oltvai, and A.-L. Barabási, PLoS Comput. Biol. **1**, e68 (2005).

[37] A. Makhorin, *GLPK – GNU Linear Programming Kit.* (2006), URL `http://www.gnu.org/software/glpk/glpk.html`.

[38] J. Czyzyk, S. Mehrotra, M. Wagner, and S. Wright, *PCx User Guide (Version 1.1), Techical Report OTC 96/01*, Optimization Technology Center, Argonne National Laboratory (1997), URL `http://www-fp.mcs.anl.gov/otc/Tools/PCx/`.

[39] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A.-L. Barabási, Nature **427**, 839 (2004).

[40] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, Trends Biotechnol. **22**, 400 (2004).

[41] M. T. Madigan and J. M. Martinko, *Brock Biology of Microorganisms.* (Prentice Hall, Upper Saddle River, New Jersey, 2006).

[42] E. Fischer and U. Sauer, Eur. J. Biochem. **270**, 880891 (2003).

[43] E. Almaas, K. Livingstone, and A. Holder, in *Optimization in Medicine and Biology*, edited by G. J. Lim (Francis & Taylor, To be published).

[44] M. Ehrgott, A. Holder, and J. Reese, Tech. Rep. 95, Trinity University Mathematics, San Antonio, TX (2005).
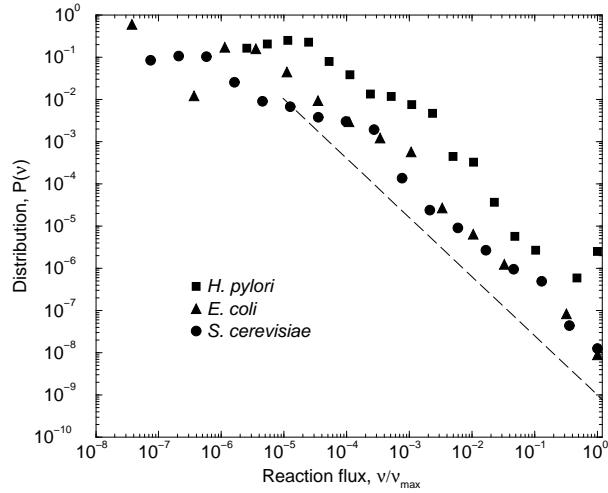
FIG. 1: Distribution of metabolic fluxes in the bacteria *H. pylori* and *E. coli*, and the yeast *S. cerevisiae* from optimization of growth using an interior-point linear programming method[38]. These results are consistent with the simplex-based optimal flux distribution of Refs. 39. The dashed line serves as a guide with power-law exponent $\alpha = 1.5$.
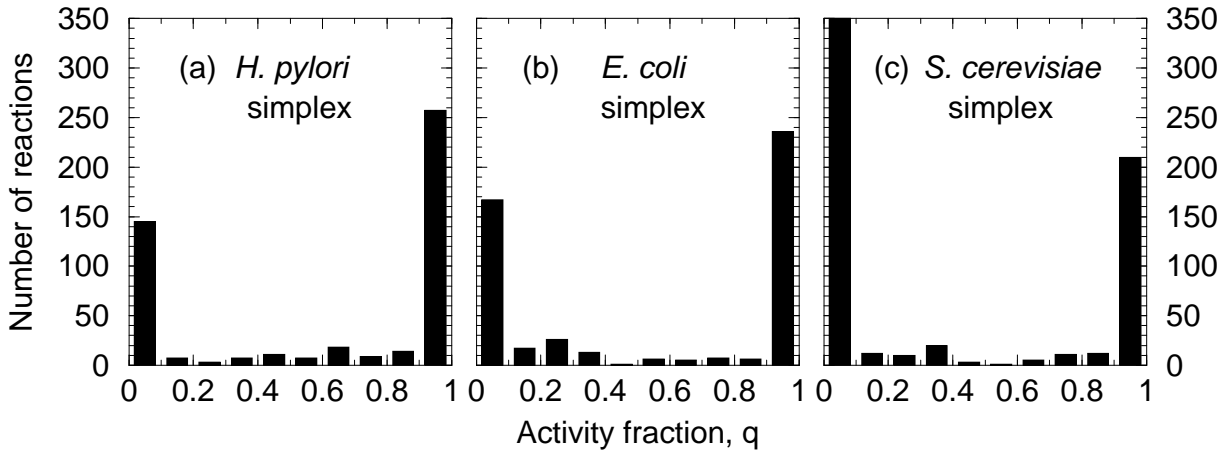
FIG. 2: Distribution of the reaction activity q for the simplex optimization approach. The activity of reaction i is $q_i = N_i/N$, the number of times reaction i carries a non-zero flux over the number of sampled environments. The "U"-shape of the *H. pylori*, *E. coli* and *S. cerevisiae* activity distributions suggests that there exists two types of reactions, independent of the metabolic network, in simplex-optimal flux states: a reaction is either almost always used or mostly inactive. The number of reactions in the first interval in panel (c) is 866.
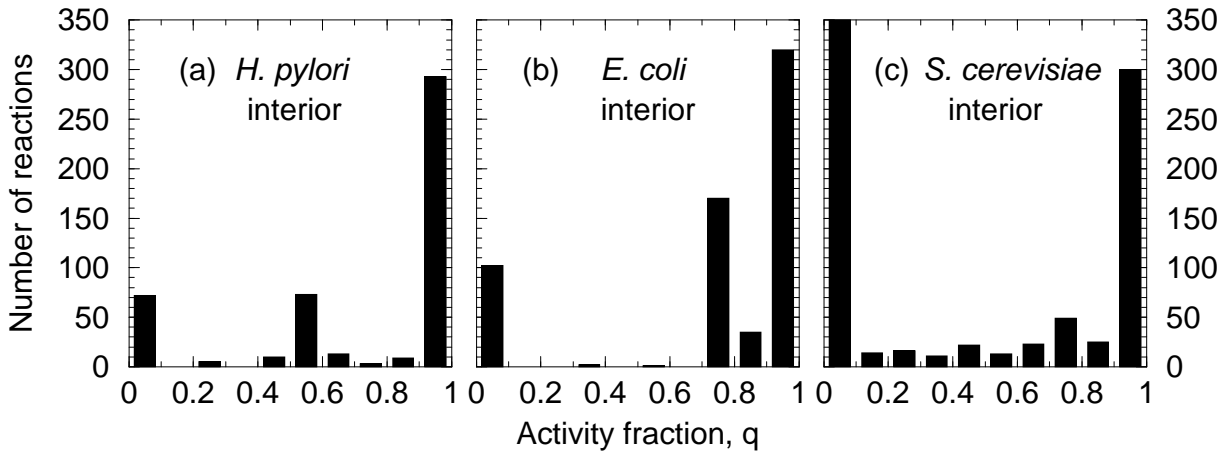

FIG. 3: Distribution of the reaction activity q for the interior-point (IP) optimization approach. Using an IP optimization method, reaction activity patterns are more complex compared to the simplex results (Fig. 2). The number of reactions in the first interval in panel (c) is 677.
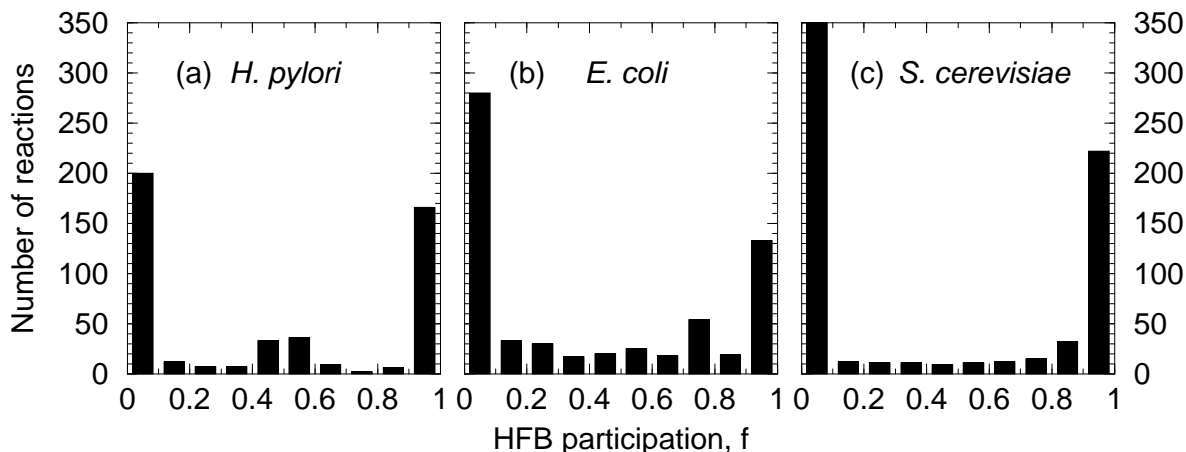
FIG. 4: Distribution of the high-flux backbone participation f for the IP optimization approach. The HFB participation of reaction i is $f_i = M_i/N$, the number of times reaction i is member of the HFB over the number of sampled environments. The approximate "U"-shape of the *H. pylori*, *E. coli* and *S. cerevisiae* $P(f)$ distributions suggests that there exists two types of reactions, independent of metabolic network: a reaction either almost always or almost never carries a locally maximal flux. The number of reactions in the first interval in panel (c) is 815.
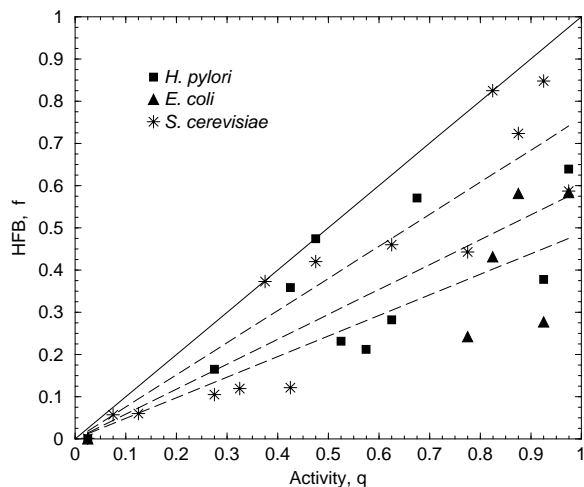

FIG. 5: Relation between average activity q and average high-flux backbone membership f. For each of the $N = 30,000$ sampled conditions we construct the HFB, and f_i is the frequency with which reaction i is an HFB member. The dashed lines indicate the best linear regression of form $y = ax$ with $a = 0.59$, 0.49, and 0.76 for *H. pylori*, *E. coli*, and *S. cerevisiae* respectively, and the solid line ($a = 1$) serves as a guide.
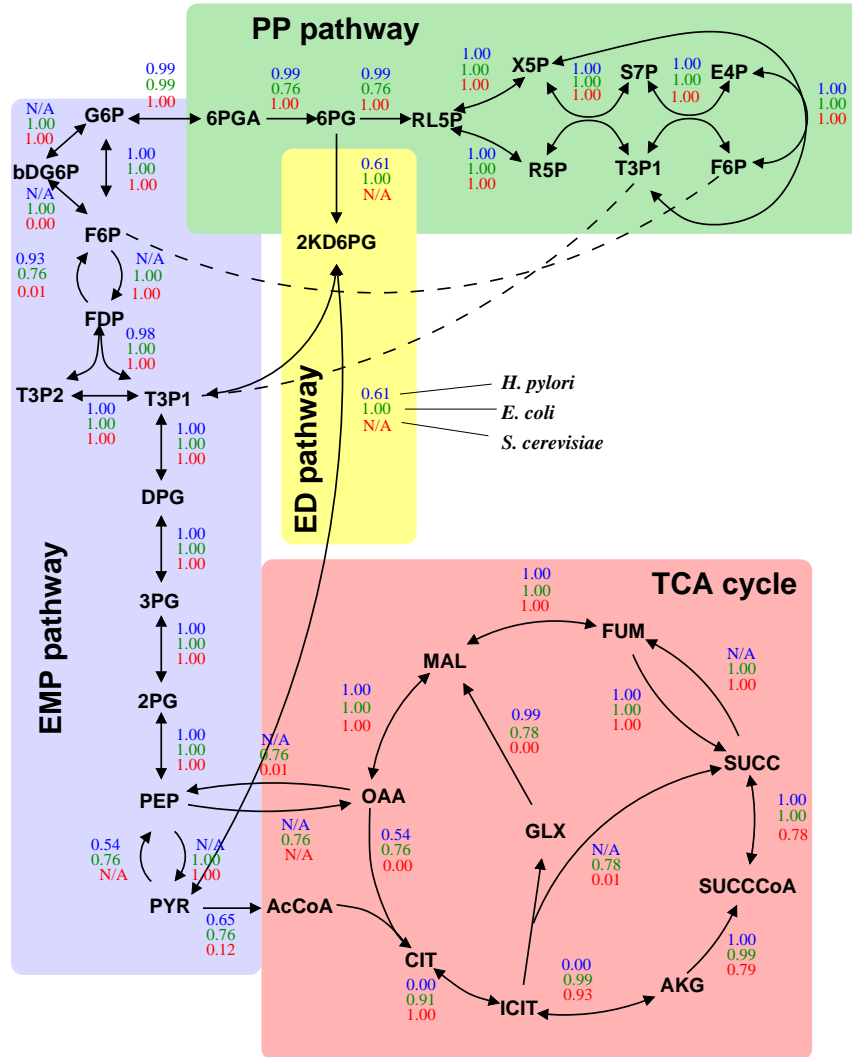
17

FIG. 6: A schematic of the central carbon metabolism, including Embden-Meyerhof-Parnas (EMP, blue background), the Tricarboxylic Acid Cycle (TCA, red), the Pentose-Phosphate pathway (PP, green), and the Entner-Doudoroff pathway (ED, yellow). The activity q_i is written next to each reaction, where the upper, middle, and lower numbers correspond to *H. pylori* (blue), *E. coli* (green), and *S. cerevisiae* (red) respectively. "N/A" indicates that an organism lacks that enzymatic capability.