OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

The Activity Reaction Core and Plasticity of Metabolic Networks

Eivind Almaas[1,2], Zoltán N. Oltvai[3*], Albert-László Barabási[2,4]

1 Microbial Systems Division, Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, United States of America, **2** Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, Indiana, United States of America, **3** Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America, **4** Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts, United States of America

Understanding the system-level adaptive changes taking place in an organism in response to variations in the environment is a key issue of contemporary biology. Current modeling approaches, such as constraint-based flux-balance analysis, have proved highly successful in analyzing the capabilities of cellular metabolism, including its capacity to predict deletion phenotypes, the ability to calculate the relative flux values of metabolic reactions, and the capability to identify properties of optimal growth states. Here, we use flux-balance analysis to thoroughly assess the activity of *Escherichia coli, Helicobacter pylori,* and *Saccharomyces cerevisiae* metabolism in 30,000 diverse simulated environments. We identify a set of metabolic reactions forming a connected metabolic core that carry non-zero fluxes under all growth conditions, and whose flux variations are highly correlated. Furthermore, we find that the enzymes catalyzing the core reactions display a considerably higher fraction of phenotypic essentiality and evolutionary conservation than those catalyzing noncore reactions. Cellular metabolism is characterized by a large number of species-specific conditionally active reactions organized around an evolutionary conserved, but always active, metabolic core. Finally, we find that most current antibiotics interfering with bacterial metabolism target the core enzymes, indicating that our findings may have important implications for antimicrobial drug-target discovery.

Citation: Almaas E, Oltvai ZN, Barabási A (2005) The activity reaction core and plasticity of metabolic networks. PLoS Comput Biol 1(7): e68.

Introduction

Constraint-based modeling approaches, such as flux-balance analysis (FBA) [1,2], have proved highly successful in analyzing the capabilities of cellular metabolism, including its capacity to predict deletion phenotypes, the ability to calculate the relative flux values of metabolic reactions, and the capability to identify properties of alternate optimal growth states in a wide range of simulated single-carbon-source environmental conditions [3–5]. Recent analyses also indicate that the deletion phenotype of a substantial subset of metabolic enzymes is growth-condition-dependent [6], arguing for a selective use of metabolic reactions in distinct environments.

Results

To thoroughly examine the utilization and relative flux rates of each metabolic reaction in a wide range of simulated environmental conditions, we sampled 30,000 randomly and uniformly chosen optimal growth conditions, as well as all single-carbon-source minimal medium conditions sufficient for growth, using FBA [1,2] on the reconstructed metabolic networks of *Helicobacter pylori*, *Escherichia coli*, and *Saccharomyces cerevisiae* [7–9] (see Materials and Methods). We found that when assuming optimal growth of the three microorganisms, their metabolism adapts to environmental changes through two distinct mechanisms. The more common mechanism is flux plasticity, involving changes in the fluxes of already active reactions when the organism is shifted from one growth condition to another. For example, changing from glucose- to succinate-rich media alters the flux of 264 *E. coli* reactions by more than 20%. Less commonly, environmental changes can also induce structural plasticity, resulting in changes in the metabolism's active wiring diagram, turning on previously zero-flux reactions and inhibiting previously active pathways. For example, when shifting *E. coli* cells from glucose- to succinate-rich media, 11 previously active reactions are turned off completely, while nine previously inactive reactions are turned on.

The two types of response mechanisms described by flux and structural plasticity imply the possible existence of a group of reactions that are not subject to structural plasticity, being active under all environmental conditions. Indeed, some metabolic reactions were found to carry non-zero fluxes in *S. cerevisiae* under nine different growth conditions [6]. In itself, however, this finding is compatible with a scenario in which the active reactions are randomly distributed: if typically a q fraction of the metabolic reactions are active under a specific growth condition, for n distinct conditions one can predict a non-zero overlap encompassing at least q^n fraction of the reactions. Indeed, we find that as we increase the number of inspected conditions, n, the number of

Received June 27, 2005; Accepted November 2, 2005; Published December 16, 2005
DOI: 10.1371/journal.pcbi.0010068

Abbreviation: FBA, flux-balance analysis

Editor: Daniel Segre, Boston University, United States of America

* To whom correspondence should be addressed. E-mail: oltvai@pitt.edu

A previous version of this article appeared as an Early Online Release on November 2, 2005 (DOI: 10.1371/journal.pcbi.0010068.eor).

Synopsis

Although cellular metabolism is among the most investigated cellular functions, it is not well understood how it changes and adapts on a systems level in response to environmental variations. In this study, the authors take advantage of the reconstructed metabolic networks of *Helicobacter pylori, Escherichia coli,* and *Saccharomyces cerevisiae* and the computational method of flux-balance analysis to investigate the functional plasticity of these three metabolic networks for a large number of simulated growth conditions.

Within this approach, the authors identify the metabolic cores of these organisms. These metabolic cores represent connected sets of reactions that are used in all tested environments. When cross-correlating the predicted cores with experimental data, the authors find that the cores display a significantly higher fraction, both of essential and evolutionary conserved enzymes, than their noncore counterparts. Additionally, the extended mRNA half-lives of the core enzymes give further support to the notion that core reactions represent main integrators of metabolic activity.

Since most of the metabolic core reactions are indispensable for the growth of the microorganism, and several existing antibiotics target select bacterial enzymes in the core, the authors suggest that the metabolic core may have important applications for antimicrobial drug-target discovery.

reactions carrying non-zero flux in each condition decreases rapidly. However, it eventually saturates at a constant value (Figure 1A–C), identifying a group of reactions that have non-zero flux under *all* 30,000 simulated growth conditions. Specifically, we find that in 138 of 381 *H. pylori* (36.2%), 90 of 758 *E. coli* (11.9%), and 33 of 1,172 *S. cerevisiae* (2.8%), metabolic reactions are always active (Figure 1D and 1E). While these reactions respond to environmental changes only

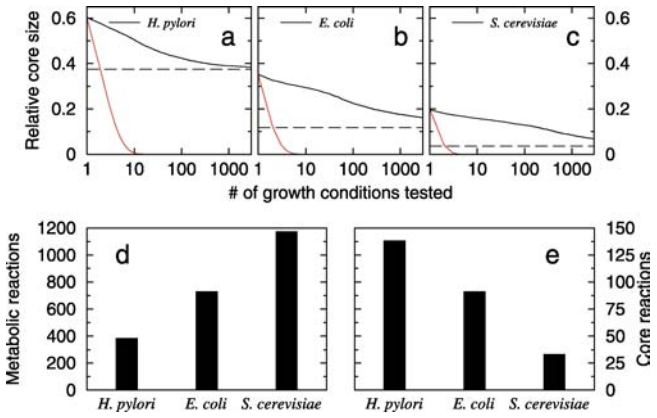

Figure 1. The Emergence of the Metabolic Core

(A–C) The average relative size of the number of reactions that are always active as a function of the number of sampled conditions (black line) for (A) *H. pylori,* (B) *E. coli,* and (C) *S. cerevisiae.* As the number of conditions increases, the curve converges to a constant denoted by the dashed line, identifying the metabolic core of an organism. The red line denotes the number of reactions that are always active if activity is randomly distributed in the metabolic network. The fact that it converges to zero indicates that the real core represents a collective network effect, forcing a group of reactions to be active in all conditions. (D and E) The number of metabolic reactions (D) and the number of metabolic core reactions (E) in the three studied organisms.

DOI: 10.1371/journal.pcbi.0010068.g001

through flux-based plasticity, the rest of the reactions are only conditionally active, being turned on only in specific growth conditions and thus being subject to both structural- and flux-based plasticity.

Figure 2 displays those metabolic reactions of *E. coli* that remain active in all 30,000 simulated growth conditions. The striking feature of this diagram is the fact that these reactions form a single connected cluster, encompassing each of the 90 reactions. This is not a unique feature of *E. coli;* in *H. pylori,* all 138 reactions form a single cluster, and in *S. cerevisiae* all 33 reactions are connected. Given the relatively low number of reactions that are always active, the likelihood that they form a single large cluster by chance alone is negligible, with $p < 1e^{-6}$ for *H. pylori* and *E. coli,* and $p \sim 2e^{-6}$ for *S. cerevisiae.* Given the compact and clustered nature of the group of reactions that are always active, we will refer to them collectively as the metabolic core.

The metabolic core contains two types of reactions: The first type consists of those that are essential for biomass formation under all environmental conditions (81 out of 90 reactions in *E. coli*), while the second type of reaction is required only to assure optimal metabolic performance (Table S1). In case of the inactivation of the second type, alternative sub-optimal pathways can be used to ensure cellular survival. Interestingly, in the compact core of *S. cerevisiae,* all 33 reactions were predicted to be indispensable for biomass formation under all growth conditions. Moreover, when assuming a 10% reduction in the growth rate (compared to optimal growth), the size and identity of the *E. coli* metabolic core remains largely unchanged, retaining 83 of the original 90 reactions, of which two are nonessential. This indicates that the concept of the metabolic core is valid under both optimal and suboptimal growth, although with some difference in the identity of individual reactions. Of note, the metabolic core represents a subset of the minimal reaction sets [10,11] and the overlap of alternative (i.e., degenerate) minimal reaction sets [4,5]. The minimal reaction set of [11] contains the metabolic core in addition, however, to reactions necessary for the sustained growth on any chosen substrate, whereas the minimal reaction set of [5] consists of the 201 reactions that are always active in *E. coli* for all 136 aerobic and anaerobic single-carbon-source minimal environments capable of sustaining optimal growth. The latter finding is in excellent agreement with Figure 1B (upper curve), and demonstrates that the analysis of a low number of growth conditions will significantly overestimate the set of reactions that are always active.

To identify some of the factors that determine the size of the metabolic core, we note that the number of core metabolic reactions systematically decreases as we move from *H. pylori* to *S. cerevisiae* (Figures 3A and 1E), the relative size of the core decreasing from 36.2% of all reactions in *H. pylori* to 11.9% in *E. coli* and 2.8% in *S. cerevisiae.* This trend can be explained as a collective network effect: the relatively small size of the *H. pylori* metabolic network leaves little flexibility for biomass production, requiring the continuous activity of a high fraction of the available metabolic reactions. Indeed, we find that on average approximately 61% of the *H. pylori* reactions are active in a given environmental condition. The larger number of metabolic reactions present in *E. coli* offers a higher degree of metabolic flexibility, allowing for a significant fraction of the biomass to be produced by

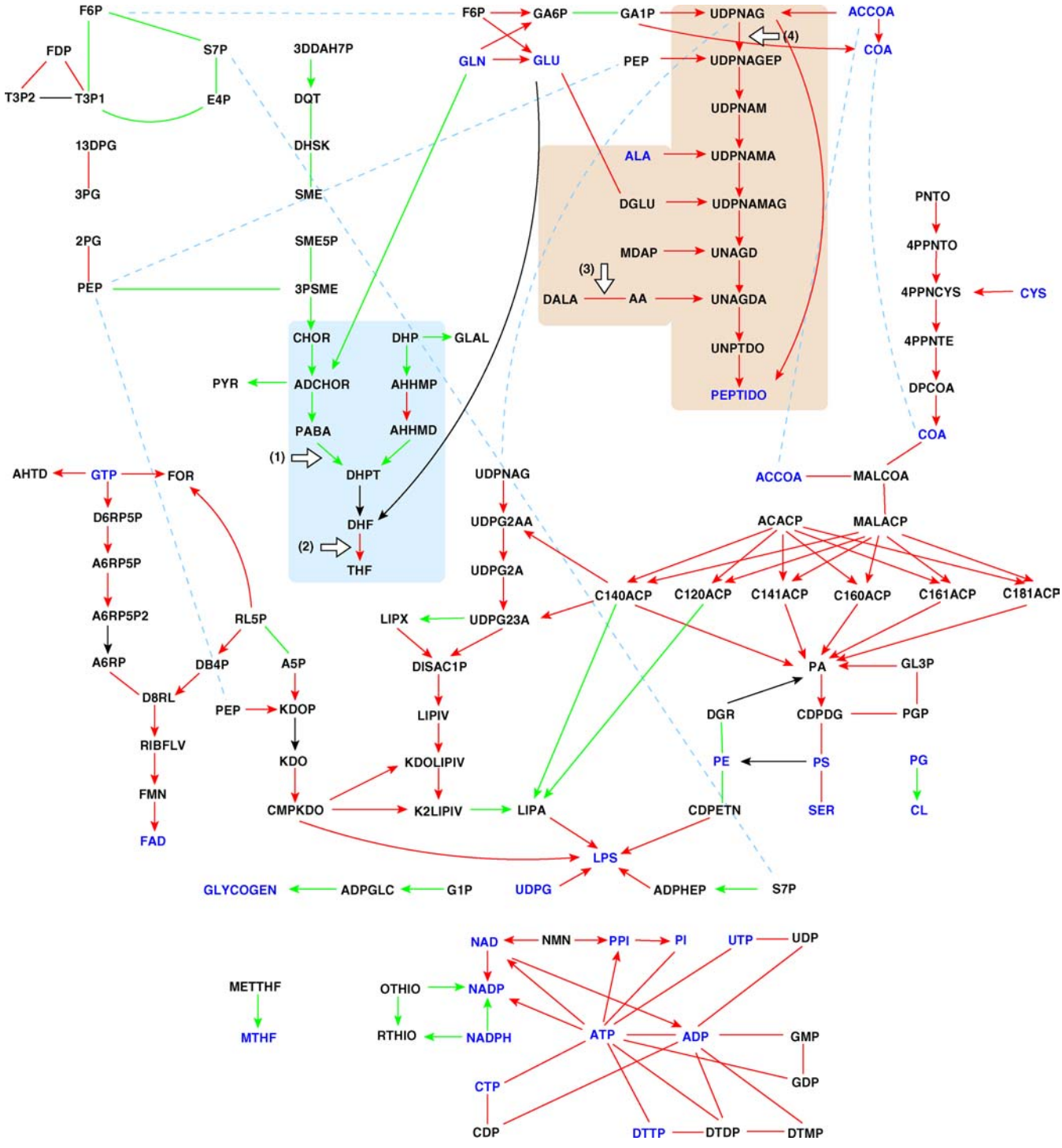

Figure 2. The Metabolic Core of *E. coli*

All reactions that are found to be active in each of the 30,000 investigated external conditions are shown. Metabolites that contribute directly to biomass formation [8] are colored blue, while core reactions (links) catalyzed by essential (or nonessential) enzymes [13] are colored red (or green). (Black-colored links denote enzymes with unknown deletion phenotype.) The blue dashed lines indicate multiple appearances of a metabolite, while links with arrows denote unidirectional reactions. Note that 20 out of the 51 metabolites necessary for biomass synthesis are not present in the core, indicating that they are produced (or consumed) in a growth-condition-specific manner. See Protocol S1 and Table S1 for the abbreviations of metabolites and a list of core reactions for *E. coli, H. pylori,* and *S. cerevisiae.* The folate and peptidoglycan biosynthesis pathways are indicated by blue and brown shading, respectively, and the white numbered arrows denote current antibiotic targets inhibited by: (1) sulfonamides, (2) trimethoprim, (3) cycloserine, and (4) fosfomycin. Note that a few reactions appear disconnected since we have omitted the drawing of cofactors.

DOI: 10.1371/journal.pcbi.0010068.g002

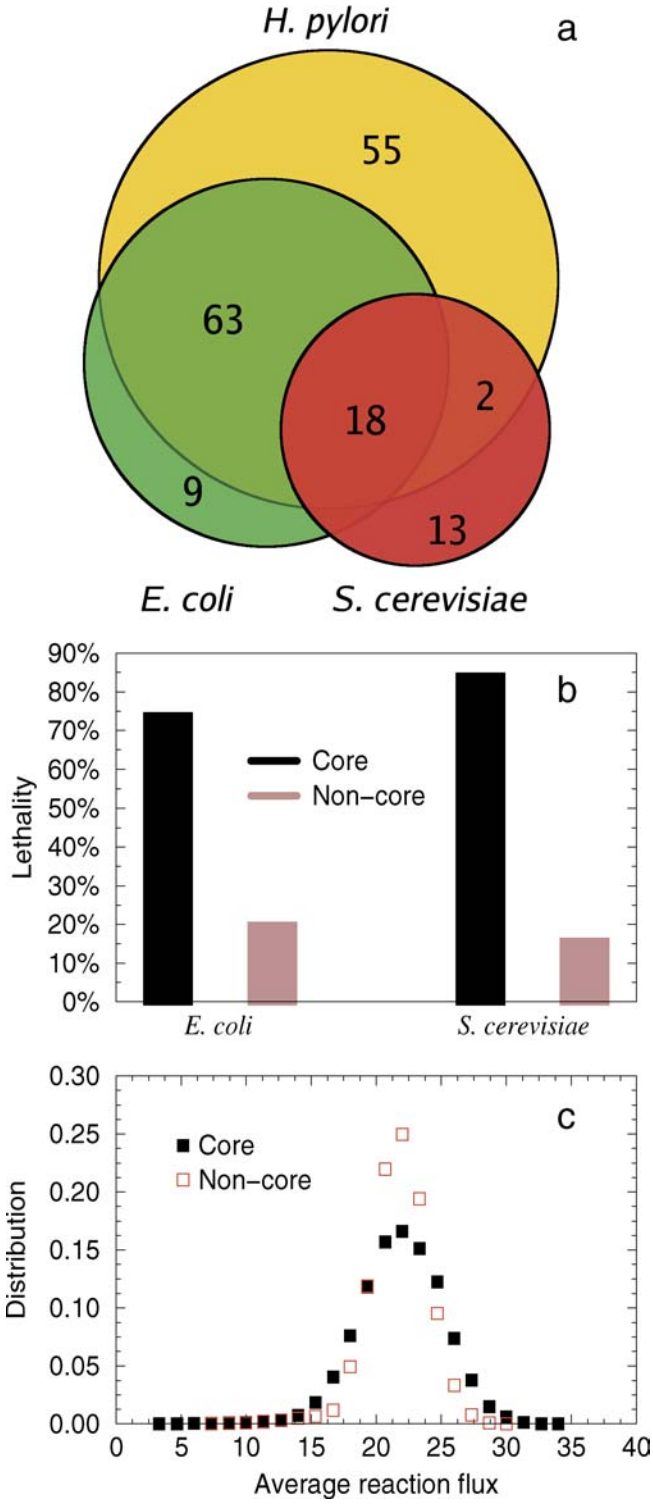

Figure 3. Characterizing the Metabolic Cores

(A) The number of overlapping metabolic reactions in the metabolic core of *H. pylori, E. coli,* and *S. cerevisiae.*

(B) The fraction of metabolic reactions catalyzed by essential enzymes in the cores (black) and outside the core in *E. coli* and *S. cerevisiae.*

(C) The distribution of average metabolic fluxes for the core and the noncore reactions in *E. coli.*

DOI: 10.1371/journal.pcbi.0010068.g003

alternative pathways. Indeed, the average utilization of the *E. coli* metabolic network in a given growth condition is only 35.3%. For *S. cerevisiae,* whose metabolic network size significantly exceeds that of both *H. pylori* and *E. coli,* there is an even higher metabolic flexibility, and the activity of only 19.7% of the reactions are required in a typical environment.

The fact that the core reactions are active under all investigated environmental conditions suggests that they must play a key role in maintaining the metabolism's overall functional integrity. Therefore, the absence of individual reactions that are part of the metabolic core may lead to significant metabolic disruptions. Indeed, using genome-scale deletion-phenotype data obtained in rich growth media [12,13], we find that 74.7% of those *E. coli* enzymes that catalyze core metabolic reactions (i.e., core enzymes) are essential, compared with a 19.6% lethality fraction characterizing the noncore enzymes. The gap between the core and noncore enzymes is also significant for *S. cerevisiae,* for which essential enzymes catalyze 84% of the core reactions, whereas the average essentiality of the conditionally active enzymes is only 15.6% (Figure 3B). Note that the likelihood of a random concentration of so many essential enzymes in the core is extremely small, with p-values of $3.3e^{-23}$ and $9.0e^{-13}$ for *E. coli* and yeast, respectively, even if we presume the presence of errors in the deletion-phenotype data (Protocol S1). Taken together, these results indicate that an organism's ability to adapt to changing environmental conditions rests, to a large extent, on the continuous activity of the metabolic core, regardless of the environmental conditions.

Intuitively, one could assume that the core represents a subset of high-flux reactions characterizing the activity of metabolic networks [3]. However, our measurements indicate that, on average, the fluxes of the core and noncore reactions are highly comparable (Figure 3C). Alternatively, we could also assume that the main biological role of the metabolic core is to ensure the continuous production of biomass under all growth conditions. In contrast, we find that 20 out of the 51 metabolites (39%) considered necessary for biomass production in *E. coli* [8] are not produced by any of the core reactions; instead—in a growth-condition-dependent fashion—they are produced by various alternative metabolic pathways. The core, however, contains a large number of reactions for selected anabolic pathways, including those of membrane lipid-, cell envelope-, and peptidoglycan-biosynthesis pathways (Table S1). These core reactions represent network bottlenecks, being the only paths for the synthesis of certain biomass components. Therefore, it appears that the composition of the metabolic core is determined by two factors. First, those metabolic reactions that directly contribute to biomass production tend to be part of the metabolic core of the organism. Second, however, this tendency is offset by network-induced redundancy: reactions or pathways whose end-products can be synthesized by at least two alternative pathways show environmental redundancy and structural plasticity, and are thus eliminated from the core. Therefore, the more reactions a metabolic network possesses (see Figure 1D), the stronger is the network-induced redundancy, and the smaller is the core (Figure 1E).

Given the important functional role played by the metabolic core in a given organism, one would expect significant parts of the core to be conserved in different

organisms. Indeed, the *E. coli* and *H. pylori* cores have 63 reactions in common, and 18 of the 33 core reactions in *S. cerevisiae* are present in both the *E. coli* and the *H. pylori* metabolic cores (see Figure 3A). Also, when considering enzyme orthologs among 32 divergent bacteria [13], we find that the metabolic core enzymes of *E. coli* display a high degree of evolutionary conservation, the average core enzyme having orthologs in 71.7% of the reference bacteria ($p < 1e^{-6}$). In contrast, the conditionally active noncore enzymes have an evolutionary retention of only 47.7% [13]. Given the observed correlation between evolutionary retention and the essentiality of gene products [13], this difference may be a simple consequence of the high-lethality fraction of the core enzymes. However, further analysis indicates that this is not the case: random selection of 90 enzymes with a 74.7% lethality ratio has an average evolutionary retention of only 63.4%. Taken together, these results indicate that major portions of the metabolic core have been conserved, displaying a higher evolutionary retention than the individual essentiality of its participating enzymes would indicate, suggesting that maintaining the core's integrity is a collective need of the organism.

The requirement for the continuous activity of the core reactions may well also impact the regulation of its catalytic enzymes. We would expect that the activity of core reactions—for which only the flux magnitude needs to be modulated—should display a higher degree of stability and a different regulatory control than the reactions outside of the core. Evidence for such regulatory effects is provided by two measurements. First, we inspected the experimentally determined mRNA half-lives of the *E. coli* metabolic enzymes when cells were grown in Luria-Bertani medium [14]. We found that the distribution of mRNA half-life times for the core enzymes displays a shift to higher values, with the average half-life time of their mRNAs being 14.0 min compared with 10.5 min for the noncore mRNAs ($p = 0.016$). Furthermore, only 2% of the core enzymes have corresponding mRNA half-life times shorter than 5 min, compared to 23% for the ones catalyzing conditionally active noncore reactions (see Materials and Methods). Therefore, the continuous metabolic need for these reactions has apparently affected the mechanisms responsible for mRNA decay of their catalytic enzymes, providing a higher dynamic inertia under environmental changes.

Additionally, for each enzyme-encoding operon, we counted the number of activating and repressive regulatory links in the *E. coli* transcriptional regulatory network [15,16]. As regulatory interactions are currently known for only 13 of the core enzyme-encoding operons, we considered the extended core, representing the set of 234 reactions that are active in more than 90% of the 30,000 simulated growth conditions. The results indicate that the fraction of repressive regulatory links in the extended core is 52.3%, while the fraction of activating interactions is only 35.7%, and the remaining 12% represents regulatory links that can either activate or suppress the enzyme's mRNA synthesis rate. In contrast, for noncore enzyme-encoding operons, there is no difference between the fraction of activating and repressing links, both representing 45% of the regulatory interactions. These results offer evidence for the co-evolution of the core metabolic network and its corresponding regulatory network: given the requirement for the continuously active non-redundant core, regulatory mechanisms have shifted both the mRNA decay rate and the nature of the regulatory interactions, offering a higher regulatory stability for the core enzymes.

The finding that the core reactions form a single cluster would suggest that the activity of the participating reactions is highly synchronized, with changes in the flux of one reaction affecting the flux of other reactions as well. Thus, we should be able to discover the core by inspecting the correlations between the activities of all metabolic reactions. To test this hypothesis, we calculated the flux correlation coefficient, C_{ij}, for each reaction pair in the *E. coli* metabolism [17] by inspecting the flux values of all reactions under each of the 30,000 simulated growth conditions. Using the C_{ij} matrix as the metric of a hierarchical clustering algorithm [18], we observed the emergence of a group of reactions whose fluxes change simultaneously under environmental shifts. Interestingly, these highly correlated reactions significantly overlap with the metabolic core (Figure 4A), indicating that when the environmental conditions are altered, the fluxes of the core reactions are increased or decreased in a synchronized fashion.

The same synchrony is observed when we inspect experimental mRNA data among the core reactions during changes in environmental conditions. Calculating the correlations between core and noncore enzymes using their corresponding mRNA copy number data (see Materials and Methods) for 41 experiments [19], we find that the correlations among mRNA copy numbers of the core enzymes are systematically higher than the observed correlations among their noncore counterparts (Figure 4B), with an average correlation of $<C> = 0.23$ for core enzymes and $<C> = 0.07$ for noncore enzymes ($p < 1e^{-4}$). Notably, this finding is different from the average correlation of mRNA expression found in 66 "correlated reaction sets" (groups of between two and nine reactions that are turned on and off together), where the frequency with which a correlated reaction set is used does not affect its mRNA correlation [5].

Discussion

Previous studies have firmly established that environmental changes induce flux plasticity, altering the flux levels of individual reactions (reviewed in [20]). Similarly, the fact that metabolism displays structural plasticity, turning on and off some reactions as the growth conditions are altered, has been observed before [3,4,6]. However, our demonstration of a group of reactions predicted to be active in all environmental conditions and forming a connected metabolic core could substantially improve our understanding of the organization and utilization of metabolic networks. The emergence of the core represents a collective network effect, channeling the production of some indispensable biomass components to a few key reactions that cannot be replaced by alternative pathways under any environmental conditions. The collective origin of the core is supported by the observed changes in the core size. While the number of biomass components that the metabolism needs to produce is comparable in the three organisms (49 for *H. pylori*, 51 for *E. coli*, and 44 for *S. cerevisiae*), the number of metabolic reactions contributing to them differs. The larger size of a metabolic network significantly increases its capacity and redundancy, decreas-

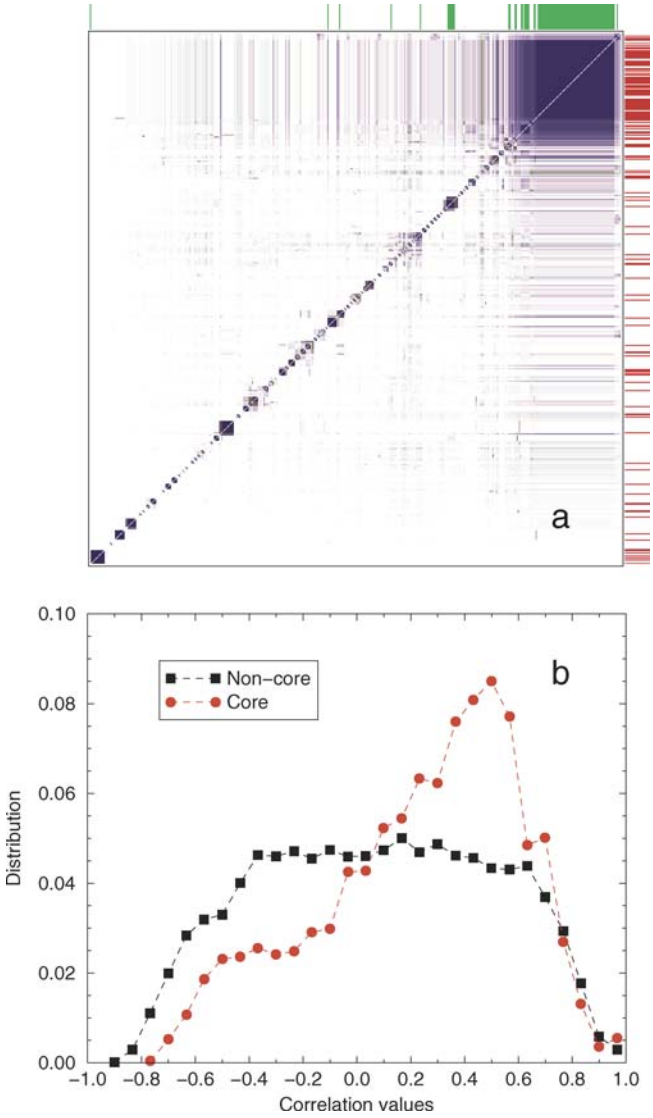

Figure 4. Correlations among *E. coli* Metabolic Reactions

(A) We calculated the Pearson correlation using flux values from 30,000 conditions for each reaction pair before grouping the reactions according to a hierarchical average-linkage clustering algorithm. The values of the flux-correlation matrix range from −1 (red) through 0 (white) to unity (blue). The horizontal color bar denotes if a reaction is a member of the core (green), and the vertical color bar denotes whether the enzymes catalyzing the reaction are essential (red).

(B) Distribution of Pearson correlation in mRNA copy numbers from 41 experiments [14]. The correlations of the core reactions are clearly shifted towards higher values, with an average correlation coefficient of $<C> = 0.23$ compared with the average noncore coefficient of $<C> = 0.07$

DOI: 10.1371/journal.pcbi.0010068.g004

ing the size of the metabolic core. Therefore, the metabolic core contains reactions that are necessary for optimal cellular performance regardless of the environmental conditions, while the conditionally active metabolic reactions represent the different ways in which the cell is capable of optimally utilizing substrates from its environment.

The identification of the metabolic core also has important practical implications: given the continuous activity and high degree of essentiality of the core enzymes, they represent potential targets for antimicrobial intervention. Specifically, while many bacterial and yeast gene products are essential, a high fraction of them are essential only in specific environments. For example, recent measurements indicate that 76% of the *S. cerevisiae* genes that are inactive in nutrient-rich conditions are in fact not only active, but are also essential in some other growth conditions [6]. However, an effective antimicrobial drug needs to be able to kill its target organism under all physiological conditions in which it can exist. Thus pharmacological interventions targeting a specific pathway will not be effective in environments where the pathway is not needed and its corresponding enzyme is turned off. Instead, the most effective antimicrobials must target the activity of the core reactions, as their disruption will have an impact on the microorganism's ability to function under all environmental conditions. Indeed, among the currently used antibiotics, fosfomycin and cycloserine act by inhibiting cell-wall peptidoglycan, while sulfonamides and trimethoprim inhibit tetrahydrofolate biosynthesis, both pathways being present in the *H. pylori* as well as the *E. coli* core (see Figure 2). In addition, core pathways involved in the synthesis of flavin adenine dinucleotide, 2-dehydro-3-deoxy-D-octonate, and lipopolysaccharide are among potential antibiotic targets [21].

This indicates that the core enzymes essential for biomass formation, both for optimal and suboptimal growth (Table S1), may prove effective antibiotic targets given the cell's need to maintain the activity of these enzymes in all conditions. The fact that not all such core reactions are shared by all bacteria offers the possibility to identify bacterium-specific drug targets. Finally, our results pertaining to the existence of the core and its characteristics are by no means limited to the three studied organisms, but the analysis can be carried over to all organisms for which high-quality metabolic reconstructions are available. Given the large number of sequenced bacterial genomes, such studies could open up new avenues for rapid in silico antimicrobial drug-target identification.

Materials and Methods

FBA. Starting from the published stoichiometric matrices of the reconstructed *E. coli* MG1655, *H. pylori*, and *S. cerevisiae* metabolic networks [7–9], the steady-state concentrations of all the internal metabolites of an organism satisfy

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} v_j = 0 \qquad (1)$$

where S_{ij} is the stoichiometric coefficient of metabolite A_i in reaction j, and v_j is the flux of reaction j. We use the convention that if metabolite A_i is a substrate (or product) in reaction j, $S_{ij} < 0$ ($S_{ij} > 0$). Any vector of fluxes $\{v_j\}$ which satisfies equation 1 corresponds to a state of the metabolic network, and hence, a potential state of operation of the cell. Using linear programming, we choose the solution that maximizes biomass production of the respective organisms. Reactions that are never active, likely reflecting annotation errors or incomplete data, are ignored in our flux analysis.

Metabolic core identification. We model all possible cellular environments in the investigated FBA models, as described before [3]. Briefly, for each metabolic-uptake reaction, we fix the uptake rate to a randomly chosen value of between 0 and 20 mmol/g/h (dry weight) before calculating the optimal fluxes for this configuration using linear programming. Each reaction is subsequently tagged as either active (non-zero flux) or inactive (zero flux). Since there are a very large number of possible combinations of the selected uptake rates, we repeat this process 30,000 times. In addition, we calculate the optimal fluxes for all single-carbon-source configurations on a minimal-uptake medium consisting of unlimited ammonia, sulfate, phosphate, carbon dioxide, potassium, and restricted oxygen for *E. coli*, *H. pylori*, and *S. cerevisiae*, plus unlimited ergosterol and zymosterol

for *S. cerevisiae*. For each reaction, we assign a value, q_i, of between zero and unity, representing the relative number of conditions for which reaction i is active. The preliminary metabolic core is defined as the set of reactions that are active in all sampled conditions ($q_i = 1$). We determine the final metabolic core by removing from the preliminary core those reactions that, when their flux rate is constrained to zero over the total set of initial conditions, leave the growth rate unchanged.

Correlations and clustering. We evaluate the FBA flux correlations by calculating the Pearson coefficient for all possible reaction pair combinations in the 30,000 different simulated growth conditions. We subsequently group the reactions by employing the correlation values as the metric in a standard hierarchical average-linkage clustering algorithm [22].

Data analysis. The annotated metabolic FBA models specify the enzymes and genes catalyzing the various metabolic reactions. We represent the mRNA activity level or half-life time of a metabolic enzyme by averaging the available experimental values, E_k, for all the catalyzing gene products, as

$$r_i = \frac{1}{N_i} \sum_{k=1,N_i} E_k \qquad (2)$$

where N_i is the number of genes. We assess the essentiality of a reaction by considering it essential if at least one of its catalyzing enzymes (or gene products) is essential for the survival of the organism (see Protocol S1 for error analysis).

Core and noncore essentiality. The deletion phenotype of *E. coli* enzymes was taken from Gerdes et al. [13]. For nonlethal core reactions, we cross-checked the deletion phenotype with the PEC database (http://www.shigen.nig.ac.jp/ecoli/pec) as given in Table S2 of Gerdes et al. [13], resulting in a total of 62 essential enzymes and seven enzymes for which we could not determine the deletion phenotype. The deletion phenotype of the yeast enzymes was taken from the CYGD database [12].

Supporting Information

Protocol S1. Supplementary Analyses and List of Metabolite Abbreviations

Found at DOI: 10.1371/journal.pcbi.0010068.sd001 (628 KB DOC).

Table S1. Metabolic Core Reactions

Found at DOI: 10.1371/journal.pcbi.0010068.st001 (57 KB XLS).

Acknowledgments

We thank S. Mobashery and S. Vakulenko for discussions. Work at the University of Notre Dame and at the University of Pittsburgh was supported by the US Department of Energy, the National Institutes of Health, and the National Science Foundation.

Competing interests. The authors have declared that no competing interests exist.

Author contributions. EA, ZNO, and ALB conceived the study. EA and ALB designed the study. EA performed all the computational work. EA, ZNO and ALB wrote the paper. ■

References

1. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97: 5528–5533.
2. Schilling CH, Letscher D, Palsson BO (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 203: 229–248.
3. Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási AL (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature 427: 839–843.
4. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic reconstructions. Genome Res 14: 301–312.
5. Reed JL, Palsson BO (2004) Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states. Genome Res 14: 1797–1805.
6. Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429: 661–664.
7. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, et al. (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. J Bacteriol 184: 4582–4593.
8. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4: R54.1–R54.12.
9. Duarte NC, Herrgard MJ, Palsson BO (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14: 1298–1309.
10. Burgard AP, Maranas CD (2001) Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. Biotechnol Bioeng 74: 364–375.
11. Burgard AP, Vaidyraman S, Maranas CD (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Prog 17: 791–797.
12. Güldener U, Münsterkötter M, Kastenmüller G, Strack N, van Helden J, et al. (2005) CYGD: The Comprehensive Yeast Genome Database. Nucleic Acids Res 33: D364–D368.
13. Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, et al. (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol 185: 5673–5684.
14. Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res 13: 216–223.
15. Shen-Orr S, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 31: 64–68.
16. Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, et al. (2004) RegulonDB (version 4.0): Transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. Nucleic Acids Res 32: D303–D306.
17. Wiback SJ, Famili I, Greenberg HJ, Palsson BO (2004) Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. J Theor Biol 228: 437–447.
18. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.
19. Allen TE, Herrgard MJ, Liu MZ, Qiu Y, Glasner JD, et al. (2003) Genome-scale analysis of the uses of the *Escherichia coli* genome: Model-driven analysis of heterogeneous data sets. J Bacteriol 185: 6392–6399.
20. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: Evaluating the consequences of constraints. Nat Rev Microbiol 2: 886–897.
21. Gerdes SY, Scholle MD, D'Souza M, Bernal A, Baev MV, et al. (2002) From genetic footprinting to antimicrobial drug targets: Examples in cofactor biosynthetic pathways. J Bacteriol 184: 4555–4572.
22. Dunn G, Everitt BS (1982) An introduction to mathematical taxonomy. Cambridge: Cambridge University Press. 152 p.