

SYSTEMS BIOLOGY

The hidden simplicity of metabolic networks is revealed by multireaction dependencies

Anika Küken¹, Damoun Langary^{1,2}, Zoran Nikoloski^{1,2*}

Understanding the complexity of metabolic networks has implications for manipulation of their functions. The complexity of metabolic networks can be characterized by identifying multireaction dependencies that are challenging to determine due to the sheer number of combinations to consider. Here, we propose the concept of concordant complexes that captures multireaction dependencies and can be efficiently determined from the algebraic structure and operational constraints of metabolic networks. The concordant complexes imply the existence of concordance modules based on which the apparent complexity of 12 metabolic networks of organisms from all kingdoms of life can be reduced by at least 78%. A comparative analysis against an ensemble of randomized metabolic networks shows that the metabolic network of *Escherichia coli* contains fewer concordance modules and is, therefore, more tightly coordinated than expected by chance. Together, our findings demonstrate that metabolic networks are considerably simpler than what can be perceived from their structure alone.

INTRODUCTION

Advances in technologies and computational approaches for genome assemblies coupled with systems biology efforts to annotate gene functions and generate omics data have propelled the development of large-scale models of metabolism not only of single cell types, organs, and organisms but also of metabolic communities of interacting organisms (1–4). In parallel with the detailed modeling of metabolic networks for specific biological systems, principles underlying the organization of metabolic reactions in metabolic networks that support steady states have also been determined. For instance, metabolic networks have been shown to exhibit bow tie structure, where few metabolites act as intermediates between a large number of precursors (i.e., nutrients) transformed in multiple building blocks of biomass (5); the bow tie structure is, in turn, reflected in the power law distribution of the number of reactions in which metabolites participate (6), in the minimal path between precursors and biomass components (7), and in the hierarchical ordering of steady-state reaction fluxes (8, 9). In addition, pairs of metabolic reactions have been grouped into different classes based on the relationships that their fluxes exhibit in every steady state that the network supports (10, 11), providing means to study the modular organization of metabolic networks (12, 13) under operational constraints.

While these findings are exclusively based on pairwise relationships between reaction fluxes, there may exist functional dependencies that include more than two reactions. However, identifying principles of functional organization of large-scale metabolic networks that go beyond pairwise relationships of reaction fluxes is challenging due to the sheer number of combinations of reactions and relations to consider. Against this background we ask: Can dependencies between multiple steady-state metabolic fluxes be identified efficiently by using structural network properties? And what are the implications of the multireaction dependencies on the simplification of the underlying networks?

To address these questions, we first define the concept of concordance of complexes that can be used to unravel dependencies between multiple steady-state reaction fluxes. We then devise a procedure for an efficient identification of concordant complexes in large-scale networks with arbitrary enzyme kinetics. We show that concordant complexes are present in genome-scale metabolic models across species from all kingdoms of life. In addition, by using the notion of concordant complexes, we provide a sound formalization of modules in a metabolic network that considers operating constraints, in contrast to previous works that rely on either the network structure or pairwise flux relations only. We then discuss additional, mild structural conditions ensuring that the identified modules can be studied in isolation—providing a tangible biochemical implications of the introduced concepts. We also show that the modules specify the extent to which a metabolic network can be simplified. Last, we also demonstrate that properties of the modules in the metabolic network of *Escherichia coli* are different in comparison to an ensemble of randomized network variants, indicating biological relevance of the introduced concepts. Together, our findings demonstrate that multireaction dependencies between fluxes highlight the elegant simplicity underlying seemingly complex metabolic networks across organisms from all kingdoms of life.

RESULTS

Concordance of complexes in biochemical networks

To define the relation of concordance of complexes, we first introduce key concepts from stoichiometric models of metabolic networks. A biochemical network is composed of reactions through which biochemical components, referred to as species, are transformed from substrates into products. The toy example network in Fig. 1A includes 10 reactions that transform six species, denoted by letters A to F (14). The network structure is described by nodes that denote complexes, corresponding to the left- and right-hand sides of the considered reactions, and directed edges representing the reactions. For instance, the network in Fig. 1A contains eight complexes connected by 10 reactions. Every reaction has a head, or substrate, and a tail, or product complex. For instance, 2A is the head and B is the tail complex of the first reaction in Fig. 1A. The

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany. ²Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany.

*Corresponding author. Email: nikoloski@mpimp-golm.mpg.de

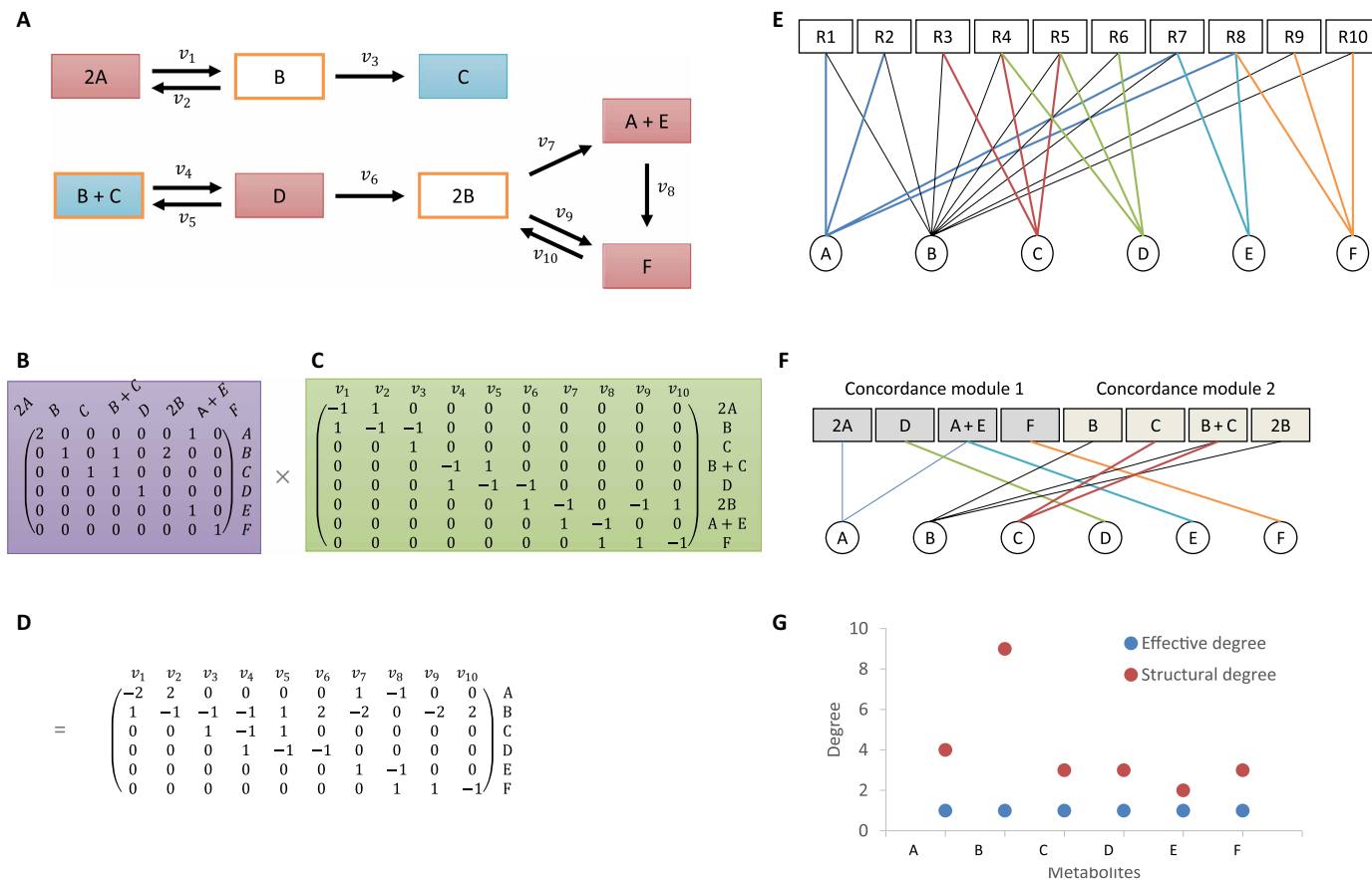


Fig. 1. Illustration of concordant complexes and implications of metabolite degree. (A) To illustrate the three different types of concordant complexes, we use a network (14) with 6 species A to F, 10 reactions with fluxes $v_1 - v_{10}$, and 8 complexes, marked with rectangular boxes. Complexes marked in red are balanced and form a concordant module. Complexes C and B + C, marked in blue, are trivially concordant because species C only participates in these two complexes. A group of nontrivial mutually concordant complexes, composed of complexes B, B + C, and 2B, is marked with orange boxes. Hence, the complexes B, C, B + C, and 2B form another concordance module. (B) Species-complex matrix Y , with rows corresponding to species and column to complexes. Each entry indicates the molarity with which a species participates in a complex. (C) Incidence matrix A of the directed graph given in (A). (D) Stoichiometry matrix N , with rows corresponding to species and columns denoting reactions, and each entry indicates the molarity with which a species is produced (positive) or consumed (negative) by a reaction. The stoichiometry matrix of the network is given by the product of species-complex matrix and incidence matrix, $N = YA$. Illustration of different networks used in the definition of metabolite degree: (E) Bipartite reaction-metabolite graph, in which metabolite degree corresponds to the number of reactions in which a metabolite participates, given by the number of edges incident on the metabolite node. (F) Bipartite complex-metabolite graph, in which metabolite degree corresponds to the number of complexes in which a metabolite participates. In the complex metabolite bipartite representation, a metabolite can participate in one or multiple concordance modules. The effective degree of a metabolite is given by the number of concordance modules that contain complexes that include the metabolite. (G) Comparison of classical definition of metabolite degree and the here-defined effective metabolite degree.

stoichiometric matrix, N , of the network is then given by the product of the matrix, Y , describing the species composition of complexes and the incidence matrix, A , of the corresponding directed graph (Fig. 1, B to D) (15, 16). Thus, column Y_j gives the species and molarity with which they form the complex j . In addition, the reactions are weighted by nonnegative numbers, which correspond to fluxes of a steady-state flux distribution, v , which satisfies $\frac{dx}{dt} = Nv = YAv = 0$, whereby the concentrations of species, gathered in the vector x , are invariant in time.

We next consider the activity of a complex, j , denoted by α_j , given by the differences between the sum of fluxes of reactions that have j as a product complex and the sum of fluxes of reactions that have j as a substrate complex. This activity of a complex is also referred to as the complex formation rate (17). Given a flux distribution, v , the

activity of complex j for that flux distribution can be succinctly written as $\alpha_j = A_j v$, where A_j corresponds to the j th row of the incidence matrix A of the network. For instance, the activity of complex C is given by $\alpha_C = v_3$, while that of complex B + C is given by $\alpha_{B+C} = v_5 - v_4$. Therefore, the activities of complexes represent multireaction relationships that correspond to linear combinations of reaction fluxes and reflect the network structure. As a result, the steady-state equations can be expressed in terms of activities of complexes, i.e., $\frac{dx}{dt} = YAv = Y\alpha = 0$.

Given a set of steady-state flux distributions $S = \{v, Nv = 0, v_{\min} \leq v \leq v_{\max}\}$, we say that complexes i and j are concordant in S if the concordance ratio, $\frac{\alpha_i}{\alpha_j}$ of their activities, α_i and α_j , is finite, nonzero, and invariant over the flux distributions in S . Formally, two complexes are concordant if there exists a nonzero constant γ_{ij} , such that

$\alpha_i - \gamma_{ij}\alpha_j = 0$ holds for every flux distribution in S . For instance, from the steady-state equation for species C, we get that $\alpha_C + \alpha_{B+C} = 0$, whereby $\frac{\alpha_C}{\alpha_{B+C}} = -1$.

Our first contribution consists of showing that concordant complexes can be efficiently identified in large-scale biochemical networks by linear fractional programming (see Methods). We note that complexes whose activity is zero in every steady state in S are referred to as balanced and have been used in reduction of metabolic networks (18). For instance, complex A + E is balanced because species E occurs in only this complex; as a result, the complex 2A is also balanced. On the basis of the definition of concordance, all balanced complexes can be considered mutually concordant.

Pairs of concordant complexes can be categorized into three groups that include (i) balanced complexes, (ii) trivially concordant complexes, and (iii) other concordant complexes that are not balanced or trivial. This categorization of concordant complexes can be readily related to properties of the species in the network. If a species k participates in only two complexes, i and j , then it must be that $y_{ki}\alpha_i + y_{kj}\alpha_j = 0$, and thereby $\frac{\alpha_i}{\alpha_j} = -\frac{y_{kj}}{y_{ki}}$. We refer to such complexes as trivially concordant. For instance, in the network in Fig. 1A, complexes C and B + C are trivially concordant because C occurs in only these two complexes in the network. In contrast, species B participates in three complexes, B, B + C, and 2B; because $\alpha_B = v_1 - v_2 - v_3 = -\alpha_{2A} - v_3 = -\alpha_C$, as the complex 2A is balanced, we obtain that these three complexes are mutually concordant. This example provides a key observation for mutually concordant complexes, whereby the activity of one of these complexes suffices to deduce the activities of the others. One can easily see that the extent to which species jointly participate in complexes is one factor that allows for concordance to occur, leading to the hypothesis that concordance of complexes may be prevalent in metabolic networks, given the interplay between species in these networks.

Concordant complexes in genome-scale metabolic networks

Having defined and illustrated the concept of concordance between complexes, we next tested the hypothesis about prevalence of concordant complexes in high-quality metabolic network models. To this end, we used the metabolic networks of 12 organisms to compare and contrast the percentage of concordant complex pairs in the following two scenarios: (i) The reactions follow the irreversibility constraints imposed in the original models, and (ii) in addition to irreversibility constraint, optimality of growth rate is imposed. We considered these two scenarios because we aimed to characterize the effects of inspecting a subset of feasible flux distributions (e.g., due to imposing additional constraints) on the determined concordant complexes. This will also indicate the extent to which optimality assumptions, often invoked in constraint-based modeling of metabolic networks, contribute to two pairs deemed as concordant. We note that reversible reactions are split into two irreversible reactions before applying the approach, and no assumptions are made about the reaction kinetics.

We found that in scenario (i), when only reaction reversibility constraints are considered, the models that exhibited the largest percentage of concordant complex pairs included *Methanosarcina barkeri* (37.8%), followed by *Methanosarcina acetivorans* (24.3%) and *Mycobacterium tuberculosis* (10.9%) (table S1 and Fig. 2A); in contrast, the smallest percentage of concordant complex pairs were identified in the models of *Natronomonas pharaonis* (1.8%) and *Arabidopsis thaliana* (0.6%) (Fig. 2A and table S1). We note that, by

the imposed convention, the considered concordant complex pairs include all pairs of balanced complexes. Upon excluding the pairs of balanced complexes from the pairs of concordant complexes, we found that the models of *N. pharaonis* (1.15%) and *Pseudomonas putida* (0.66%) exhibited the largest percentage of concordant complex pairs (table S1 and Fig. 2A). In scenario (ii), when optimal specific growth rate is imposed as an additional constraint, we observed, as expected, an increase in the percentage of concordant complex pairs. For the models of *A. thaliana*, *N. pharaonis*, *P. putida*, and *Thermotoga maritima*, the additional constraint resulted in at least 14% increase in the percentage of concordant complex pairs, ranging from 1.1% in *A. thaliana* to 3.7% for *N. pharaonis* and 4.5% for *P. putida* (Fig. 2A and table S1). In other words, imposing restrictive operational constraint—under which a metabolic network of interest may function—leads to higher concordance of activities of the network complexes.

Investigating the percentage of complexes that are in concordance relation with at least one other complex, we found that the largest fraction in scenario (i) corresponds to the models of *M. barkeri* (83%) and *N. pharaonis* (70%) (Fig. 2B and table S1). In contrast, the models of *A. thaliana* (42%), *Saccharomyces cerevisiae* (40%), *Chlamydomonas reinhardtii* (39%), and *Escherichia coli* (36%) showed the smallest percentage of complexes that are in concordance relation with at least one other complex. These findings hold in the scenario when optimality of the specific growth rate is imposed as a constraint (table S1).

As illustrated above, the concordance ratio for trivially concordant complexes is necessarily negative. Therefore, we next determined the percentage of complex pairs that are trivially concordant and also looked for those complex pairs whose concordance ratios is positive. Our findings indicated that the percentage of trivially concordant complex pairs ranges from 0.005% in *S. cerevisiae* to 0.06% in *T. maritima*, with an average of 0.02% across the investigated models (table S1 and Fig. 2A). Furthermore, the percentage of concordant complex pairs with positive concordance ratio ranges from 0.0001% in *E. coli* to 0.2% in *N. pharaonis* (table S1). Together, our results demonstrated that genome-scale metabolic networks harbor nontrivial concordant complexes that arise as a result of the interplay between network structure and operational constraints.

Concordance modules and distribution of their sizes across metabolic networks

Because concordance is an equivalence relation (Methods), it partitions the set of complexes into classes of mutually concordant complexes that give rise to concordance modules in a metabolic network. From the definition of concordance module, it follows that knowledge of the activity of one complex suffices to obtain the activities of all other complexes in the module. For instance, the paradigmatic network in Fig. 1A is composed of two concordance modules (Fig. 1F): The first is given by all balanced complexes, namely, 2A, D, A + E, and F, while the second is composed of B, C, B + C, and 2B. Therefore, the notion of concordance modules can be used to quantify the modularity of functional metabolic networks.

We were next interested to examine whether concordance modules are related to metabolic pathways, as defined in biochemistry textbooks (19). To this end, we used the information of metabolic subsystems in the analyzed models and investigated the number of modules that include more than half of complexes of each metabolic subsystem. For instance, in the model of *E. coli*, more than half of complexes in glycolysis/gluconeogenesis as well as valine, leucine,

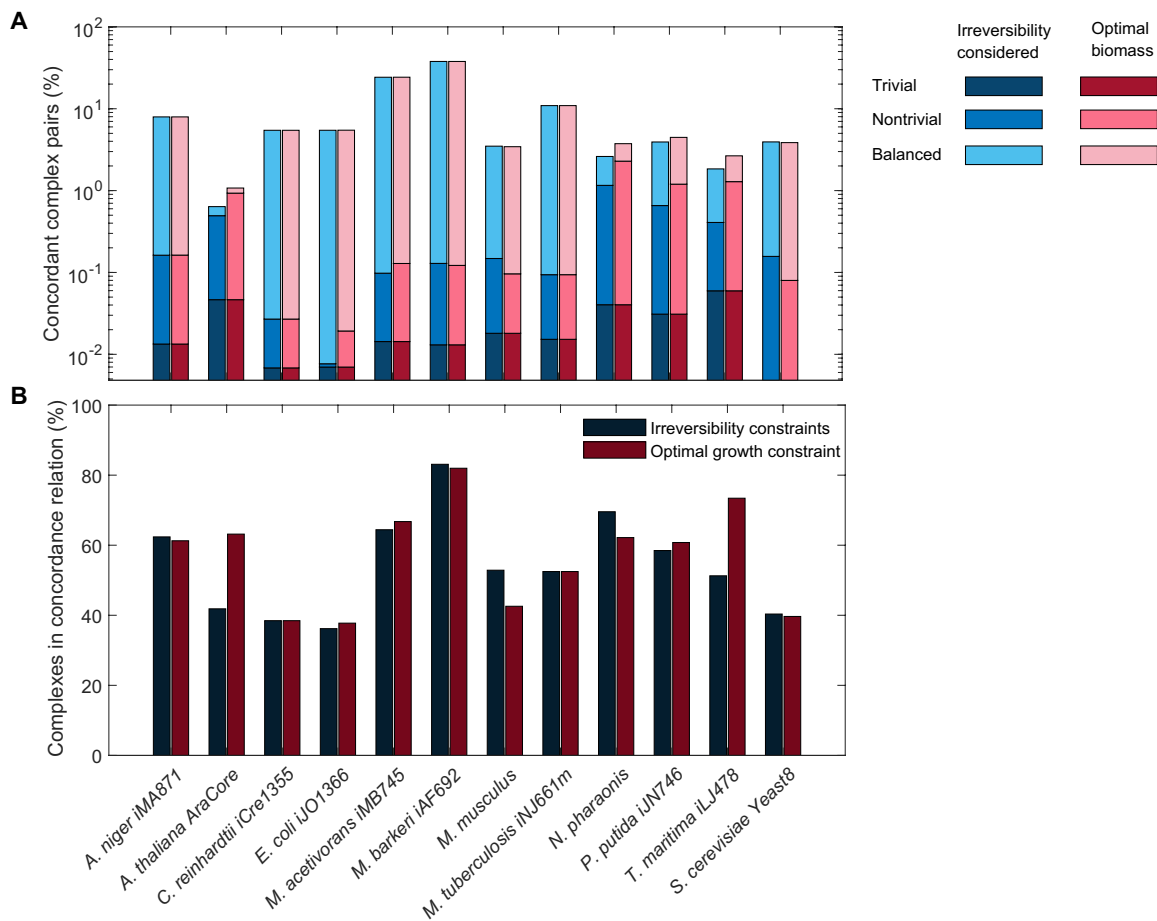


Fig. 2. Prevalence of concordant complexes in metabolic networks from 12 organisms. Genome-scale metabolic models of 12 organisms from all kingdoms of life are analyzed for occurrence of concordant complexes (i) assuming reaction reversibility as given in the original model and (ii) when optimal specific growth rate is imposed as additional constraint. **(A)** Percentage of concordant complex pairs from the total number of unique complex pairs. Concordant complex pairs can be categorized into three groups (legend): concordance of balanced complexes (balanced), trivially concordant complexes including a species that occurs in these two complexes only, and other concordant complexes (that are not balanced or trivial). **(B)** Percentage of complexes in concordance relation to at least one other complex relative to the total number of model complexes.

and isoleucine metabolism are covered by eight modules (Fig. 3). In *S. cerevisiae*, more than half of complexes in the metabolic subsystems of fatty acid biosynthesis, atrazine degradation, carbapenem biosynthesis, and sulfur relay system are grouped in individual modules (fig. S1). In *A. thaliana*, this is the case for 10 metabolic subsystems, including light reactions, carbon fixation, and fatty acid synthesis (fig. S2). Furthermore, for aspartate synthesis, oxidative phosphorylation, proline synthesis, pyruvate decarboxylation, serine synthesis, and sulfur assimilation, only two modules comprise more than half of the complexes. To statistically assess whether complexes of the same subsystem are overrepresented in a concordance module, we used the hypergeometric test (Methods). Considering only concordance modules including at least one complex of the tested subsystems, we observed significant enrichment ($P \leq 0.05$) for 6, 4, and 2% of the subsystems in *E. coli*, *S. cerevisiae*, and *A. thaliana*, respectively (table S2). Furthermore, for 42, 33, and 8% of the subsystems, we found significant enrichment in at least 50% of the concordance modules that contain complexes of that subsystem. Therefore, these findings indicated that concordance modules, inferred in an automated

fashion from the network structure in combination with operational constraints, are partly in line with textbook boundaries of metabolic pathways and emphasize the relation between these pathways.

Next, we investigated the distribution of the size of concordance modules across all used models in the two considered scenarios. For scenario (i), without considering balanced complexes, we found that these distributions do not follow power law, except for the networks of *M. tuberculosis*, *P. putida*, and *S. cerevisiae* (scaling coefficients in [5.96, 8.92]; table S3D and fig. S3). These distributions can be hence classified in the Super-Weak category, indicating that although power law seems to be a better fit compared to some tested alternatives, it is not necessarily a statistically plausible choice in these cases (20). With consideration of balanced complexes, the size of concordance modules does not follow power law for any models (table S2C). The distribution of the size of concordance modules is better described by stretched exponential distributions of the form $y = e^{-\left(\frac{x}{b}\right)^a}$, with parameters $a = 1$ and $b = 0.01$ (table S3, C and D) (20). Under scenario (ii), without considering balanced complexes, the size of concordance modules follows Super-Weak power law in the networks

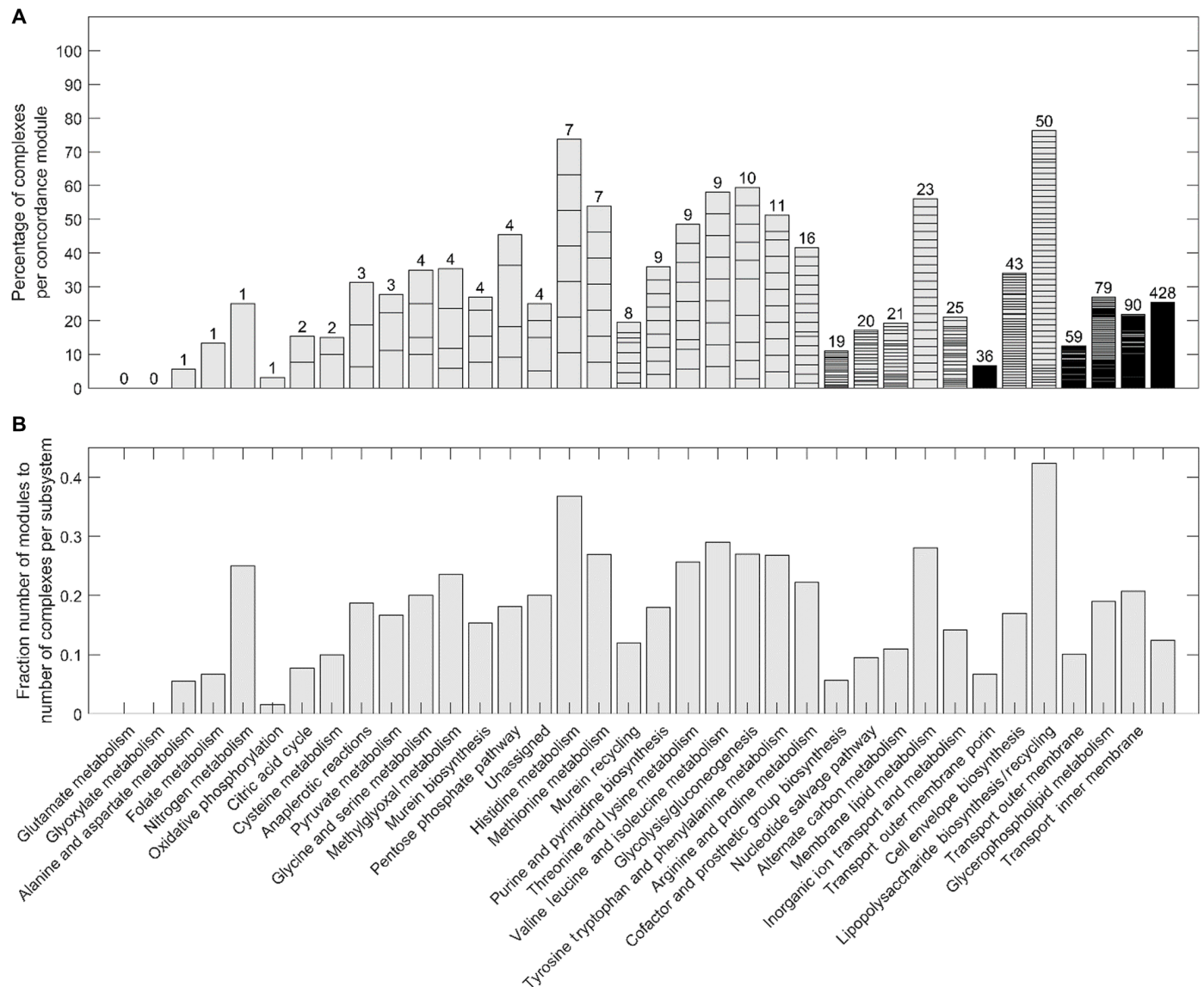


Fig. 3. Concordant module structure of metabolic subsystems in the model of *E. coli*. (A) Percentage of complexes assigned to a subsystem grouped in the same concordance module. For subsystems that do not sum up to 100%, the remaining complexes are not in concordance relation with any other complex. A horizontal line is added in each bar to mark the contribution of a module to a metabolic subsystem. The number on top of each bar indicates the number of concordance modules per subsystem. (B) Fraction of the number of concordance modules per subsystem to the number of complexes per subsystem.

of *A. thaliana*, *E. coli*, *Mus musculus*, *M. tuberculosis*, *P. putida*, and *T. maritima* (table S3D). Similar to scenario (i), when we take balanced complexes into account, no distribution follows power law (table S3C). In this case, too, the size of concordance modules is better described by stretched exponential distributions with estimated of $a = 1$ and $b = 0.01$ for the parameters (table S3, C and D) (20). These findings are in line with the presence of few large and many small concordance modules in the analyzed large-scale metabolic networks.

Implications of concordance modules on the metabolite degree

Because activities of complexes are the building blocks for the steady-state equations, we also examined the effect of modularity on the

structure of these equations and, thereby, on the complexity of the metabolic networks at steady state. The number of reactions in which a metabolite participates corresponds to the number of reaction fluxes appearing in the steady-state equation for that metabolite and is termed nominal metabolite degree. For instance, for the network in Fig. 1E, species A participates in four reactions, while species B participates in nine reactions, corresponding to their nominal degrees. Previous work has indicated that the nominal metabolite degree is associated with chemical properties of metabolites (e.g., molecular solubility) and their concentrations (21) and that metabolites of larger nominal degree may exhibit smaller variability of concentrations (22). It has been demonstrated that the number of reactions in which metabolites participate follows power law distribution and that this evidence is the strongest in comparison to

degree distributions in different types of biological networks analyzed to date (6, 20). However, because activities of complexes are used to specify the steady-state equation of a metabolite, we hereby define the degree of a metabolite as the number of complexes in which it participates. For instance, in the network in Fig. 1A, species A is of degree two as it participates in complexes 2A and A + E, while species B is of degree three due to its appearance in complexes B, B + C, and 2B (Fig. 1F). We found that for 67% of networks, the metabolite degree distribution follows power law (20) with exponents ranging from 1.96 in the *M. acetivorans* to 2.65 in the *A. thaliana* network (Fig. 4 and table S3A). Furthermore, the degree distributions of *Aspergillus niger* and *T. maritima* can be classified in the Strongest scale-free category [see (20) for classification details], as Strong in case of *A. thaliana* and Weak for *E. coli*, *M. acetivorans*, *M. barkeri*, and *M. musculus*. In contrast, the degree distributions of *C. reinhardtii*, *M. tuberculosis*, *N. pharaonis*, *P. putida*, and *S. cerevisiae* are not power law and can be better described by stretched exponential functions of the form $y =$

$e^{-\left(\frac{x}{0.01}\right)^1}$ (Fig. 4). In line with the analysis based on the number of reactions, ubiquitous metabolites, such as H^+ and water, followed by adenosine triphosphate, phosphate, adenosine diphosphate, and nicotinamide adenine dinucleotides, participate in the largest number of complexes and thus exhibit the largest degree across the analyzed metabolic networks (table S4).

To better analyze the implications of the concordance modules on the structure of the underlying equations and the degree of metabolites (see fig. S4 for distributions of metabolite degrees), let us first define the following notions: We say that a metabolite appears in a concordance module if the module includes at least one complex containing the metabolite. The effective degree of a metabolite is then given by the number of concordance modules in which it appears. For instance, while species B is of degree three, as it participates in three complexes in the network in Fig. 1A, its effective degree is one, because all these complexes are contained in one concordance module (Fig. 1G). Therefore, on the basis of the algebraic structure alone, metabolites of large degree may, in fact, have small

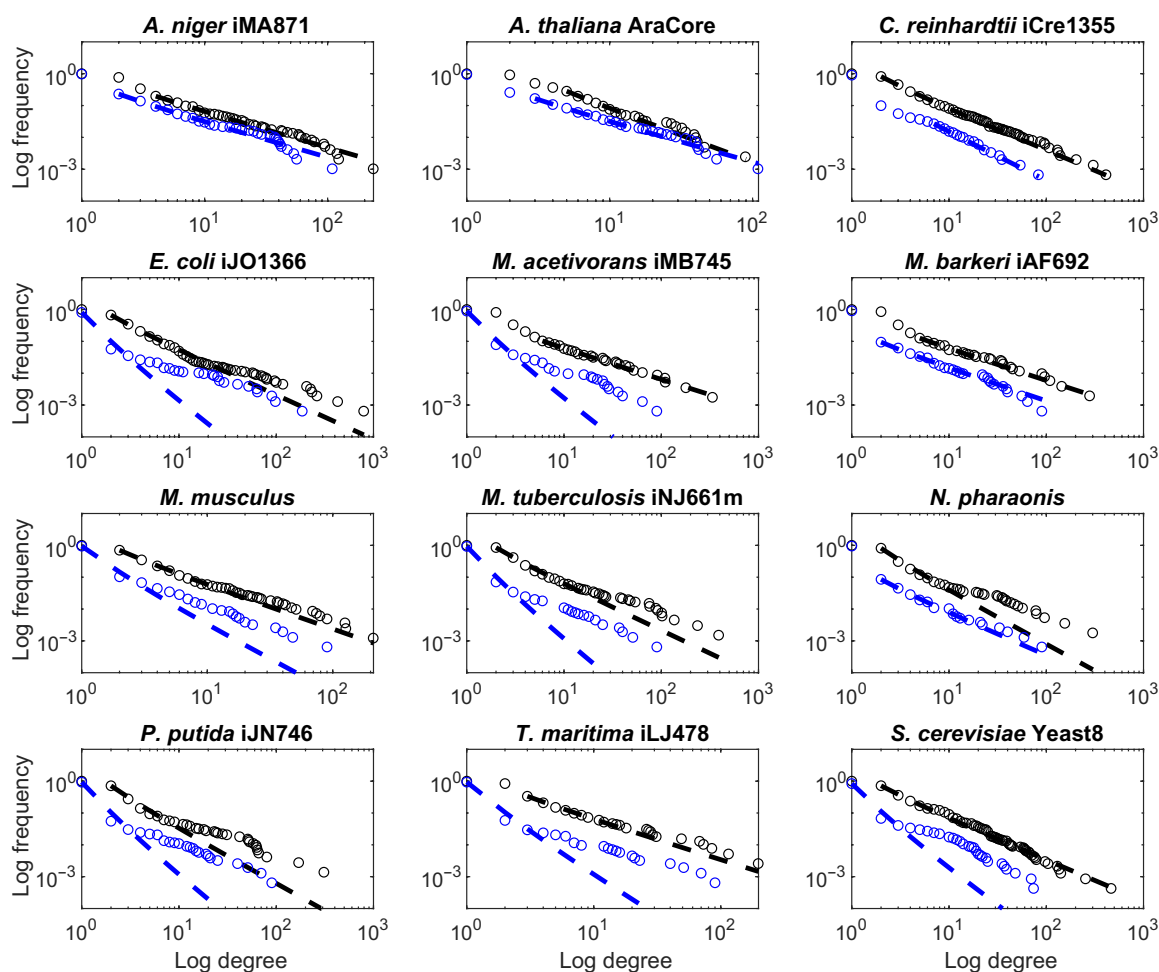


Fig. 4. Log-log plot of distributions of metabolite degree and effective degree in networks of 12 organisms. We fit power law distributions to metabolite degree, i.e., the number of complexes in which a metabolite participates, and effective degree, i.e., the number of concordance modules in which a metabolite participates, obtained for scenario (i). The distributions of effective degrees are drawn in blue, while those of the metabolite degrees are drawn in black. While 58% of distributions of degree and 58% of distributions of effective degree follow power law, the classification shows power law to be less plausible for the effective degree (table S3, A and B).

effective degree, when steady state and other operational constraints are imposed—which is a key implication of concordance modules. Our analysis demonstrated that the effective degree follows Super-Weak power law only in the metabolic networks of *A. thaliana*, *M. acetivorans*, *M. barkeri*, *M. tuberculosis*, and *N. pharaonis* for both scenarios (table S3B). Furthermore, the effective degree follows power law for the network of *M. musculus* for scenario (i) and *T. maritima* for scenario (ii) (Fig. 4 and table S3B). Therefore, unlike the case with nominal degree distributions, power law is hardly a plausible fit when it comes to the distribution of effective degrees—that capture the operational constraints of the network. For at least 42% of the analyzed models in both scenarios (i) and (ii), the effective degree is better described by stretched exponential distributions of the form $y = e^{-\left(\frac{x}{b}\right)^a}$, with estimated parameters a ranging from 0.8 to 1 and b ranging from 0.01 to 0.02 (table S3B) (20). These findings motivate the following analysis of reducibility of metabolic networks.

Reducibility index highlights the hidden simplicity of metabolic networks

We note that the effective degree implies simplification of the steady-state equations, because it denotes the smallest number of complexes that suffice to obtain mathematically equivalent descriptions of the steady-state equations. For instance, the steady-state equation of species B in terms of activities of complexes is given by $\alpha_B + \alpha_{B+C} + 2\alpha_{2B} = 0$, which simplifies to $\alpha_B + \alpha_{2B} = 0$, because $\frac{\alpha_B}{\alpha_{B+C}} = 1$. As a result, we define the reducibility index of a metabolic network as $1 - \frac{m}{c}$, with m denoting the number of concordance modules and c representing the total number of complexes. A larger value for the index indicates a higher reducibility of the metabolic network. For instance, the reducibility index for the network in Fig. 1 is 0.75, because there are two concordance modules and eight complexes. Our results showed that the reducibility index across the 12 metabolic networks analyzed ranges from 0.82 to 0.94 in scenario (i) and from 0.85 to 0.96 in scenario (ii) (fig. S5). In addition, we investigated the reducibility index when balanced complexes were not considered and observed ranges from 0.78 to 0.94 in scenario (i) and from 0.78 to

0.96 in scenario (ii) (Fig. 5), indicating that metabolic networks can be effectively reduced at steady state. Together, the concordance modules that arise because of the interplay of the network structure, steady-state constraints, and flux capacity constraints reveal the simplicity of seemingly complex metabolic networks.

Concordance modules and reducibility of randomized metabolic networks

To obtain insights into how different the properties of concordance modules are in networks that obey physicochemical constraints (e.g., mass balancing, number of substrates, and products of reactions) on the same set of metabolites, we created an ensemble of randomized network variants from the metabolic network model of *E. coli* (23). We then determined the concordance modules in each of the network variants and investigated the following statistics: the number of concordance modules, the size of the largest concordance module, the mean concordance module size, and the reducibility index, all associated with the degree of coordination between activity of complexes; furthermore, we determined the mean and maximum effective metabolite degree, hinting at the simplicity of the existing metabolic network of *E. coli*. Assuming that the null distributions generated for each property based on the ensemble of metabolic networks are normal (fig. S6), we then used the z score to calculate the significance of the observed values. By solving ~275 million large-scale linear programs (see Methods), we found that the number of concordance modules ($P = 2 \times 10^{-9}$) as well as their mean ($P = 1 \times 10^{-16}$) and maximum size ($P = 0.01$) was statistically smaller than expected at random. These findings suggest that the real-world metabolic network of *E. coli* has experienced evolutionary pressure toward higher coordination of the activity of complexes. In addition, we also observed a statistically smaller mean ($P = 4 \times 10^{-22}$) and maximum ($P = 0.002$) effective metabolite degree than expected by chance, indicating that the higher coordination in metabolism of *E. coli* is associated with larger simplicity of its metabolic network. This is in line with the larger reducibility index of the observed metabolic network in comparison to the randomized network variants ($P = 2 \times 10^{-31}$).

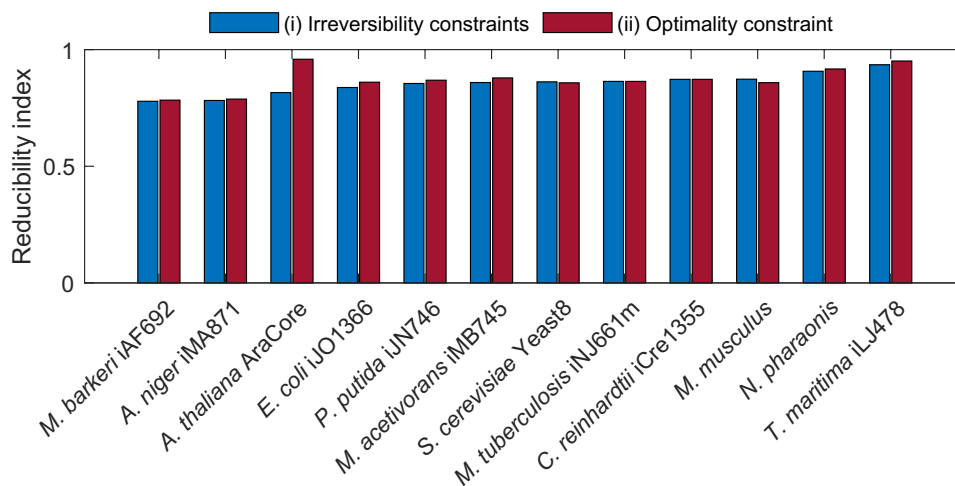


Fig. 5. Reducibility index in metabolic networks from species across kingdoms of life. Genome-scale metabolic models of 12 organisms from all kingdoms of life are analyzed for their reducibility index when balanced complexes were not considered. We considered two scenarios: (i) assuming reaction reversibility as given in the original model and (ii) when optimal growth is imposed as an additional constraint.

Classification of concordance modules and their implications

Next, we show that the concept of concordance module has important implications with respect to decomposability of metabolic networks. First, we note that a balanced complex that has out-degree of one can be removed without affecting the steady-state supported by the wired network (18), because this amounts to substitution of the flux of the outgoing reaction by the sum of fluxes of the incoming reactions. On the basis of the number of reactions incoming and outgoing to a concordance module in the network obtained by removal of such balanced complexes, we define four classes of concordance modules: (i) source modules that have no input from any complex outside of the concordant module but have output to other concordant modules, (ii) sink modules that have no output to any complex outside of the concordant module but have some inputs from other concordant modules, (iii) intermediate modules that have input and output from complexes outside of the concordant module, and (iv) closed modules that have no input or output from any complex outside of the concordant module. For instance, the network obtained upon removal of the balanced complexes in Fig. 1A is composed of balanced complex D, which cannot be removed from the network without further assumptions on reaction kinetics, and concordance module composed of B, C, B + C, and 2B (see fig. S7). Because there are two reactions incoming to the module from complex D and one reaction outgoing from the module to complex D, this represents an intermediate module. Moreover, a module that only includes species that do not appear in any other module will be called independent, because it paves the way to analyze the corresponding steady-state equations in isolation from the rest of the network. For instance, the intermediate module in fig. S7, composed of B, C, B + C, and 2B complexes, is an independent module. In addition, we call a module pseudo-independent, if it only includes species that do not appear in any other module or species that can be assumed to be buffered (i.e., their concentration can be considered constant over different environmental conditions) under the analyzed scenarios.

Equipped with these definitions and motivated by the suggested bow tie structure of metabolism (6), we next ask whether the concordance modules have similar interconnections across the analyzed metabolic networks in the simpler scenario (i). Unexpectedly, we found that 7 of 12 networks do not contain source and sink modules, and the remaining networks include less than 5% of source or sink modules (table S5). Intermediate modules can be found to a large extent across all networks, ranging from 37.3% in the network of *A. niger* to 100% in networks of *T. maritima* and *S. cerevisiae*. In addition, we identified that between 0.3 and 58.1% of modules are closed, with the highest percentage found in networks of *A. niger* (58.1%), *M. barkeri* (56.0%), and *N. pharaonis* (36.8%) and the lowest percentage found in networks of *C. reinhardtii* (0.3%), *M. tuberculosis* (1.5%), *M. acetivorans* (2.3%), and *P. putida* (4.2%). Closed modules are small, with a median number of two complexes across all networks and maximum number of complexes ranging from two complexes in the networks of *C. reinhardtii* and *M. tuberculosis* to 48 complexes in the network of *A. niger* (table S5). These findings indicate that metabolic networks differ in the interconnectedness of their concordance modules, which either are tightly linked or fully detached from the rest of the modules. However, additional assumptions on the reaction rate laws are needed to make further claims regarding the concentration of metabolites.

Because we did not find any independent modules among the identified modules, we next investigated the existence of

pseudo-independent modules, assuming that currency metabolites are of buffered concentration. We identified pseudo-independent modules in 5 of 12 analyzed networks, i.e., 11 in the network of *A. niger*, 8 in the network of *E. coli*, 5 such modules in the network of *M. barkeri*, 4 in the network of *N. pharaonis*, and only 1 pseudo-independent module in the network of *M. musculus* (table S5). These pseudo-independent modules are of small size and include up to four complexes. We also found that these pseudo-independent modules include a few (three to eight) metabolites (tables S5 and S6 and fig. S8) from pathways like alternate carbon metabolism, citric acid cycle, glycolysis/gluconeogenesis, cell envelope biosynthesis, and glycerophospholipid metabolism in *E. coli* or nucleotide metabolism, cysteine metabolism, and alanine and aspartate metabolism in *M. barkeri*. With the assumption that the kinetic rates in these pseudo-independent modules depend only on the metabolites in the respective modules, our findings show that these modules can be analyzed in isolation from the rest of the network.

DISCUSSION

Network science has sparked interest in characterizing the complexity and self-organizing capacity of networks across different domains, largely by contrasting seminal properties, like the degree distribution and average path length, of real-world networks with classical models of random graphs (24). However, these analyses with metabolic networks neglect physicochemical and functional constraints that such networks must obey. Here, we focused on properties of steady-state flux distributions along with the implications that they have on multireaction dependencies that arise because of the interplay between the network structure, physio-chemical, and functional constraints. By using the representation of the steady-state equations in terms of activities of complexes, we defined the notion of concordance of complexes. The concordance of complexes appears to have a loose connection to full coupling of fluxes (11), whereby two fluxes have an invariant nonzero ratio for any steady-state flux distribution. We showed that all concordant complexes can be efficiently identified in large-scale metabolic networks. The concordance relation allows us to identify concordance modules in metabolic networks. The presence of concordance modules indicates the possibility for network simplifications, because fewer activities of complexes suffice to obtain equivalent steady-state equations.

On the basis of the network concepts of effective degree of a metabolite and reducibility index and by using metabolic models across species from all kingdoms of life, we showed that (i) the effective degree does not follow power law degree distribution for the majority of studied networks; (ii) the metabolic networks can be effectively reduced due to the presence of concordance complexes; (iii) the network properties based on concordance modules significantly differ between the metabolic network of *E. coli* and an ensemble of randomized network variants, hinting at larger degree of coordination of activity of complexes than expected at random; and (iv) characterization of concordance modules based on their connectedness allows us to identify (pseudo)independent modules that pave the way for their analysis without considering the network context. All these results hold irrespective of the enzyme kinetics that the reactions may assume. Because the findings are obtained on the basis of simple techniques from convex optimization, our study opens the possibility to study effective complexity in

other types of networks where constraints other than those coming from the network structure dictate the network function.

METHODS

Models and their processing

The genome-scale metabolic models of 12 organisms (table S1) were obtained from their original publications (23, 25–35). Each reversible reaction was split into two irreversible reactions. The lower bounds for the irreversible reactions were set to zero, while the upper bounds were fixed to corresponding values in the original model. In the next step, the blocked reactions, i.e., reactions with absolute flux values below 10^{-6} mmol/gDW per hour in any feasible steady state supported by network [as determined by Flux Variability Analysis (36)], were removed from the original models. Note that the threshold to consider a reaction blocked can influence the number of identified balanced complexes [compare results of (18) where balanced complexes are identified using threshold 10^{-9}]. Optimum specific growth rate was determined per flux balance analysis (37) with the assumed reaction reversibility.

Identification of concordant complexes

Let \mathbf{Y} denote the nonnegative matrix of complexes, with rows denoting species and columns representing complexes. The entry y_{ij} denotes the stoichiometry with which species i enters the complex j . Let \mathbf{A} denote the incidence matrix of the directed graph with nodes representing complexes and edges denoting reactions. The rows of \mathbf{A} denote the complexes, and its columns stand for the reactions. Because the graph is directed, each column of \mathbf{A} has precisely one -1 and one 1 entry, corresponding to the substrate and product complexes of the respective reaction. The stoichiometry matrix is then given by $\mathbf{N} = \mathbf{YA}$. The activity of a complex i is the sum of fluxes around the complex, given by the i th entry of the vector \mathbf{Av} , denoted by $[\mathbf{Av}]_i$, where \mathbf{v} is a flux distribution. Two complexes i and j are concordant in the set of flux distributions $S = \{\mathbf{v} \mid \mathbf{Nv} = \mathbf{0}, \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}\}$ if, for every $\mathbf{v} \in S$, it holds that $[\mathbf{Av}]_i - \gamma_{ij}[\mathbf{Av}]_j = 0$, where γ_{ij} is a nonzero constant. This condition can be verified by determining that the minimum and maximum values of $[\mathbf{Av}]_i/[\mathbf{Av}]_j$ over all $\mathbf{v} \in S$ are nonzero and equal to each other. The latter can be ensured for two complexes i and j , which are not balanced, by solving two linear fractional programs

$$\min/\max [\mathbf{Av}]_i/[\mathbf{Av}]_j$$

subject to (s.t.)

$$\mathbf{YA}\mathbf{v} = \mathbf{N}\mathbf{v} = \mathbf{0}$$

$$\mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}$$

which can be transformed into four linear programs following the Charnes-Cooper transformation (38), leading to.

Min/ max $[\mathbf{Aw}]_i$	Min/ max $[\mathbf{Aw}]_j$
s.t.	s.t.
$\mathbf{YA}\mathbf{w} = \mathbf{N}\mathbf{w} = \mathbf{0}$	$\mathbf{YA}\mathbf{w} = \mathbf{N}\mathbf{w} = \mathbf{0}$
$[\mathbf{Aw}]_i = 1$	$[\mathbf{Aw}]_j = 1$
$\mathbf{v}_{\min}^t \leq \mathbf{w} \leq \mathbf{v}_{\max}^t$	$\mathbf{v}_{\min}^t \leq \mathbf{w} \leq \mathbf{v}_{\max}^t$
$t \geq 0$	$t \geq 0$

The complexes i and j are then concordant if the optima from the four linear programs coincide. The implementation of this approach is available at https://github.com/ankueken/concordant_complexes.

The concordance relation is reflexive (i.e., the ratio of the activity of a complex with itself is always of value one and, hence, constant), symmetric (i.e., the ratio of the activities of two complexes is constant if and only if the reciprocal ratio is constant), and transitive (i.e., if complex i is concordant with complex j , and complex j is concordant with complex k , then so are complexes i and k) and therefore represents an equivalence relation.

Fit of power law distributions

To test whether distributions of (i) the metabolite degree, as the number of complexes in which the metabolite participates; (ii) effective degree of a metabolite, as the number of concordance modules in which the metabolite participates; and (iii) the size of concordance modules, given by the number of complexes per module, follow power law distribution, we used the implementation of Broido and Clauset (20). In addition, the software is used to test whether other distributions (i.e., log-normal distribution, power law with exponential cutoff, and exponential and stretched exponential distribution) can better fit the abovementioned data.

Enrichment analysis

Enrichment analysis is performed on the basis of hypergeometric test to determine whether complexes participating in a given metabolic subsystem (extracted from the models) are overrepresented in concordance modules. The resulting P values are corrected for multiple hypotheses testing using the Benjamini-Hochberg procedure.

Randomized network variants

Using mass-balanced randomization (39), we create 58 random network variants for which we determine concordance modules. We test for differences in the number of concordance modules, the size of the largest concordance module, the mean concordance module size, the mean effective metabolite degree, the maximum effective metabolite degree, and the reducibility index using z scores, assuming that the null distributions over the ensemble of randomized metabolic networks are normal.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abl6962>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. F. Zorrilla, K. R. Patil, A. Zelezniak, metaGEM: Reconstruction of genome scale metabolic models directly from metagenomes. *bioRxiv*, 2020.12.31.424982 (2021).
2. C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, S. Y. Lee, Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).
3. M. Masid, M. Ataman, V. Hatzimanikatis, Analysis of human metabolism by reducing the complexity of the genome-scale models using redHUMAN. *Nat. Commun.* **11**, 2821 (2020).
4. I. Thiele, S. Sahoo, A. Heinken, J. Hertel, L. Heirendt, M. K. Aurich, R. Mt Fleming, Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol. Syst. Biol.* **16**, e8982 (2020).
5. T. Friedlander, A. E. Mayo, T. Tlustý, U. Alon, Evolution of bow-tie architectures in biology. *PLOS Comput. Biol.* **11**, e1004055 (2015).
6. H. Jeong, B. Tombor, R. Albert, Z. N. Oltval, A. L. Barabási, The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
7. E. Noor, E. Eden, R. Milo, U. Alon, Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* **39**, 809–820 (2010).

8. E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A. L. Barabási, Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843 (2004).
9. S. Robaina-Estévez, Z. Nikoloski, Flux-based hierarchical organization of *Escherichia coli*'s metabolic network. *PLoS Comput. Biol.* **16**, e1007832 (2020).
10. G. Basler, Z. Nikoloski, A. Larhlimi, A. L. Barabási, Y. Y. Liu, Control of fluxes in metabolic networks. *Genome Res.* **26**, 956–968 (2016).
11. A. P. Burgard, E. V. Nikolaev, C. H. Schilling, C. D. Maranas, Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* **14**, 301–312 (2004).
12. A. C. Müller, A. Bockmayr, Flux modules in metabolic networks. *J. Math. Biol.* **69**, 1151–1179 (2014).
13. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabási, Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
14. G. Shinar, M. Feinberg, Structural sources of robustness in biochemical reaction networks. *Science* **327**, 1389–1391 (2010).
15. S. Müller, G. Regensburger, Generalized mass action systems: Complex balancing equilibria and sign vectors of the stoichiometric and kinetic-order subspaces. *SIAM J. Appl. Math.* **72**, 1926–1947 (2012).
16. J. Neigenfind, S. Grimbs, Z. Nikoloski, On the relation between reactions and complexes of (bio)chemical reaction networks. *J. Theor. Biol.* **317**, 359–365 (2013).
17. M. Feinberg, in *Applied Mathematical Sciences (Switzerland)* (Springer, 2019), vol. 202, pp. i–454.
18. A. Küken, P. Wending, D. Langary, Z. Nikoloski, A structural property for reduction of biochemical networks. *Sci. Rep.* **11**, 17415 (2021).
19. D. Voet, J. G. Voet, *Biochemistry* (Wiley, ed. 4, 2010), pp. 1520.
20. A. D. Broido, A. Clauset, Scale-free networks are rare. *Nat. Commun.* **10**, 1017 (2019).
21. Q. Zhu, T. Qin, Y.-Y. Jiang, C. Ji, D.-X. Kong, B.-G. Ma, H.-Y. Zhang, Chemical basis of metabolic network organization. *PLoS Comput. Biol.* **7**, e1002214 (2011).
22. A. Akbari, J. T. Yurkovich, D. C. Zielinski, B. O. Palsson, The quantitative metabolome is shaped by abiotic constraints. *Nat. Commun.* **12**, 3178 (2021).
23. J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, B. Ø. Palsson, A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011).
24. D. J. Watts, S. H. Strogatz, Collective dynamics of “small-world” networks. *Nature* **393**, 440–442 (1998).
25. M. R. Andersen, M. L. Nielsen, J. Nielsen, Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol. Syst. Biol.* **4**, 178 (2008).
26. A. Arnold, Z. Nikoloski, Bottom-up metabolic reconstruction of *Arabidopsis* and its application to determining the metabolic costs of enzyme production. *Plant Physiol.* **165**, 1380–1391 (2014).
27. S. Imam, S. Schäuble, J. Valenzuela, A. López García De Lomana, W. Carter, N. D. Price, N. S. Balliga, A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. *Plant J.* **84**, 1239–1256 (2015).
28. M. N. Benedict, M. C. Gonnerman, W. W. Metcalf, N. D. Price, Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *J. Bacteriol.* **194**, 855–865 (2012).
29. A. M. Feist, J. C. M. Scholten, B. Palsson, F. J. Brockman, T. Ideker, Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol. Syst. Biol.* **2**, 2006.0004 (2006).
30. L.-E. Quek, L. K. Nielsen, On the reconstruction of the *Mus musculus* genome-scale metabolic network model. *Genome Inform.* **21**, 89–100 (2008).
31. X. Fang, A. Wallqvist, J. Reifman, Development and analysis of an in vivo-compatible metabolic network of *Mycobacterium tuberculosis*. *BMC Syst. Biol.* **4**, 160 (2010).
32. O. Gonzalez, T. Oberwinkler, L. Mansueto, F. Pfeiffer, E. Mendoza, R. Zimmer, D. Oesterheld, Characterization of growth and metabolism of the haloalkaliphile *Natronomonas pharaonis*. *PLoS Comput. Biol.* **6**, e1000799 (2010).
33. J. Nogales, B. Palsson, I. Thiele, A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: *i* JN746 as a cell factory. *BMC Syst. Biol.* **2**, 79 (2008).
34. Y. Zhang, I. Thiele, D. Weekes, Z. Li, L. Jaroszewski, K. Ginalski, A. M. Deacon, J. Wooley, S. A. Lesley, I. A. Wilson, B. Palsson, A. Osterman, A. Godzik, Three-dimensional structural view of the central metabolic network of *thermotoga maritima*. *Science* **325**, 1544–1549 (2009).
35. H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marčišauskas, P. M. Anton, D. Lappa, C. Lieven, M. E. Beber, N. Sonnenschein, E. J. Kerkhoven, J. Nielsen, A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).
36. S. Gudmundsson, I. Thiele, Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).
37. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
38. A. Charnes, W. W. Cooper, Programming with linear fractional functionals. *Nav. Res. Logist. Q.* **9**, 181–186 (1962).
39. G. Basler, O. Ebenhöf, J. Selbig, Z. Nikoloski, Mass-balanced randomization of metabolic networks. *Bioinformatics* **27**, 1397–1403 (2011).

Acknowledgments

Funding: This work was supported by Human Frontier Science Program grant RGP0046/2018 (to A.K. and Z.N.). **Author contributions:** Conceptualization: Z.N. Data collection: A.K. Formal analysis: A.K. and D.L. Investigation: A.K. and Z.N. Software: A.K. Validation: A.K. and Z.N. Writing—original draft: Z.N. Writing—reviewing and editing: A.K., D.L., and Z.N. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials and at <https://doi.org/10.5281/zenodo.5752480> or https://github.com/ankueken/concordant_complexes.

Submitted 29 July 2021

Accepted 8 February 2022

Published 30 March 2022

10.1126/sciadv.abl6962