Flux-dependent graphs for metabolic networks

Mariano Beguerisse-Díaz,[1, 2, *] Gabriel Bosque,[3, †] Diego Oyarzún,[2] Jesús Picó,[3] and Mauricio Barahona[2, ‡]

[1]*Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK*
[2]*Department of Mathematics, Imperial College London, London, SW7 2AZ, UK*
[3]*Institut Universitari d'Automàtica i Informàtica Industrial,*
Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain.

Cells adapt their metabolic fluxes in response to changes in the environment. We present a systematic flux-based framework for the construction of graphs to represent organism-wide metabolic networks. Our graphs encode the directionality of metabolic fluxes via links that represent the flow of metabolites from source to target reactions. The methodology can be applied in the absence of a specific biological context by modelling fluxes probabilistically, or can be tailored to different environmental conditions by incorporating flux distributions computed from constraint-based modelling such as Flux Balance Analysis. We illustrate our approach on the central carbon metabolism of *Escherichia coli*, and on a larger metabolic model of human hepatocytes, and study the proposed graphs under various environmental conditions and genetic perturbations. The results reveal drastic changes in the topological and community structure of the metabolic graphs, which capture the re-routing of metabolic fluxes under each growth and genetic condition. By integrating constraint-based models and tools from network science, our framework allows for the interrogation of context-specific metabolic responses beyond fixed, standard pathway descriptions.

I. INTRODUCTION

Metabolic reactions enable cellular function by converting nutrients into energy, and by assembling macromolecules that sustain the cellular machinery [1]. Cellular metabolism is usually thought of as a collection of pathways comprising enzymatic reactions associated with broad functional categories. Yet metabolic reactions are highly interconnected: enzymes convert multiple reactants into products with other metabolites acting as co-factors; enzymes can catalyse several reactions, and some reactions are catalysed by multiple enzymes, and so on. This enmeshed web of reactions is thus naturally amenable to network analysis, an approach that has been successfully applied to different aspects of cellular and molecular biology, e.g., protein-protein interactions [2], transcriptional regulation [3], or protein structure [4, 5].

Tools from network theory [6] have previously been applied to the analysis of structural properties of metabolic networks, including their degree distribution [7–10], the presence of metabolic roles [11], and their community structure [12–15]. A central challenge, however, is that there are multiple ways to construct a (mathematical) graph from a metabolic network [16]. For example, one can create a graph with metabolites as nodes and edges representing the reactions that transform one metabolite into another [7, 8, 17, 18]; a graph with reactions as nodes and edges corresponding to the metabolites shared among them [19–21]; or even a bipartite graph with both reactions and metabolites as nodes [22]. Importantly, the conclusions of graph-theoretical analyses are highly dependent on the chosen graph construction [23].

A key feature of metabolic reactions is the directionality of flux: metabolic networks contain both irreversible and reversible reactions, and reversible reactions can change their direction depending on the cellular and environmental contexts [1]. Many of the existing graph constructions, however, lead to undirected graphs that disregard such directional information, which is central to metabolic function [8, 16]. Furthermore, current graph constructions are usually derived from the whole set of metabolic reactions in an organism, and thus correspond to a generic metabolic 'blueprint' of the cell. However, cells switch specific pathways 'on' and 'off' to sustain their energetic budget in different environments [24]. Hence, such blueprint graphs might not capture the specific metabolic connectivity in a given environment, thus limiting their ability to provide biological insights in different growth conditions.

In this paper, we present a flux-based approach to construct metabolic graphs that encapsulate the directional flow of metabolites produced or consumed through enzymatic reactions. The proposed graphs can be tailored to incorporate flux distributions under different environmental conditions. To introduce our approach, we proceed in two steps. We first define the *Probabilistic Flux Graph* (PFG), a weighted, directed graph with reactions as nodes,

* beguerisse@maths.ox.ac.uk

† gabbosch@upv.es
‡ m.barahona@imperial.ac.uk

edges that represent supplier-consumer relationships between reactions, and weights given by the probability that a metabolite chosen at random is produced/consumed by the source/target reaction. This graph can be used to carry out graph-theoretical analyses of organism-wide metabolic organisation independent of cellular context or environmental conditions. We then show that this formalism can be adapted seamlessly to construct the *Metabolic Flux Graph* (MFG), a directed, environment-dependent, graph with weights computed from Flux Balance Analysis (FBA) [25], the most widespread method to study genome-scale metabolic networks.

Our formulation addresses several drawbacks of current constructions of metabolic graphs. Firstly, in our flux graphs, an edge indicates that metabolites are produced by the source reaction and consumed by the target reaction, thus accounting for metabolic directionality and the natural flow of chemical mass from reactants to products. Secondly, the Probabilistic Flux Graph discounts naturally the over-representation of pool metabolites (e.g., adenosine triphosphate (ATP), nicotinamide adenine dinucleotide (NADH), protons, water, and other co-factors) that appear in many reactions and tend to obfuscate the graph connectivity. Our construction avoids the removal of pool metabolites from the network, which can change the graph structure drastically [26–30]. Finally, the Metabolic Flux Graph incorporates additional biological information reflecting the effect of the environmental context into the graph construction. In particular, since the weights in the MFG correspond directly to fluxes (in units of mass per time), different biological scenarios can be analysed using balanced fluxes (e.g., from different FBA solutions) under different carbon sources and other environmental perturbations [16, 25, 31, 32].

After introducing the mathematical framework, we showcase our approach with two examples. Firstly, in the absence of environmental context, our analysis of the PFG of the core model of *Escherichia coli* metabolism [33] reveals the importance of including directionality and appropriate edge weights in the graph to understand the modular organisation of metabolic sub-systems. We then use FBA solutions computed for several relevant growth conditions for *E. coli*, and show that the structure of the MFG changes dramatically in each case (e.g., connectivity, ranking of reactions, community structure), thus capturing the environment-dependent nature of metabolism. Secondly, we study a model of human hepatocyte metabolism evaluated under different conditions for the wild-type and in a mutation found in primary hyperoxaluria type 1, a rare metabolic disorder [34], and show how the changes in network structure of the MFGs reveal new information that is complementary to the analysis of fluxes predicted by FBA.

II. RESULTS

A. Definitions and background

Consider a metabolic network composed of n metabolites with concentrations x_i $(i = 1, \ldots, n)$ that participate in m reactions

$$R_j : \quad \sum_{i=1}^{n} \alpha_{ij} x_i \rightleftharpoons \sum_{i=1}^{n} \beta_{ij} x_i, \quad j = 1, 2, \ldots, m, \tag{1}$$

where α_{ij} and β_{ij} are the stoichiometric coefficients of species i in reaction j. We can then define the n-dimensional vector of metabolite concentrations: $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^T$. Each reaction takes place at a rate $v_j(\mathbf{x}, t)$, measured in units of concentration per time [35].

The mass balance of the system can then be represented compactly by the system

$$\dot{\mathbf{x}} = \mathbf{S} \mathbf{v}, \tag{2}$$

where we have defined $\mathbf{v}(t) = (v_1(t), \ldots, v_m(t))^T$, the m-dimensional vector of reaction rates. The $n \times m$ matrix $\mathbf{S}$ is the stoichiometric matrix with entries $S_{ij} = \beta_{ij} - \alpha_{ij}$, i.e., the net number of x_i molecules produced (positive S_{ij}) or consumed (negative S_{ij}) by the j-th reaction. To illustrate the different schemes and graphs described in this paper, we use a toy example [32] of a metabolic network including nutrient uptake, biosynthesis of metabolic intermediates, secretion of waste products, and biomass production (Figure 1A).

Starting from the stoichiometry $\mathbf{S}$, there are several ways to construct a graph for a given set of metabolic reactions [16]. A common approach [16] is to define the *unipartite graph* with reactions as nodes and the $m \times m$ adjacency matrix

$$\mathbf{A} = \widehat{\mathbf{S}}^T \widehat{\mathbf{S}}, \tag{3}$$

where $\widehat{\mathbf{S}}$ is the boolean version of $\mathbf{S}$ (i.e., $\hat{S}_{ij} = 1$ when $S_{ij} \neq 0$ and $\hat{S}_{ij} = 0$ otherwise). In this *Reaction Adjacency Graph* (RAG), two reactions (nodes) are connected if they share metabolites, either as reactants or products, and self-loops represent the total number of metabolites that participate in a reaction (Fig. 1B).
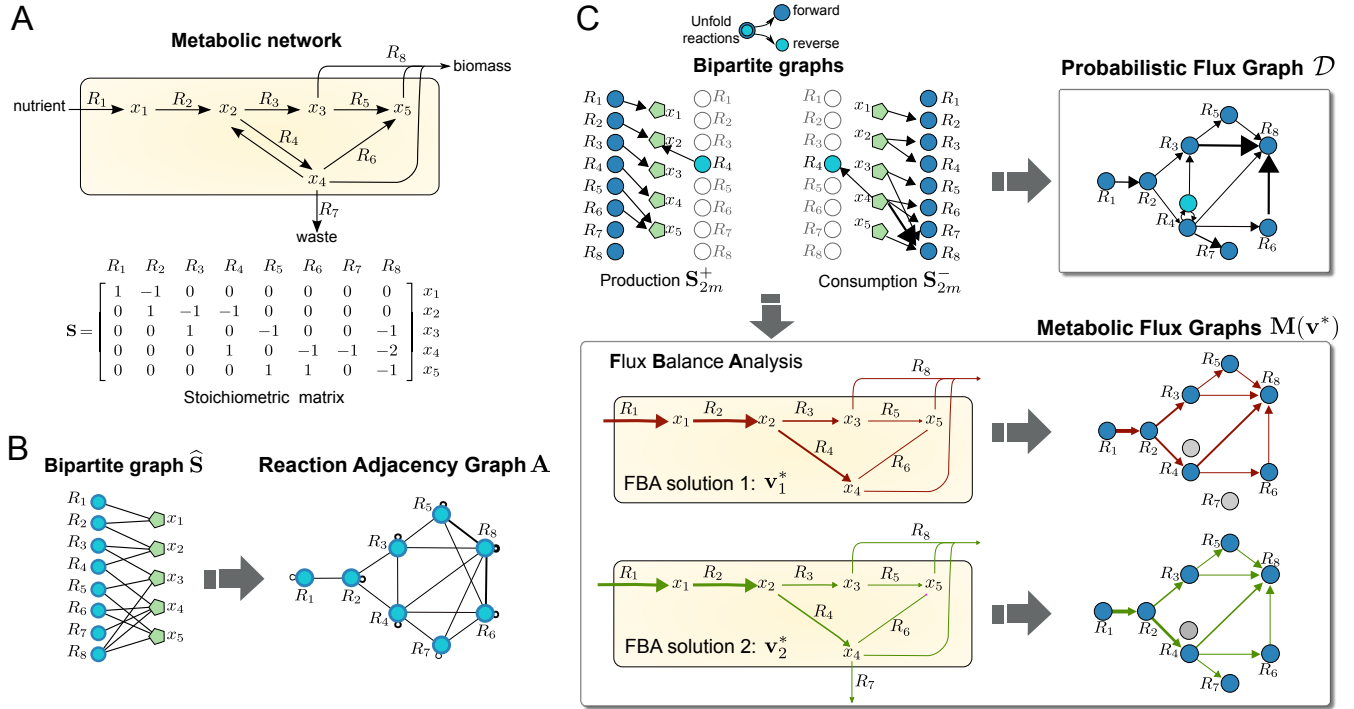
FIG. 1. **Graphs for metabolic networks.** (A) Toy metabolic network describing nutrient uptake, biosynthesis of metabolic intermediates, secretion of waste products, and biomass production [32]. The biomass reaction is $R_8 : x_3 + 2x_4 + x_5$. (B) Bipartite graph associated with the boolean stoichiometric matrix $\widehat{\mathbf{S}}$, and the standard Reaction Adjacency Graph (RAG [16]) with adjacency matrix $\mathbf{A} = \widehat{\mathbf{S}}^T \widehat{\mathbf{S}}$. The undirected edges of $\mathbf{A}$ indicate the number of shared metabolites among reactions. (C) The Probabilistic Flux Graph (PFG, $\mathcal{D}$) and two Metabolic Flux Graphs (MFG, $\mathbf{M}(\mathbf{v}^*)$) are constructed from the unfolded consumption and production stoichiometric matrices (5). Note that the reversible reaction R_4 is unfolded into two nodes. The PFG (7) is a directed graph with weights representing the probability that the source reaction produces a metabolite consumed by the target reaction. The MFGs (11) are constructed from two Flux Balance Analysis solutions ($\mathbf{v}_1^*$ and $\mathbf{v}_2^*$) obtained by optimizing a biomass objective function under different flux constraints representing different environmental or cellular contexts (see Sec. SI 2 in the Supplementary Information for details). The edges of the MFGs represent mass flow from source to target reactions, with weights in units of metabolic flux. The computed FBA solutions translate into different connectivity in the resulting MFGs.

Though widely studied [8, 16], the RAG has known important limitations, as it overlooks key aspects of the connectivity of metabolic networks. In particular, the RAG does not distinguish between forward and backward fluxes, nor does it incorporate information on the irreversibility of reactions ($\mathbf{A}$ is a symmetric matrix). Furthermore, the structure of $\mathbf{A}$ is dominated by the large number of connections introduced by pool metabolites (e.g., water, ions or enzymatic cofactors) that appear in many reactions. Although computational schemes have been designed to ameliorate the pool metabolite bias [27], their justification does not follow from biophysical considerations. Finally, the construction of the graph $\mathbf{A}$ from rigid topological criteria is not easily extended to incorporate the effect of varying environmental contexts.

B. Metabolic graphs that incorporate flux directionality and context

To address the limitations of the standard reaction adjacency graph $\mathbf{A}$, given in Eq. (3), and aiming at enhanced biophysical and biological interpretability, we propose a graph formulation that follows from a flux-based perspective. To construct our graph, we unfold each reaction into separate forward and reverse directions, and redefine the presence of links between reaction nodes to reflect producer-consumer relationships, i.e., two reactions are connected if one produces a metabolite that is consumed by the other. As shown below, this definition leads to graphs that naturally account for the reversibility of reactions and allows for a seamless integration of different biological contexts modelled

through FBA.

Inspired by matrix formulations of chemical reaction network kinetics [36], we unfold the reactions into forward and backward components. Specifically, let us rewrite the reaction rate vector introduced in (2) as:

$$\mathbf{v} := \mathbf{v}^+ - \mathbf{v}^- = \mathbf{v}^+ - \text{diag}(\mathbf{r})\,\mathbf{v}^-,$$

where $\mathbf{v}^+$ and $\mathbf{v}^-$ are non-negative vectors containing the forward and backward reaction rates, respectively, and $\mathbf{r}$ is the m-dimensional reversibility vector with components $r_j = 1$ if reaction R_j is reversible and $r_j = 0$ if it is irreversible. The $m \times m$ matrix $\text{diag}(\mathbf{r})$ contains $\mathbf{r}$ in its main diagonal. With these definitions, we rewrite the metabolic model in Eq. (2) in terms of the unfolded $2m$-dimensional vector of reaction rates, $\mathbf{v}_{2m} := [\mathbf{v}^+ \ \mathbf{v}^-]^T$, to obtain:

$$\dot{\mathbf{x}} = \mathbf{S}\mathbf{v} = \underbrace{\begin{bmatrix} \mathbf{S} & -\mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & 0 \\ 0 & \text{diag}(\mathbf{r}) \end{bmatrix}}_{\mathbf{S}_{2m}} \begin{bmatrix} \mathbf{v}^+ \\ \mathbf{v}^- \end{bmatrix} := \mathbf{S}_{2m}\,\mathbf{v}_{2m}, \tag{4}$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{S}_{2m}$ is an unfolded version of the stoichiometric matrix corresponding to the $2m$ forward and reverse reactions.

1. Probabilistic Flux Graph: a directional blueprint of metabolism

The unfolding into forward and backward fluxes leads us to the definition of *production* and *consumption* stoichiometric matrices:

$$\text{Production:} \quad \mathbf{S}_{2m}^+ = \frac{1}{2}\left(\text{abs}(\mathbf{S}_{2m}) + \mathbf{S}_{2m}\right)$$

$$\text{Consumption:} \quad \mathbf{S}_{2m}^- = \frac{1}{2}\left(\text{abs}(\mathbf{S}_{2m}) - \mathbf{S}_{2m}\right), \tag{5}$$

where $\text{abs}(\mathbf{S}_{2m})$ is the matrix whose entries are the absolute values of the corresponding entries of $\mathbf{S}_{2m}$. Note that each entry of the matrix $\mathbf{S}_{2m}^+$, denoted s_{ij}^+, gives the number of molecules of metabolite x_i produced by reaction R_j. Conversely, the entries of $\mathbf{S}_{2m}^-$, denoted s_{ij}^-, correspond to the number of molecules of metabolite x_i consumed by reaction R_j.

Within our directional flux framework, it is natural to consider a probabilistic description of producer-consumer relationships between reactions, as follows. Given a stoichiometric matrix $\mathbf{S}$, and in the absence of further biological information, the probability that metabolite x_k is produced by reaction R_i and consumed by reaction R_j is:

$$P\left(\text{a molecule of } x_k \text{ is produced by } R_i \text{ and consumed by } R_j\right) = \frac{s_{ki}^+}{w_k^+}\frac{s_{kj}^-}{w_k^-},$$

where $w_k^+ = \sum_{h=1}^{2m} s_{kh}^+$ and $w_k^- = \sum_{h=1}^{2m} s_{kh}^-$ are the total number of molecules of x_k produced and consumed by all reactions.

We thus define the weight of the edge between reaction nodes R_i and R_j as the probability that *any* metabolite chosen at random is produced by R_i and consumed by R_j. Summing over all metabolites and normalizing, the edge weight is defined as

$$\mathcal{D}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \frac{s_{ki}^+}{w_k^+}\frac{s_{kj}^-}{w_k^-}. \tag{6}$$

These edge weights are the entries of the adjacency matrix of the

$$\textbf{Probabilistic Flux Graph (PFG):} \quad \mathcal{D} = \frac{1}{n}\left(\mathbf{W}_+^\dagger \mathbf{S}_{2m}^+\right)^T \left(\mathbf{W}_-^\dagger \mathbf{S}_{2m}^-\right), \tag{7}$$

where $\mathbf{W}_+^\dagger = \text{diag}(\mathbf{S}_{2m}^+ \mathbb{1}_{2m})^\dagger$, $\mathbf{W}_-^\dagger = \text{diag}(\mathbf{S}_{2m}^- \mathbb{1}_{2m})^\dagger$, $\mathbb{1}_{2m}$ is a vector of ones, and $\dagger$ denotes the Moore-Penrose pseudoinverse. The PFG is a weighted, directed graph with a double-stochastic adjacency matrix ($\sum_{i,j} \mathcal{D}_{ij} = 1$). It provides a directional blueprint of the whole metabolic model, and naturally scales the contribution of pool metabolites to flux transfer. In Figure 1C we illustrate the creation of the PFG for a toy network. Note that our PFG is distinct from a directed graph directly comparable to the RAG $\mathbf{A}$, and with similar shortcomings, which can be constructed from boolean versions of the production and consumption stoichiometric matrices as shown in Sec. SI 1.

We now extend the idea behind the construction of the PFG to account for specific environmental context or growth conditions.

2. Metabolic Flux Graphs: incorporating information of the biological context

Cells adjust their metabolic fluxes to respond to the availability of nutrients and environmental requirements. Flux Balance Analysis (FBA) is a widely used method to predict environment-specific flux distributions. FBA computes a vector of metabolic fluxes $\mathbf{v}^*$ that maximise a cellular objective (e.g., biomass, growth or ATP production). The FBA solution is obtained assuming steady state conditions and subject to constraints that describe the availability of nutrients and other extracellular compounds [16]. The core elements of FBA are briefly summarised in Section A 1.

To incorporate the biological information afforded by FBA solutions into the structure of a metabolic graph, we again define the graph edges in terms of production and consumptions fluxes. Similarly to Eq. (4), we unfold the FBA solution vector $\mathbf{v}^*$ into forward and backward components, so that positive entries in the FBA solution correspond to forward fluxes, while negative entries in the FBA solution correspond to backward fluxes. From the unfolded fluxes

$$\mathbf{v}_{2m}^* = \begin{bmatrix} \mathbf{v}^{*+} \\ \mathbf{v}^{*-} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathrm{abs}\,(\mathbf{v}^*) + \mathbf{v}^* \\ \mathrm{abs}\,(\mathbf{v}^*) - \mathbf{v}^* \end{bmatrix},$$

we compute the vector of production and consumption fluxes as

$$\mathbf{j}(\mathbf{v}^*) = \mathbf{S}_{2m}^+ \mathbf{v}_{2m}^* = \mathbf{S}_{2m}^- \mathbf{v}_{2m}^*. \tag{8}$$

The k-th entry of $\mathbf{j}(\mathbf{v}^*)$ is the flux at which metabolite x_k is produced and consumed. Note that production and consumption fluxes are identical because of the steady state condition ($\dot{\mathbf{x}} = 0$ in Eq. (2)).

We now construct the flux graph by defining the weight of the edge between reactions R_i and R_j as the *total flux of metabolites produced by R_i that are consumed by R_j*. Assuming that the amount of metabolite produced by one reaction is distributed among the reactions that consume it in proportion to their flux, the flux of metabolite x_k from reaction R_i to R_j is given by

$$\text{Flux of } x_k \text{ from } R_i \text{ to } R_j = (\text{flux of } x_k \text{ produced by } R_i) \times \left(\frac{\text{flux of } x_k \text{ consumed by } R_j}{\text{total consumption flux of } x_k} \right). \tag{9}$$

For example, if the total flux of metabolite x_k is 10 mmol/gDW/h, with reaction R_i producing x_k at a rate 1.5 mmol/gDW/h and reaction R_j consuming x_k at a rate 3.0 mmol/gDW/h, then the flux of x_k from R_i to R_j is 0.45 mmol/gDW/h. Summing Eq. (9) over all metabolites, we then obtain the edge weight relating reactions R_i and R_j as

$$M_{ij}(\mathbf{v}^*) = \sum_{k=1}^{n} s_{ki}^+ v_{2m\,i}^* \times \left(\frac{s_{kj}^- v_{2m\,j}^*}{\sum_{j=1}^{2m} s_{kj}^- v_{2m\,j}^*} \right). \tag{10}$$

The edge weights are collected into the adjacency matrix of the

Metabolic Flux Graph (MFG): $\quad \mathbf{M}(\mathbf{v}^*) = \left(\mathbf{S}_{2m}^+ \mathbf{V}^* \right)^T \mathbf{J}_v^\dagger \left(\mathbf{S}_{2m}^- \mathbf{V}^* \right),$ $\qquad\qquad$ (11)

where $\mathbf{V}^* = \mathrm{diag}\,(\mathbf{v}_{2m}^*)$, $\mathbf{J}_v = \mathrm{diag}\,(\mathbf{j}(\mathbf{v}^*))$ and $\dagger$ denotes the matrix pseudoinverse. The MFG is a directed graph with weights in units of mmol/gDW/h. Self loops describe the metabolic flux of autocatalytic reactions, i.e., those in which products are also reactants.

The MFG provides a versatile framework to create environment-specific metabolic graphs from FBA solutions. In Figure 1C we illustrate the creation of MFGs for a toy network. In each case we compute FBA solutions under a fixed uptake flux and constrain the remaining fluxes to account for different biological scenarios. In scenario 1 the fluxes are constrained to be strictly positive and no larger than the nutrient uptake flux, while in scenario 2 we impose a positive lower bound on reaction R_7. The graph in scenario 2 displays an extra edge between reactions R_4 and R_7 and distinct edge weights, as compared to scenario 1 (see Sec. SI 2). The results thus illustrate how changes in the FBA solutions translate into different graph connectivities and edge weights.

C. Graphs of *Escherichia coli* metabolism

To illustrate our framework, we construct and analyse flux graphs (PFG and MFGs) of the core metabolic model of *E. coli* [33]. This model (Fig. 2A) contains 72 metabolites and 95 reactions, grouped into 11 pathways, which describe the main biochemical routes in central carbon metabolism [38–40]. See the Supplemental Spreadsheet for details about the reactions and metabolites in this model.
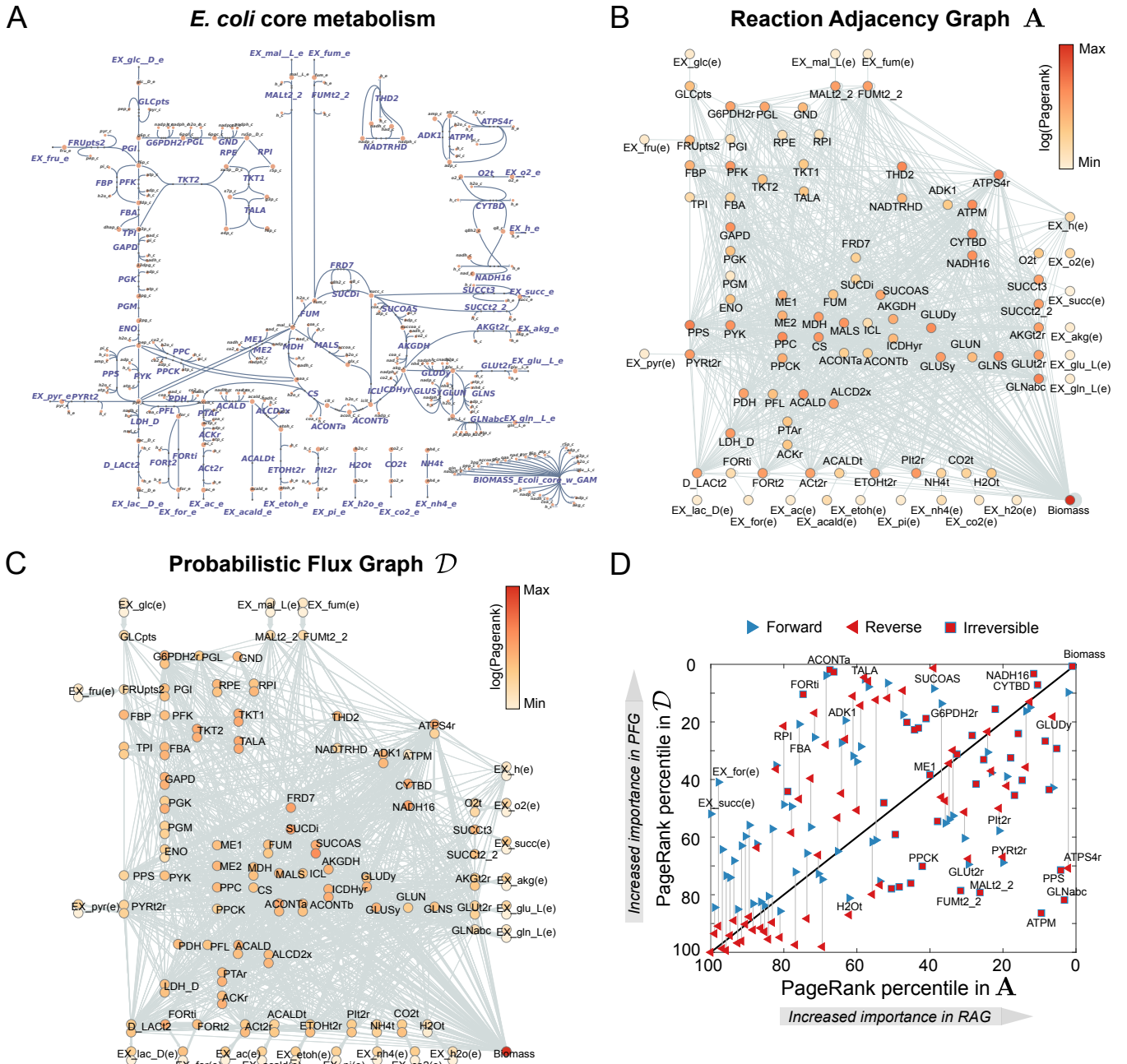
FIG. 2. **Graphs for the core metabolism of** *Escherichia coli*. (A) Map of the *E. coli* core metabolic model created with the online tool Escher [33, 37]. (B) The standard Reaction Adjacency Graph **A**, as given by Eq. (3). The nodes represent reactions; two reactions are linked by an undirected edge if they share reactants or products. The nodes are coloured according to their PageRank score, a measure of their centrality in the graph. (C) The directed Probabilistic Flux Graph **𝒟**, as computed from Eq. (7). The reversible reactions are unfolded into two overlapping nodes (one for the forward reaction, one for the backward). The directed links indicate flow of metabolites produced by the source node and consumed by the target node. The nodes are coloured according to their PageRank score. (D) Comparison of PageRank percentiles of reactions in **A** and **𝒟**. Reversible reactions are represented by two triangles connected by a line; both share the same PageRank in **A**, but each has its own PageRank in **𝒟**. Reactions that appear above (below) the diagonal have increased (decreased) PageRank in **𝒟** as compared to **A**.

1. Probabilistic Flux Graph: the impact of directionality

To highlight the effect of flux directionality on the constructed graphs, Figure 2 compares the standard undirected Reaction Adjacency Graph (**A**) and our proposed Probabilistic Flux Graph ($\mathcal{D}$) for the same metabolic model. The **A** graph has 95 nodes and 1,158 undirected edges, while the $\mathcal{D}$ graph has 154 nodes and 1,604 directed and weighted edges. The increase in node count is due to the unfolding of forward and backward reactions into separate nodes. Unlike the **A** graph, where the edges represent shared metabolites between two reactions, the directed edges of the $\mathcal{D}$ graph represent the flow of metabolites from a source to a target reaction. A salient feature of both graphs is their high connectivity, which is not apparent from the traditional pathway representation in Figure 2A.

The impact of directionality becomes apparent when comparing the importance of reaction nodes on the overall connectivity of each graph, as measured by the PageRank score introduced in the original Google algorithm [41, 42]. Figure 2B–D shows that the PageRank of reactions is substantially different in **A** and $\mathcal{D}$. The overall ordering is maintained: exchange reactions tend to have low PageRank, whereas core metabolic reactions have high PageRanks in both graphs—indeed, the biomass reaction has the highest rank in both cases (see Supplemental Spreadsheet). However, we observe a dramatic change in the importance of some reactions. For example, the reactions for ATP maintenance (ATPM, irreversible), phosphoenolpyruvate synthase (PPS, irreversible) and ABC-mediated transport of L-glutamine (GLNabc, irreversible) drop from being among the top 10% most important reactions in the **A** graph to the bottom percentiles in the $\mathcal{D}$ graph. Conversely, other reactions such as aconitase A (ACONTa, irreversible), transaldolase (TALA, reversible) and succinyl-CoA synthetase (SUCOAS, reversible), and formate transport via diffusion (FORti, irreversible) gain substantial importance in the $\mathcal{D}$ graph. For instance, FORti is the sole consumer of formate, which is produced by pyruvate formate lyase (PFL), a reaction that is highly connected to the rest of the network. Importantly, in most of the reversible reactions, such as ATP synthase (ATPS4r), there is a wide gap between the PageRank of the forward and backward reactions, suggesting a marked asymmetry in the importance of metabolic flows.

Community detection is a technique that is frequently used for the analysis of complex graphs: nodes are clustered into tightly-related communities in order to reveal the coarse-grained structure of the graph, potentially at different levels of resolution [45–47]. Indeed, the community structure of graphs derived from metabolic networks has been the subject of several analyses [12, 14, 45]. However, most existing community detection methods are only applicable to undirected graphs and fail to capture the directionality of the edges, a key feature in metabolism. In order to account for directionality, we use the Markov Stability community detection framework [47–49], which employs diffusions on graphs to detect groups of nodes where flows are retained persistently across time scales. Due to its use of diffusive dynamics, Markov Stability is ideally suited to find multi-resolution community structure [46], and naturally incorporates edge directionality, if present [47, 50] (see Sec. A 2). When applied to metabolic graphs, Markov Stability can thus reveal groups of reactions that are tightly linked by the flow of metabolites they produce and consume.

Figure 3 highlights the strong differences between the community structure of the undirected RAG and the directed PFG of the core metabolism of *E. coli*, underscoring the importance of directionality in these graphs. When applied to the **A** graph, Markov Stability reveals a robust partition into seven communities (Figure 3B, see also Sec. SI 3). The reaction communities obtained are largely determined by the edges created by abundant pool metabolites. For example, community C1(**A**) is mainly composed of reactions that consume or produce ATP and water. Note, however, that the biomass reaction (the largest consumer of ATP) is not a member of C1(**A**) because, in the **A** graph construction, any connection involving ATP has equal weight. Other communities in **A** are also determined by pool metabolites, e.g. C2(**A**) is dominated by H^+, and C3(**A**) is dominated by NAD^+ and $NADP^+$, as shown by the word clouds representing the relative frequency of metabolites that appear in the reactions contained in each community. The community structure in **A** thus reflects the limitations of this graph construction due to the absence of biological context and the large number of uninformative links introduced by pool metabolites.

In contrast, we found a robust partition into five communities for the $\mathcal{D}$ graph (Figure 3C, see also Sec. SI 3). These communities comprise reactions related consistently by biochemical pathways. Community C1($\mathcal{D}$) contains the reactions in the pentose phosphate pathway together with the first steps of glycolysis involving D-fructose, D-glucose, or D-ribulose. Community C2($\mathcal{D}$) contains the main reactions that produce ATP from substrate level as well as oxidative phosphorylation and the biomass reaction. Community C3($\mathcal{D}$) includes the core of the citric acid cycle, anaplerotic reactions related to malate syntheses, as well as the intake of cofactors such as CO_2. Community C4($\mathcal{D}$) contains reactions that are secondary sources of carbon (such as malate and succinate), as well as oxidative phosphorilation reactions. Finally, community C5($\mathcal{D}$) contains reactions that are part of the pyruvate metabolism subsystem, as well as transport reactions for the most common secondary carbon metabolites such as lactate, formate, acetaldehyde and ethanol. Altogether, the communities of the $\mathcal{D}$ graph reflect metabolite flows associated with specific cellular functions, a key benefit and consequence of including flux directionality in the graph construction. As seen in Fig. 3C, the communities are no longer exclusively determined by pool metabolites (e.g., water is no longer dominant and protons are spread among all communities). For a more detailed explanation and comparison of the communities
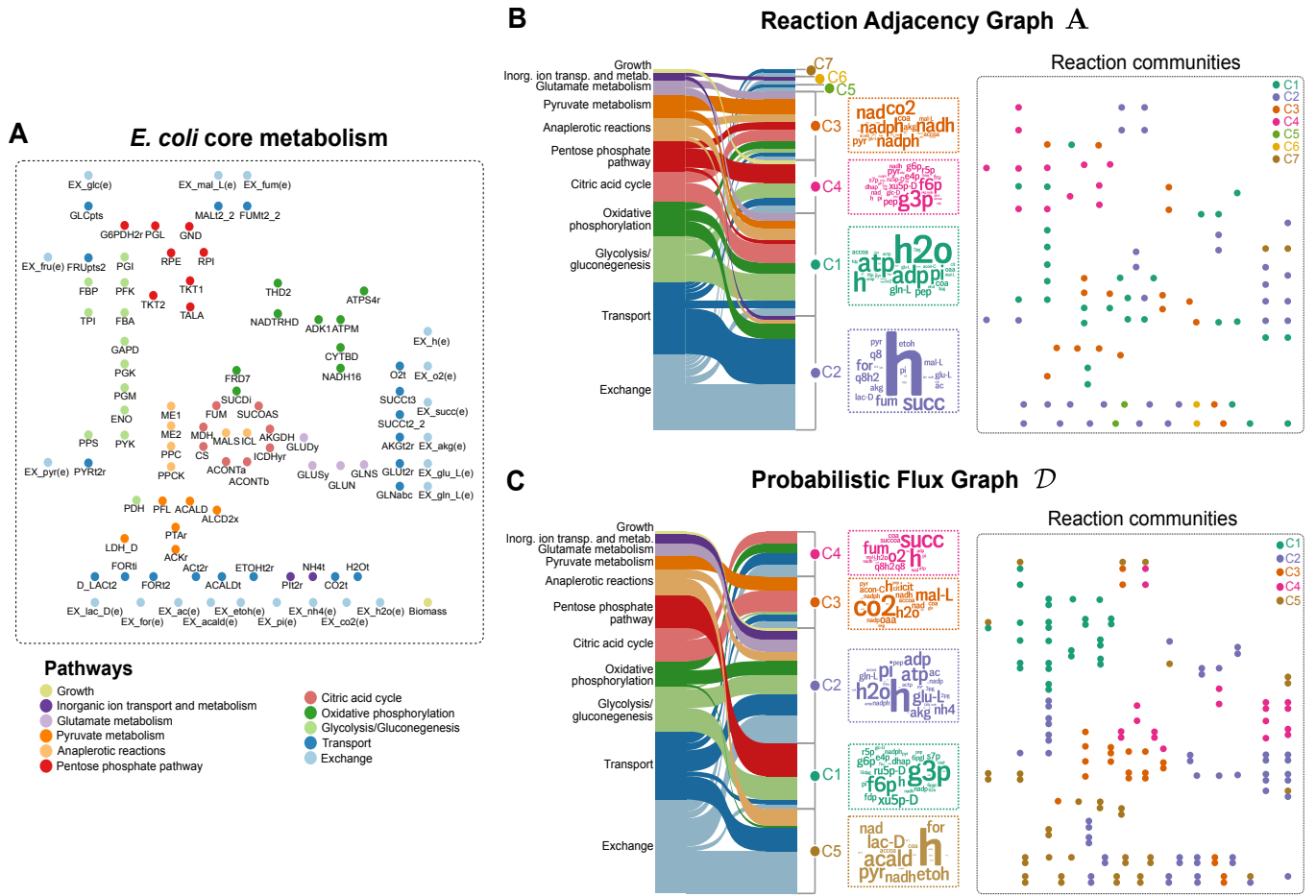
FIG. 3. **Directionality and community structure of graphs for *Escherichia coli* metabolism.** (A) Reactions of the core model of *E. coli* metabolism grouped into eleven biochemical pathways. (B–C) Graphs **A** and $\mathcal{D}$ from Fig. 2B–C partitioned into communities computed with the Markov Stability method; for clarity, the graph edges are not shown. The Sankey diagrams [43, 44] show the correspondence between biochemical pathways and the communities found in each graph. The word clouds contain the metabolites that participate in the reactions each community, and the word size is proportional to the number of reactions in which each metabolite participates.

found in the **A** and $\mathcal{D}$ graphs, see Section SI 3 and the Supplementary Spreadsheet.

2. *Metabolic Flux Graphs: the impact of growth conditions and biological context*

To incorporate the impact of environmental context in our graphs, we construct different Metabolic Flux Graphs (11), using flux distributions obtained from Flux Balance Analysis applied to the core model of *E. coli* metabolism under several growth conditions: aerobic growth in rich media with glucose or ethanol, aerobic growth in glucose but phosphate- and ammonium-limited, and anaerobic growth in glucose.

The results, summarised in Figure 4, reveal how changes in metabolite flows induced by different biological contexts are reflected in our graph construction. In all cases, the MFGs have fewer nodes than the blueprint graph $\mathcal{D}$ because the FBA solutions contain numerous reactions with zero flux. The different environments also affect the graph connectivity and the relative node importance, as measured by their PageRank score. Furthermore, the community structure of the MFGs for the four environmental conditions, as obtained with the Markov Stability framework, reflect the distinct usage of functional pathways by the cell in response to growth requirements under specific environments. We briefly describe the salient features of the analysis; a more detailed discussion can be found in Section SI 4 and Fig. SI 2 in the Supplementary Information, and the Supplemental Spreadsheet.

FIG. 4. **Metabolic Flux Graphs for *Escherichia coli* in different growth conditions.** The MFGs are computed from Eq. (11) and the FBA solutions in four different environments: (A) aerobic growth in D-glucose, (B) aerobic growth in ethanol, (C) anaerobic growth in D-glucose, and (D) aerobic growth in D-glucose but with limited ammonium and phosphate. Each subfigure shows: (left) flux map obtained with Escher [37], where the increased red colour of the arrows indicates increased flux; (centre) Metabolic Flux Graph with nodes coloured according to their PageRank (zero flux reactions are in grey; thickness of connections proportional to fluxes); (right) community structure computed with the Markov Stability method together with Sankey diagrams showing the correspondence between biochemical pathways and MFG communities.

a. Aerobic growth in D-glucose ($\mathbf{M}_{glc}$). We observe a robust partition into three communities with an intuitive biological interpretation (Fig. 4A and Fig. SI 2A). Firstly, C1($\mathbf{M}_{glc}$) can be thought of as a carbon-processing community, comprising reactions that process carbon from D-glucose to pyruvate including most of the glycolysis and pentose phosphate pathways, together with related transport and exchange reactions. Secondly, C2($\mathbf{M}_{glc}$) harbours the bulk of reactions related to oxidative phosphorylation and the production of energy in the cell, including the electron transport chain of NADH dehydrogenase, cytochrome oxidase, and ATP synthase, as well as transport reactions for phosphate and oxygen intake and proton balance. The growth reaction is also included in community C2($\mathbf{M}_{glc}$), consistent with ATP being the main substrate for both the ATP maintenance (ATPM) requirement and the biomass reaction in this biological scenario. Finally, C3($\mathbf{M}_{glc}$) contains reactions related to the citric acid cycle (TCA) and the production of NADH and NADPH (i.e., the cell's reductive power), as well as routes that take phosphoenolpyruvic acid (PEP) as a starting point, thus highlighting carbon intake routes that are strongly linked to the TCA cycle.

b. Aerobic growth in ethanol ($\mathbf{M}_{etoh}$). We found a robust partition into three communities that resemble those found in $\mathbf{M}_{glc}$, with subtle yet important differences (Fig. 4B and Fig. SI 2B). The most salient differences are observed in the carbon-processing community C1($\mathbf{M}_{etoh}$), which clearly reflects the switch of carbon source from D-glucose to ethanol. This community contains gluconeogenic reactions (instead of glycolytic), due to the reversal of flux induced by the change of carbon source, as well as anaplerotic reactions and reactions related to glutamate metabolism. The main role of the reactions in this community is the production of bioprecursors such as PEP, pyruvate, 3-phospho-D-glycerate (3PG), glyceraldehyde-3-phosphate (G3P), D-fructose-6-phosphate (F6P), and D-glucose-6-phosphate, all of which are substrates for growth. Consequently, the biomass reaction is also grouped within C1($\mathbf{M}_{etoh}$) due the increased metabolic flux of precursors relative to ATP production in this biological scenario. The other two reaction communities (energy-generation C2($\mathbf{M}_{etoh}$) and citric acid cycle C3($\mathbf{M}_{etoh}$)) display less prominent differences relative to the $\mathbf{M}_{glc}$ graph, with additional pyruvate metabolism and anaplerotic reactions as well as subtle ascriptions of reactions involved in NADH/NADPH balance and the source for acetyl-CoA.

c. Anaerobic growth in D-glucose ($\mathbf{M}_{anaero}$). The absence of oxygen has a profound impact on the metabolic balance of the cell and the MFG captures the drastic changes in this new regime effectively (Fig. 4C and Fig. SI 2C). Both the connectivity and the reaction communities in this MFG are different from the aerobic scenarios, with a much diminished presence of oxidative phosphorylation pathways and the absence of the first two steps of the electron transport chain (CYTBD and NADH16). We found that $\mathbf{M}_{anaero}$ has a robust partition into four communities. C1($\mathbf{M}_{anaero}$) still contains carbon processing (glucose intake and glycolysis), yet these reactions are decoupled from the pentose phosphate pathway, which is now part of community C3($\mathbf{M}_{anaero}$) grouped with the citric acid cycle (now incomplete) and the biomass reaction. C3($\mathbf{M}_{anaero}$) includes the growth precursors in this scenario, including alpha-D-ribose-5-phosphate (r5p), D-erythrose-4-phosphate (e4p), 2-oxalacetate and NADPH. The other two communities are specific to the anaerobic context: C2($\mathbf{M}_{anaero}$) contains the conversion of PEP into formate (more than half of the carbon secreted by the cell becomes formate [51]); and C4($\mathbf{M}_{anaero}$) includes NADH production and consumption via reactions linked to glyceraldehyde-3-phosphate dehydrogenase (GAPD).

d. Aerobic growth in D-glucose but limited phosphate and ammonium ($\mathbf{M}_{lim}$). Under these growth-limiting conditions, we found a robust partition into three communities (Fig. 4D and Fig. SI 2D). The community structure reflects *overflow metabolism* [52], which occurs when the cell takes in more carbon than it can process. As a consequence, the excess carbon is secreted from the cell, leading to a strong decrease in growth and a partial shutdown of the citric acid cycle. This is reflected in the reduced weight of the TCA pathway in C3($\mathbf{M}_{lim}$) and its grouping with the secretion routes of acetate and formate. Hence, C3($\mathbf{M}_{lim}$) comprises reactions that would not be strongly coupled in more favorable growth conditions, yet are linked together by metabolic responses appearing due to the limited availability of ammonium and phosphate. Furthermore, the carbon-processing community C1($\mathbf{M}_{lim}$) contains the glycolytic pathway, but detached from the pentose phosphate pathway, as in the $\mathbf{M}_{anaero}$ graph, highlighting its role in precursor formation. The bioenergetic machinery is contained in community C2($\mathbf{M}_{lim}$), including the pentose phosphate pathway, with a smaller role for the electron transport chain (21.8% of the total ATP as compared to 66.5% in $\mathbf{M}_{glc}$).

3. Multiscale organisation of metabolic flux graphs

Another advantage of flux-based MFGs is the possibility of applying network-theoretic tools to detect natural groupings of reactions at different levels of resolution, as well as their hierarchical relationship across scales. The Markov Stability framework [48, 53] can be used to detect multi-resolution community structure in directed graphs (Sec. A 2), thus allowing the exploration of the modular multiscale organisation of metabolic reaction networks.

Figure 5 illustrates this multiscale analysis on the metabolic flux graph of *E. coli* under aerobic growth in glucose ($\mathbf{M}_{glc}$). By varying the Markov time t, a parameter in the Markov Stability method, we scanned the community structures at different resolutions. Our results show that, as we move from finer to coarser resolutions, the MFG can
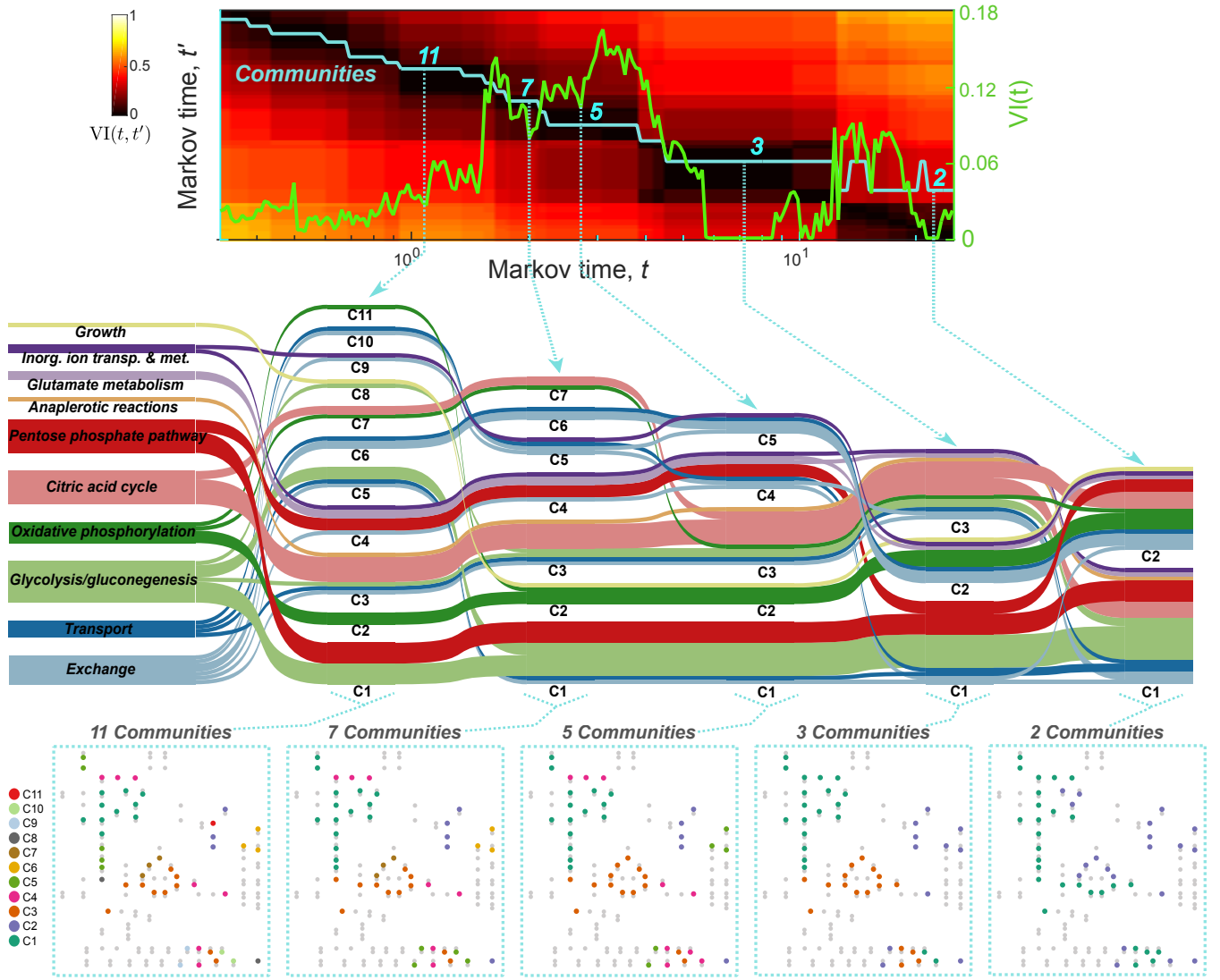
FIG. 5. **Community structure of flux graphs across different scales.** We applied the Markov Stability method to partition the metabolic flux graph for *E. coli* aerobic growth in glucose ($\mathbf{M}_{\text{glc}}$) across levels of resolution. The top panel shows the number of communities of the optimal partition (blue line) and two measures of its robustness ($VI(t)$ (green line) and $VI(t, t')$ (colour map)) as a function of the Markov time t (see text and Methods section). The five Markov times selected correspond to robust partitions of the graph into 11, 7, 5, 3, and 2 communities, as signalled by extended low values of $VI(t, t')$ and low values (or pronounced dips) of $VI(t)$. The Sankey diagram (middle panel) visualises the multiscale organisation of the communities of the flux graph across Markov times, and the relationship of the communities with the biochemical pathways. The bottom panel shows the five partitions at the selected Markov times. The partition into 3 communities corresponds to that in Figure 4A.

be partitioned into 11, 7, 5, 3, and 2 communities which have high robustness across Markov time (extended plateaux of optimality over t, as shown by the low values of $VI(t, t')$) and are highly robust within the optimisation ensemble (as shown by dips in $VI(t)$). For further details, see Section A 2 and Refs. [46–48, 53].

The Sankey diagram in Fig. 5 allows us to visualise the pathway composition of the graph partitions and their relationships across different resolutions. As we decrease the resolution (longer Markov times), the reactions in different pathways assemble and split into different groupings, reflecting both specific relationships and general organisation principles associated with this growth condition. A general observation is that glycolysis is grouped together with oxidative phosphorylation across most scales, underlining the fact that those two pathways function as cohesive metabolic sub-units in aerobic conditions. In contrast, the exchange and transport pathways appear spread among

multiple partitions across all resolutions. This is expected, as these are enabling functional pathways in which reactions do not interact amongst themselves but rather feed substrates to other pathways.

Other reaction groupings reflect more specific relationships. For example, the citric acid cycle (always linked to anaplerotic reactions) appears as a cohesive unit across most scales, and only splits in two in the very final flux grouping, reflecting the global role of the TCA cycle in linking to both glycolysis and oxidative phosphorylation. The pentose phosphate pathway, on the other hand, is split into two groups (one linked to glutamate metabolism and another one linked to glycolysis) across early scales, only merging into the same community towards the final groupings. This suggests a more interconnected flux relationship of the different steps of the penthose phosphate pathway with the rest of metabolism. Figure SI 2 contains a multiscale analyses of the communities for the other three growth scenarios.

D. Hepatocyte metabolism in wild type and PH1 mutant human cells

To showcase the applicability of our framework to larger metabolic models, we analyse a model of human hepatocyte (liver) metabolism with 777 metabolites and 2589 reactions [34], which extends the widely used HepatoNet1 model [54] with 50 reactions and 8 metabolites. This model was used in Ref. [34] to compare wild-type cells (WT) and cells lacking alanine:glyoxylate aminotransferase (AGT) as a result of a genetic mutation in the rare disease primary hyperoxaluria type 1 (PH1). The enzyme AGT is found in peroxisomes and its mutation decreases the breakdown of glyoxylate, with subsequent accumulation of calcium oxalate that leads to liver damage.

Following [34], we obtain 442 FBA solutions for different sets of metabolic objectives for both the wild type (WT) model and the PH1 model lacking AGT (reaction r2541). We then generate the corresponding 442 MFGs for WT and 442 MFGs for PH1, and we obtain the averages over each ensemble: $\overline{\mathbf{M}}_{\mathrm{WT}}$ and $\overline{\mathbf{M}}_{\mathrm{PH1}}$. Of the 2589 reactions in the model, 2448 forward and 1362 reverse reactions are present in at least one of the FBA solutions. Hence the average MFGs have 3810 nodes each (see the Supplementary Spreadsheet for details about the reactions).

Figure 6A shows the MFG for the wild type ($\overline{\mathbf{M}}_{\mathrm{WT}}$) coloured according to a robust partition into 7 communities obtained with Markov Stability. The seven communities are broadly linked to amino acid metabolism (C0), energy metabolism (C1 and C5), glutathione metabolism (C2), fatty acid and bile acid metabolism (C3 and C4) and cholesterol metabolism and lipoprotein particle assembly (C6). As expected, the network community structure of the MFG is largely preserved under the AGT mutation: the Sankey diagramme in Fig. 6A shows a remarkable match between the partitions of $\overline{\mathbf{M}}_{\mathrm{WT}}$ and $\overline{\mathbf{M}}_{\mathrm{PH1}}$ found independently with Markov Stability. Despite this similarity, our method also identified subtle but important differences between the healthy and diseased networks. In particular, C3 in $\overline{\mathbf{M}}_{\mathrm{PH1}}$ receives 60 reactions, almost all taking place in the peroxisome and linked to mevalonate and iso-pentenyl pathways, as well as highly central transfer reactions of PP_i, O_2 and H_2O_2 between the peroxisome and the cytosol (r1152, r0857, r2577) .

Similarly, the network centrality of most reactions is relatively unaffected by the PH1 mutation. Figure 6B shows a good overall correlation between the PageRank percentiles in $\overline{\mathbf{M}}_{\mathrm{WT}}$ and $\overline{\mathbf{M}}_{\mathrm{PH1}}$ but with some notable exceptions. Indeed, the reactions that exhibit the largest change in network centrality (labelled in Fig. 6B) provide biological insights related to the disease state. Specifically, the reactions that undergo the largest decrease in centrality in the PH1 network are largely related to VLDL-pool reactions, whereas the four reactions (r0916, r1088, r2384, r2374) that have the largest increase in centrality in the PH1 network are related to the transfer of citrate out of the cytosol in exchange for oxalate and PEP. Note that although oxalate and citrate reactions are directly linked to metabolic changes associated with the PH1 diseased state, none of them exhibits large changes in their flux predicted by FBA, yet they show large changes in network centrality.

These observations underscore that the information provided by our network analysis is complementary to the analysis drawn from FBA predictions. As shown in Figure 6C, a group of reactions exhibit large gains or decreases in their flux under the PH1 mutation with relatively small changes in their centrality scores. Closer inspection reveals that most of these reactions are close to the AGT reaction (r2541) in the perturbed pathway and involve the conversion of glycolate, pyruvate, glycine, alanine and serine. These observed changes in flux follow from the local rearrangement of network flows consequence of the deletion of reaction r2541. On the other hand, the citrate and oxalate reactions discussed above, which have large increases or decreases of centrality with small changes in flux, reflect global changes in the flow structure of the network. Interestingly, the transport reactions of O_2, H_2O_2, serine and hydroxypyruvate between cytosol and peroxisome (r0857, r2577, r2583, r2543) undergo large changes both in centrality and flux, highlighting the importance of peroxisome reactions in PH1. We provide a full spreadsheet with these analyses as Supplementary Material for the interested reader.
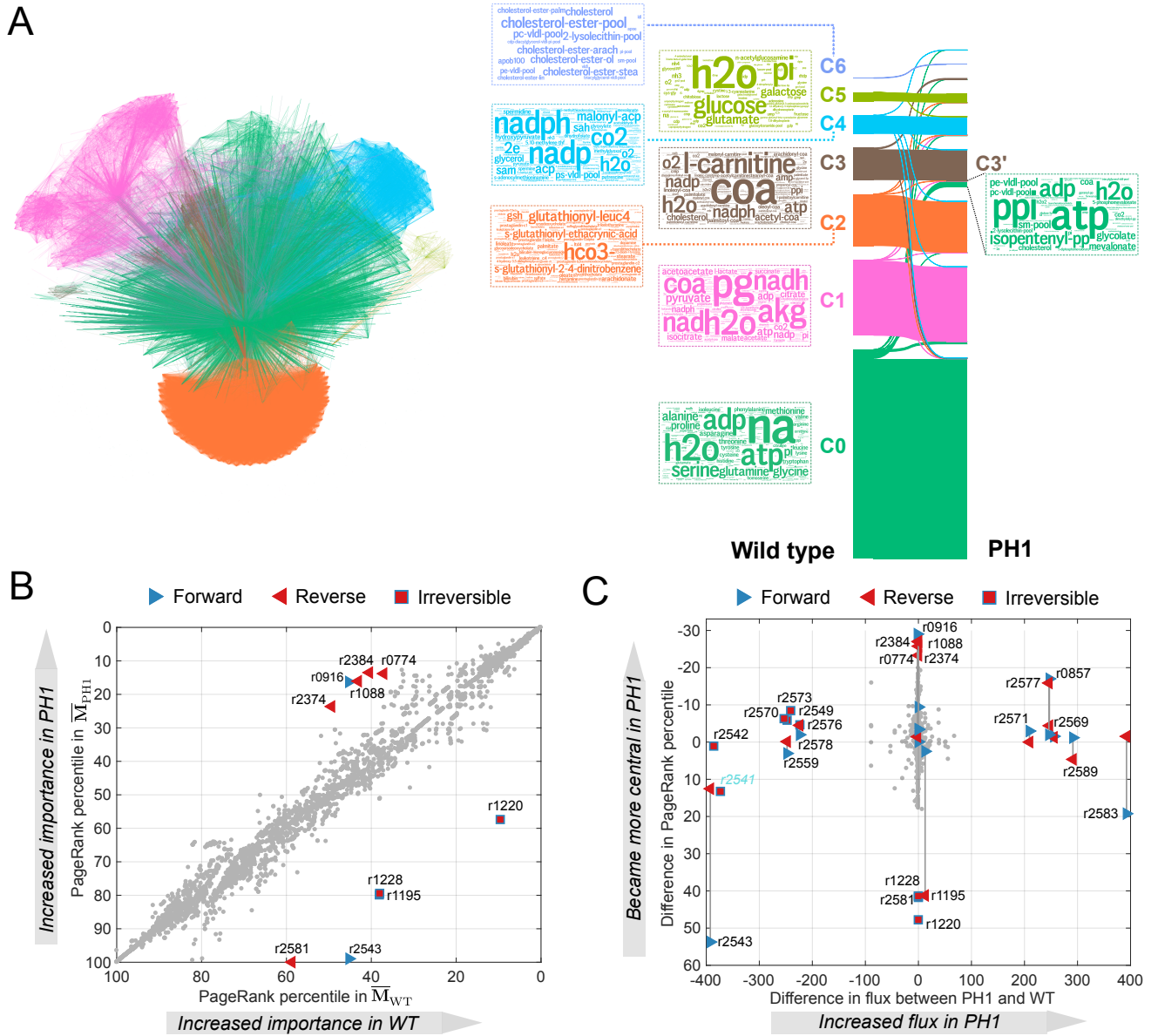
FIG. 6. MFG analysis of a model of human hepatocyte metabolism and the genetic condition PH1.
(A) Average MFG of hepatocytes in wild-type cells over 442 metabolic objectives. The reaction nodes are coloured according to their community in a robust partition obtained with Markov Stability. The Sankey diagramme shows the consistency between the communities independently found in the MFGs of the wild type and the mutated PH1 cells. The word clouds of the most frequent metabolites in the reactions of the communities in WT reveal functional groupings (see text). The main reorganisation of the community structure under the PH1 mutation is summarised by the word cloud of the metabolites that join C3' in PH1 from other communities in WT. (B) Comparison of the PageRank percentiles in the WT and PH1 MFGs. Reactions whose rank changes by more than 20 percentiles are labelled and shown in colour. (C) Difference in average FBA flux between WT and PH1 vs difference in PageRank percentile between WT and PH1. The reactions whose flux difference is greater than 100mmol/gDW/h or whose change in PageRank percentile is greater than 20 are labelled and shown in colour. The differences in centrality PageRank score provide complementary information revealing additional important reactions affected by the PH1 mutation which affects reaction r2541 (indicated with italics).

III. DISCUSSION

Metabolic reactions are commonly understood in terms of functional pathways that are heavily interconnected to form metabolic networks, i.e., metabolites linked by arrows representing enzymatic reactions between them [37] (Figs. 2 and 4). However, such standard representations are not amenable to rigorous graph-theoretic analysis. Importantly, there are fundamentally different graphs that can be constructed from the metabolic reaction information depending on the chosen representation of species/interactions as nodes/edges, e.g., reactions as nodes; metabolites as nodes; or both reaction and metabolites as nodes [16]. Each one of these graphs can be directed or undirected, and with weighted links computed according to different rules. The choices and subtleties in graph construction are crucial both to capture the relevant metabolic information and to interpret their topological properties [10, 17].

Here, we have presented a flux-based strategy to build graphs for metabolic networks. Our graphs have reactions as nodes and directed edges representing the flux of metabolites produced by a source reaction and consumed by a target reaction. This principle can be applied to build both 'blueprint' graphs (PFG), which summarise the probabilistic fluxes of the whole metabolism of an organism, as well as context-specific graphs (MFGs), which reflect specific environmental conditions. The blueprint Probabilistic Flux Graph with edge weights equal to the probability that the source/target reactions produce/consume a molecule of a metabolite chosen at random, naturally tames the over-representation of pool metabolites without the need to remove them from the graph arbitrarily, as is often done in the literature [26, 28–30]. The PFG can be used to construct networks when the stoichiometric matrix is the only information available. The context-specific Metabolic Flux Graphs (MFGs) incorporate the effect of the environment, as edge weights correspond to the total flux of metabolites between reactions as calculated by Flux Balance Analysis (FBA). Computing FBA solutions for different environments allows us to build metabolic graphs systematically for different growth media.

The two proposed graphs provide complementary tools for studying the organisation of metabolism and can be embedded into virtually any FBA-based modelling pipeline. Specifically, the PFG relies on the availability of a well-curated stoichiometric matrix, which is produced with metabolic reconstruction techniques that typically precede the application of FBA. The MFG, on the other hand, explicitly uses the FBA solutions in its construction. Both methods provide a systematic framework to convert genome-scale metabolic models into a directed graph on which powerful analysis tools from network theory can be applied.

To exemplify our approach, we built and analysed PFG and MFGs for the core metabolism of *E. coli*. Through the analysis of topological properties and community structure of these graphs, we highlighted the importance of weighted directionality in metabolic graph construction and revealed the flux-mediated relationships between functional pathways under different environments. In particular, the MFGs capture specific metabolic adaptations such as the glycolytic-gluconeogenic switch, overflow metabolism, and the effects of anoxia. We note that although we have illustrated our analysis on the core metabolism of *E. coli*, the proposed graph construction can be readily applied to large genome-scale metabolic networks [12, 19, 21, 22, 38].

To illustrate the scalability of our analyses to larger metabolic models, we studied a genome-scale model of a large metabolic model of human hepatocytes with around 3000 reactions in which we compared the wild type and a mutated state associated with the disease PH1 under more than 400 metabolic conditions [34]. Our network analysis of the corresponding MFGs revealed a consistent organisation of the reactions, which is highly preserved under the mutation, but also indentified notable changes in the network centrality and community structure of certain reactions that could be linked to key biological processes associated with PH1. Importantly, the network measures computed from the MFGs reveal complementary information to that provided by the analysis of perturbed fluxes predicted by FBA.

Our flux graphs provide a systematic connection between network theory and constraint-based methods widely employed in metabolic modelling [21, 22, 25, 32], thus opening avenues towards environment-dependent, graph-based analyses of cell metabolism. An area of interest would be to use MFGs to study how the community structure of flux graphs across scales can help characterise metabolic conditions that maximise the efficacy of drug treatments or disease-related distortions, e.g., cancer-related metabolic signatures [55–58]. In particular, MFGs can quantify metabolic robustness via graph statistics upon removal of reaction nodes [22].

The proposed graph construction framework can be extended in different directions. The core idea behind our framework is the distinction between production and consumption fluxes, and how to encode both in the links of a graph. This general principle can also be used to build other potentially useful graphs. For example, two other graphs that describe relationships between reactions are:

$$\text{Competition flux graph:} \qquad \boldsymbol{\mathcal{D}}_{\text{c}} = \frac{1}{n}\, \mathbf{S}_{2m}^{-T} \left(\mathbf{W}_{-}^{\dagger}\right)^2 \mathbf{S}_{2m}^{-} \tag{12}$$

$$\text{Synergy flux graph:} \qquad \boldsymbol{\mathcal{D}}_{\text{s}} = \frac{1}{n}\, \mathbf{S}_{2m}^{+T} \left(\mathbf{W}_{+}^{\dagger}\right)^2 \mathbf{S}_{2m}^{+}. \tag{13}$$

The competition and synergy graphs are undirected and their edge weights represent the probability that two reac-

tions consume ($\boldsymbol{\mathcal{D}}_{\mathrm{c}}$) or produce ($\boldsymbol{\mathcal{D}}_{\mathrm{s}}$) metabolites picked at random. There exist corresponding FBA versions of the competition and synergy flux graphs, which follow from (11), and the definitions (12) and (13). These graphs could help reveal further relationships between metabolic reactions in the cell and will be the subject of future studies.

Our approach could also be extended to include dynamic adaptations of metabolic activity, e.g., by using dynamic extensions of FBA [59–61], by incorporating static [62] and time-varying [63] enzyme concentrations, or by considering full kinetic models to generate reaction fluxes. Of particular interest to metabolic modelling, we envision that MFGs could provide a novel route to evaluate the robustness of FBA solutions [25, 64] by exploiting the non-uniqueness of the MFG from each FBA solution in the space of graphs. Such results could enhance the interface between network science and metabolic analysis, allowing for the systematic exploration of the system-level organisation of metabolism in response to environmental constraints and disease states.

Appendix A: Methods

1. Flux balance analysis

Flux Balance Analysis (FBA) [25, 32] is a widely-adopted approach to analyse metabolism and cellular growth. FBA calculates the reaction fluxes that optimise growth in specific biological contexts. The main hypothesis behind FBA is that cells adapt their metabolism to maximise growth in different biological conditions. The conditions are encoded as constraints on the fluxes of certain reactions; for example, constraints reactions that import nutrients and other necessary compounds from the exterior.

The mathematical formulation of the FBA is described in the following constrained optimisation problem:

$$\begin{aligned} \text{maximise:} \quad & \mathbf{c}^T\mathbf{v} \\ \text{subject to} \ & \left\{ \begin{array}{c} \mathbf{S}\mathbf{v} = 0 \\ \mathbf{v}_{\mathrm{lb}} \leq \mathbf{v} \leq \mathbf{v}_{\mathrm{ub}}, \end{array} \right. \end{aligned} \tag{A1}$$

where $\mathbf{S}$ is the stoichiometry matrix of the model, $\mathbf{v}$ the vector of fluxes, $\mathbf{c}$ is an indicator vector (i.e., $c(i) = 1$ when i is the biomass reaction and zero everywhere else) so that $\mathbf{c}^T\mathbf{v}$ is the flux of the biomass reaction. The constraint $\mathbf{S}\mathbf{v} = 0$ enforces mass-conservation at stationarity, and $\mathbf{v}_{\mathrm{lb}}$ and $\mathbf{v}_{\mathrm{ub}}$ are the lower and upper bounds of each reaction's flux. Through these vectors, one can encode a variety of different scenarios [33]. The biomass reaction represents the most widely-used flux that is optimised, although there are others can be used as well [31, 65].

In our simulations, we set the individual carbon intake rate to 18.5 mmol/gDW/h for every source available in each scenario. We allowed oxygen intake to reach the maximum needed in to consume all the carbon except in the anaerobic condition scenario, in which the upper bound for oxygen intake was 0 mmol/gDW/h. In the scenario with limited phosphate and ammonium intake, the levels of NH_4 and phosphate intake were fixed at 4.5 mmol/gDW/h and 3.04 mmol/gDW/h respectively (a reduction of 50% compared to a glucose-fed aerobic scenario with no restrictions).

2. Markov Stability community detection framework

We extract the communities in each network using the Markov Stability community detection framework [48, 49]. This framework uses diffusion processes on the network to find groups of nodes (i.e., communities) that retain flows for longer than one would expect on a comparable random network; in addition, Markov Stability incorporates directed flows seamlessly into the analysis [47, 50].

The diffusion process we use is a continuous-time Markov process on the network. From the adjacency matrix $\mathbf{G}$ of the graph (in our case, the RAG, PFG or MFG), we construct a rate matrix for the process: $\mathbf{M} = \mathbf{K}_{\mathrm{out}}^{-1}\mathbf{G}$, where $\mathbf{K}_{\mathrm{out}}$ is the diagonal matrix of out-strengths, $k_{\mathrm{out},i} = \sum_j g_{i,j}$. When a node has no outgoing edges then we simply let $k_{\mathrm{out},i} = 1$. In general, a directed network will not be strongly-connected and thus a Markov process on $\mathbf{M}$ will not have a unique steady state. To ensure the uniqueness of the steady state we must add a *teleportation* component to the dynamics by which a random walker visiting a node can follow an outgoing edge with probability λ or jump (teleport) uniformly to any other node in the network with probability $1 - \lambda$ [41]. The rate matrix of a Markov process with teleportation is:

$$\mathbf{B} = \lambda\mathbf{M} + \frac{1}{N}\left[(1 - \lambda)\mathbf{I}_N + \lambda\,\mathrm{diag}(\mathbf{a})\right]\mathbf{1}\mathbf{1}^T, \tag{A2}$$

where the $N \times 1$ vector $\mathbf{a}$ is an indicator for dangling nodes: if node i has no outgoing edges then $a_i = 1$, and $a_i = 0$ otherwise. Here we use $\lambda = 0.85$. The Markov process is described by the ODE:

$$\dot{\mathbf{x}} = -\mathbf{L}^T \mathbf{x}, \tag{A3}$$

where $\mathbf{L} = \mathbf{I}_N - \mathbf{B}$. The solution of (A3) is $\mathbf{x}(t) = e^{-t\mathbf{L}^T}\mathbf{x}(0)$ and its stationary state (i.e., $\dot{\mathbf{x}} = 0$) is $\mathbf{x} = \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the leading left eigenvector of $\mathbf{B}$.

A hard partition of the graph into C communities can be encoded into the $N \times C$ matrix $\mathbf{H}$, where $h_{ic} = 1$ if node i belongs to community c and zero otherwise. The $C \times C$ *clustered autocovariance matrix* of (A3) is

$$\mathbf{R}(t, \mathbf{H}) = \mathbf{H}^T \left(\boldsymbol{\Pi} e^{-t\mathbf{L}^T} - \boldsymbol{\pi}\boldsymbol{\pi}^T \right) \mathbf{H}, \tag{A4}$$

and the entry (c, s) of $\mathbf{R}(t, \mathbf{H})$ measures how likely it is that a random walker that started the process in community c finds itself in community s after time t when at stationarity. The diagonal elements of $\mathbf{R}(t, \mathbf{H})$ thus record how good the communities in $\mathbf{H}$ are at retaining flows. The *Markov stability of the partition* is then defined as

$$r(t, \mathbf{H}) = \text{trace}\, \mathbf{R}(t, \mathbf{H}). \tag{A5}$$

The optimised communities are obtained by maximising the cost function (A5) over the space of all partitions for every time t to obtain an optimised partition $\widehat{\mathcal{P}}(t)$. This optimisation is NP-hard; hence with no guarantees of optimality. Here we use the Louvain greedy optimisation heuristic [66], which is known to give high quality solutions $\widehat{\mathcal{P}}(t)$ in an efficient manner. The value of the Markov time t, i.e. the duration of the Markov process, can be understood as a resolution parameter for the partition into communities [46, 48]. In the limit $t \to 0$, Markov stability will assign each node to its own community; as t grows, we obtain larger communities because the random walkers have more time to explore the network [49]. We scan through a range of values of t to explore the multiscale community structure of the network. The code for Markov Stability can be found at `github.com/michaelschaub/PartitionStability`.

To identify the important partitions across time, we use two criteria of robustness [46]. Firstly, we optimise (A5) 100 times for each value of t and we assess the consistency of the solutions found. A relevant partition should be a robust outcome of the optimisation, i.e., the ensemble of optimised solutions should be similar as measured with the normalised variation of information [67]:

$$VI(\mathcal{P}, \mathcal{P}') = \frac{2\Omega(\mathcal{P}, \mathcal{P}') - \Omega(\mathcal{P}) - \Omega(\mathcal{P}')}{\log(n)}, \tag{A6}$$

where $\Omega(\mathcal{P}) = -\sum_{\mathcal{C}} p(\mathcal{C}) \log p(\mathcal{C})$ is a Shannon entropy and $p(\mathcal{C})$ is the relative frequency of finding a node in community $\mathcal{C}$ in partition $\mathcal{P}$. We then compute the average variation of information of the ensemble of solutions from the $\ell = 100$ Louvain optimisations $\mathcal{P}_i(t)$ at each Markov time t:

$$VI(t) = \frac{1}{\ell(\ell - 1)} \sum_{i \neq j} VI(\mathcal{P}_i(t), \mathcal{P}_j(t)). \tag{A7}$$

If all Louvain runs return similar partitions, then $VI(t)$ is small, indicating robustness of the partition to the optimisation. Hence we select partitions with low values (or dips) of $VI(t)$. Secondly, relevant partitions should also be optimal across Markov time, as indicated by a low values of the cross-time variation of information:

$$VI(t, t') = VI(\widehat{\mathcal{P}}(t), \widehat{\mathcal{P}}(t')). \tag{A8}$$

Therefore, we also search for partitions with extended low value plateaux of $VI(t, t')$ [46, 47, 53].

Supplementary Information

Appendix S1: Relation of the PFG with a directed version of the RAG

A directed version of the RAG (3) could in principle be obtained from the boolean production/consumption matrices $\widehat{\mathbf{S}}_{2m}^{+}$ and $\widehat{\mathbf{S}}_{2m}^{-}$ as follows. Projecting onto the space of reactions gives the $2m \times 2m$ (asymmetric) adjacency matrix

$$\mathbf{D} = \widehat{\mathbf{S}}_{2m}^{+\,T}\,\widehat{\mathbf{S}}_{2m}^{-}, \tag{S1}$$

where the entries D_{ij} represent the total number of metabolites *produced* by reaction R_i that are *consumed* by reaction R_j. A directed version of the Reaction Adjacency Graph on m nodes (directly comparable to the standard RAG) is then

$$\mathbf{A}_{\mathrm{dir}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{I}_m \end{bmatrix} \mathbf{D} \begin{bmatrix} \mathbf{I}_m \\ \mathbf{I}_m \end{bmatrix}. \tag{S2}$$

Clearly, when the metabolic model contains only reversible reactions, (i.e., the reversibility vector is all ones, $\mathbf{r} = \mathbb{1}_m$), it follows that $\mathbf{A}_{\mathrm{dir}} = \mathbf{A}$.

Although $\mathbf{A}_{\mathrm{dir}}$ does not include spurious edges introduced by non-existent backward reactions, its structure is still obscured by the effect of uninformative connections created by pool metabolites.

Appendix S2: Details of the toy metabolic network

As an illustration of the graph construction, the toy metabolic network in Fig. 1 was taken from Ref. [32]. The graph matrices for this model are as follows:

- Reaction Adjacency Graph, Eq. (3):

$$\mathbf{A} = \widehat{\mathbf{S}}^{T}\widehat{\mathbf{S}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 2 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 2 & 1 & 3 \end{bmatrix}.$$

- Probabilistic Flux Graph, Eq. (7):

$$\mathcal{D} = \frac{1}{n}\,\mathbf{S}_{2m}^{+\,T}\left(\mathbf{W}_{+}^{\dagger}\mathbf{W}_{-}^{\dagger}\right)\mathbf{S}_{2m}^{-} = \begin{array}{c|ccccccccc} & R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_7 & R_8 & R_{4r} \\ R_1 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ R_2 & 0 & 0 & 0.05 & 0.05 & 0 & 0 & 0 & 0 & 0 \\ R_3 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0.1 & 0 \\ R_4 & 0 & 0 & 0 & 0 & 0 & 0.04 & 0.04 & 0.08 & 0.04 \\ R_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0 \\ R_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0 \\ R_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ R_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ R_{4r} & 0 & 0 & 0.05 & 0.05 & 0 & 0 & 0 & 0 & 0 \end{array}.$$

- Metabolic Flux Graph for FBA scenario 1, Eq. (11):

$$\begin{array}{c|ccc} & \mathbf{v}_{\mathrm{lb}_1} & \mathbf{v}_{\mathrm{ub}_1} & \mathbf{v}_1^{*} \\ R_1: & 10 & 10 & 10 \\ R_2: & 0 & 10 & 10 \\ R_3: & 0 & 10 & 4.992 \\ R_4: & -10 & 10 & 5.008 \\ R_5: & 0 & 10 & 2.492 \\ R_6: & 0 & 10 & 0.008 \\ R_7: & 0 & 10 & 0 \\ R_8: & 0 & 10 & 2.5 \\ R_{4r}: & -10 & 10 & 0 \end{array} \qquad \mathbf{M}(\mathbf{v}_1^{*}) = \begin{array}{c|ccccccc} & R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_8 \\ R_1 & 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ R_2 & 0 & 0 & 4.992 & 5.008 & 0 & 0 & 0 \\ R_3 & 0 & 0 & 0 & 0 & 2.492 & 0 & 2.5 \\ R_4 & 0 & 0 & 0 & 0 & 0 & 0.008 & 5 \\ R_5 & 0 & 0 & 0 & 0 & 0 & 0 & 2.492 \\ R_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0.008 \\ R_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}.$$

- Metabolic Flux Graph for FBA scenario 2, , Eq. (11):

$$
\begin{array}{c}
\begin{array}{cccc}
 & \mathbf{v}_{\mathrm{lb}_2} & \mathbf{v}_{\mathrm{ub}_2} & \mathbf{v}_2^* \\
R_1: & 10 & 10 & 10 \\
R_2: & 0 & 10 & 10 \\
R_3: & 0 & 10 & 3.877 \\
R_4: & -10 & 10 & 6.123 \\
R_5: & 0 & 10 & 1.877 \\
R_6: & 0 & 10 & 0.123 \\
R_7: & 2 & 10 & 2 \\
R_8: & 0 & 10 & 2 \\
R_{4r}: & -10 & 10 & 0 \\
\end{array}
\qquad
\mathbf{M}(\mathbf{v}_2^*) =
\begin{array}{c}
\\
R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8
\end{array}
\left[
\begin{array}{cccccccc}
R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_7 & R_8 \\
0 & 10 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3.877 & 6.123 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1.877 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0.123 & 2 & 4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.877 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.123 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\right].
\end{array}
$$

Appendix S3: Reaction communities in context-free graphs of the core *E. coli* metabolic model

1. Reaction Adjacency Graph, A

A robust partition into seven communities in the RAG was found at Markov time $t = 6.01$ (Fig. S1A). The communities at this resolution (Fig. 3E) are:

- Community C1($\mathbf{A}$) contains all the reactions that consume or produce ATP and water (two pool metabolites). Production of ATP comes mostly from oxidative phosphorylation (ATPS4r) and substrate level phosphorylation reactions such as phosphofructokinase (PFK), phosphoglicerate kinase (PGK) and succinil-CoA synthase (SU-COAS). Reactions that consume ATP include glutamine synthetase (GLNS) and ATP maintenance equivalent reaction (ATPM). The reactions L-glutamine transport via ABC system (GLNabc), acetate transport in the form of phosphotransacetilase (PTAr), and acetate kinase (ACKr) are also part of this community. Additionally, C1($\mathbf{A}$) (green) contains also reactions that involve H_2O. Under normal conditions water is assumed to be abundant in the cell, thus the biological link that groups these reactions together is tenuous.

- Community C2($\mathbf{A}$) includes the reactions NADH dehydrogenase (NADH16), cytochrome oxidase (CYTBD), and transport and exchange reactions. These two reactions involve pool metabolites (such as H^+) which create a large number of connection. Other members include fumarate reductase (FR7) and succinate dehydrogenase (SUCDi) which couple the TCA cycle with the electron transport chain (through ubiquinone-8 reduction and ubiquinol-8 oxidation). Reactions that include export and transport of most secondary carbon sources (such as pyruvate, ethanol, lactate, acetate, malate, fumarate, succinate or glutamate) are included in the community as well. These reactions are included in the community because of their influence in the proton balance of the cell. Most of these reactions do not occur under normal circumstances. This community highlights the fact that in the absence of biological context, many reactions that do not normally interact can be grouped together.

- Community C3($\mathbf{A}$) contains reactions that produce or consume nicotinamide adenine dinucleotide (NAD^+), nicotinamide adenine dinucleotide phosphate ($NADP^+$), or their reduced variants NADH and NADPH. The main two reactions of the community are NAD(P) transhydrogenase (THD2) and NAD^+ transhydrogenase (NADTRHD). There are also reactions related to the production of NADH or NADPH in the TCA cycle such as isocitrate dehydrogenase (ICDHyr), 2-oxoglutarate dehydrogenase (AKGDH) and malate dehydrogenase (MDH). The community also includes reactions that are not frequently active such as malic enzyme NAD (ME1) and malic enzyme NADH (ME2) or acetate dehydrogenase (ACALD) and ethanol dehydrogenase (ALCD2x).

- Community C4($\mathbf{A}$) contains the main carbon intake of the cell (glucose), the initial steps of glycolysis, and most of the pentose phosphate shunt. These reactions are found in this community because the metabolites involved in these reactions (e.g., alpha-D-ribose-5-phosphate (r5p) or D-erythrose-4-phosphate (e4p)) are only found in these reactions. This community includes the biomass reaction due to the number of connections created by growth precursors.

- Communities C5($\mathbf{A}$), C6($\mathbf{A}$) and C7($\mathbf{A}$) are small communities that contain oxygen intake, ammonium intake and acetaldehyde secretion reactions, respectively.
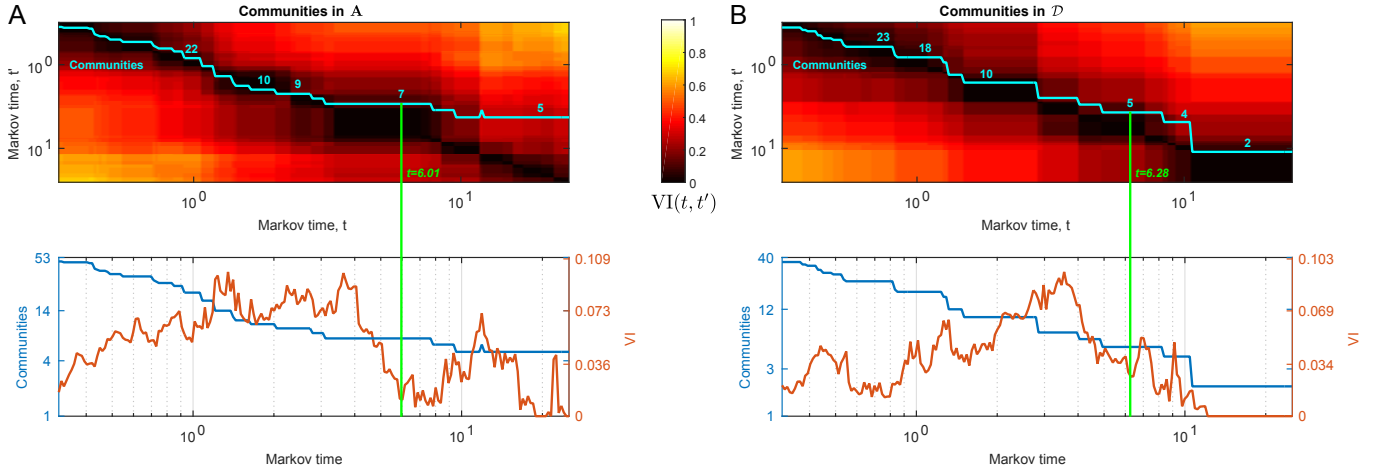
FIG. S1. Community structure in the template networks A and $\mathcal{D}$. (A) Communities in **A**. Top plot: Variation of Information (VI) of the best partition found at Markov time t with every other partition at time t'. Bottom plot: Number of communities and VI of the ensemble of solutions found at each Markov time. A robust partition into seven communities is found at $t = 6.01$. (B) Communities and VI in $\mathcal{D}$. A robust partition into five communities is found at $t = 6.28$.

2. Probabilistic Flux Graph, $\mathcal{D}$

A robust partition into five communities in the PFG was found at Markov time $t = 6.28$ (Fig. S1B). The communities at this resolution (Fig. 3C) are:

- Community C1($\mathcal{D}$) includes the first half of the glycolysis and the complete pentose phosphate pathway. The metabolites that create the connections among these reactions such as D-fructose, D-glucose, or D-ribulose.

- Community C2($\mathcal{D}$) contains the main reaction that produces ATP through substrate level (PGK, PYK, ACKr) and oxidative phosphorylation (ATPS4r). The flow of metabolites among the reactions in this community includes some pool metabolites such as ATP, ADP, H_2O, and phosphate. However, there are connections created by metabolites that only appear in a handful of reactions such as adenosine monophosphate (AMP) whose sole producer is phosphoenolpyruvate synthase (PPS) and its sole consumer is ATPS4r. This community also contains the biomass reaction.

- Community C3($\mathcal{D}$) includes the core of the citric acid (TCA) cycle such as citrate synthase (CS), aconitase A/B (ACONTa/b), and anaplerotic reactions such as malate synthase (MALS), malic enzyme NAD (ME1), and malic enzyme NADP (ME2). This community also includes the intake of cofactors such as CO_2.

- Community C4($\mathcal{D}$) contains reactions that are secondary sources of carbon such as malate and succinate, as well as oxidative phosphorilation reactions.

- Community C5($\mathcal{D}$) contains some reactions part of the pyruvate metabolism subsystem such as D-lactate dehydrogenase (LDH-D), pyruvate formate lyase (PFL) or acetaldehyde dehydrogenase (ACALD). In addition, it also includes the tranport reaction for the most common secondary carbon metabolites such as lactate, formate, acetaldehyde and ethanol.

Appendix S4: Reaction communities in Metabolic Flux Graphs of *E. coli* metabolism under different biological scenarios

1. M_{glc}: aerobic growth under glucose

This graph has 48 reactions with nonzero flux and 227 edges. At Markov time $t = 7.66$ (Fig. S2A) this graph has a partition into three communities (Fig. 4A):

- Community C1($\mathbf{M}_{\text{glc}}$) comprises the intake of glucose and most of the glycolysis and pentose phosphate pathway. The function of the reactions in this community consists of carbon intake and processing glucose into phospho-enolpyruvate (PEP). This community produces essential biocomponents for the cell such as alpha-D-Ribose 5-phosphate (rp5), D-Erythrose 4-phosphate (e4p), D-fructose-6-phosphate (f6p), glyceraldehyde-3-phosphate (g3p) or 3-phospho-D-glycerate (3pg). Other reactions produce energy ATP and have reductive capabilities for catabolism.

- Community C2($\mathbf{M}_{\text{glc}}$) contains the electron transport chain which produces the majority of the energy of the cell. In the core *E coli* metabolic model the chain is represented by the reactions NADH dehydrogenase (NADH16), cytochrome oxidase BD (CYTBD) and ATP synthase (ATPS4r). This community also contains associated reactions to the electron transport such as phosphate intake (EXpi(e), PIt2), oxygen intake (EXo2(e), O2t) and proton balance (EXh(e)). This community also includes the two reactions that represent energy maintenance costs (ATPM), and growth (biomass); this is consistent with the biological scenario because ATP is the main substrate for both ATPM, and the biomass reaction.

- Community C3($\mathbf{M}_{\text{glc}}$) contains the TCA cycle at its core. The reactions in this community convert PEP into ATP, NADH and NADPH. In contrast with C1($\mathbf{M}_{\text{glc}}$), there is no precursor formation here. Beyond the TCA cycle, pyruvate kinase (PYK), phosphoenolpyruvate carboxylase (PPC) and pyruvate dehydrogenase (PDH) appear in this community. These reactions highlight the two main carbon intake routes in the cycle: oxalacetate from PEP through phosphoenol pyruvate carboxylase (PPC), and citrate from acetyl coenzyme A (acetyl-CoA) via citrate synthase (CS). Furthermore, both routes begin with PEP, so it is natural for them to belong to the same community along with the rest of the TCA cycle. Likewise, the production of L-glutamate from 2-oxoglutarate (AKG) by glutamate dehydrogenase (GLUDy) is strongly coupled to the TCA cycle.

2. $\mathbf{M}_{\text{etoh}}$: aerobic growth under ethanol

This graph contains 49 reactions and 226 edges. At Markov time $t = 6.28$ (Fig. S2B) this graph has a partition into three communities (Fig. 4B):

- Community C1($\mathbf{M}_{\text{etoh}}$) in this graph is similar to its counterpart in $\mathbf{M}_{\text{glc}}$, but with important differences. For example, the reactions in charge of the glucose intake (EXglc(e) and GLCpts) are no longer part of the network (i.e., they have zero flux), and reactions such as malic enzyme NAPD (ME2) and phosphoenolpyruvate caboxykinase (PPCK), which now appear in the network, belong to this community. This change in the network reflect the cell's response to a new biological situation. The carbon intake through ethanol has changed the direction of glycolysis into gluconeogenesis [1] (the reactions in C1($\mathbf{M}_{\text{glc}}$) in Fig. 4A are now operating in the reverse direction in Fig. 4B). The main role of the reactions in this community is the production of bioprecursors such as PEP, pyruvate, 3-phospho-D-glycerate (3PG) glyceraldehyde-3-phosphate (G3P), D-fructose-6-phosphate (F6P), and D-glucose-6-phosphate, all of which are substrates for growth. Reactions ME2 and PPCK also belong to this community due to their production of PYR and PEP. Reactions that were in a different community in $\mathbf{M}_{\text{glc}}$, such as GLUDy and ICDHyr which produce precursors L-glutamate and NADPH respectively, are now part of C1($\mathbf{M}_{\text{etoh}}$). This community also includes the reactions that produce inorganic substrates of growth such as NH_4, CO_2 and H_2O.

- Community C2($\mathbf{M}_{\text{etoh}}$) contains the electron transport chain and the bulk of ATP production, which is similar to C2($\mathbf{M}_{\text{glc}}$). However, there are subtle differences that reflect changes in this new scenario. Ethanol intake and transport reactions (EXetoh(e) and ETOHt2r) appear in this community due to their influence in the proton balance of the cell. In addition, C2($\mathbf{M}_{\text{etoh}}$) contains NADP transhydrogenase (THD2) which is in charge of NADH/NADPH balance. This reaction is present here due to the NAD consumption involved in the reactions ACALD and ethanol dehydrogenase (ALCD2x), which belong to this community as well.

- Community C3($\mathbf{M}_{\text{etoh}}$) contains most of the TCA cycle. The main difference between this community and C1($\mathbf{M}_{\text{glc}}$) is that here acetyl-CoA is extracted from acetaldehyde (which comes from ethanol) by the reaction acetaldehyde dehydrogenase reaction (ACALD), instead of the classical pyruvate from glycolysis. The glycoxylate cycle reactions isocitrate lyase (ICL) and malate synthase (MALS) which now appear in the network, also belong to this community. These reactions are tightly linked to the TCA cycle and appear when the carbon intake is acetate or ethanol to prevent the loss of carbon as CO_2.

3. M$_{\text{anaero}}$: anaerobic growth

This graph contains 47 reactions and 212 edges. At Markov time $t = 6.01$ (Fig. S2C) this graph has a partition into four communities (Fig. 4C):

- Community C1($\mathbf{M}_{\text{anaero}}$) contains the reactions responsible D-glucose intake (EXglc) and most of the glycolysis. The reaction that represents the cellular maintenance energy cost, ATP maintenance requirement (ATPM), is included in this community because of the increased strength of its connection to the substrate-level phosphorilation reaction phosphoglycerate kinase (PGK). Also note that reactions in the pentose phosphate pathway do not belong to the same community as the glycolysis reactions (unlike in $\mathbf{M}_{\text{glc}}$ and $\mathbf{M}_{\text{etoh}}$).

- Community C2($\mathbf{M}_{\text{anaero}}$) contains the conversion of PEP into formate through the sequence of reactions PYK, PFL, FORti and EXfor(e). More than half of the carbon secreted by the cell becomes formate.

- Community C3($\mathbf{M}_{\text{anaero}}$) includes the biomass reaction and the reactions in charge of supplying it with substrates. These reactions include the pentose phosphate pathway (now detached from C1($\mathbf{M}_{\text{glc}}$)), which produce essential growth precursors such as alpha-D-ribose-5-phosphate (r5p) or D-erythrose-4-phosphate (e4p). The TCA cycle is present as well because its production of two growth precursors: 2-oxalacetate and NADPH. Finally, the reactions in charge of acetate production (ACKr, ACt2r and EXac(e)) are also members of this community through the ability of ACKr to produce ATP. Glutamate metabolism reaction GLUDy is also included in this community. It is worth mentioning that the reverse of ATP synthase (ATPS4r) is present in this community because here, unlike in $\mathbf{M}_{\text{glc}}$, ATPS4r consumes ATP instead of producing it. When this flux is reversed, then ATPS4r is in part responsible for pH homeostasis.

- Community C4($\mathbf{M}_{\text{anaero}}$) includes the main reactions involved in NADH production and consumption, which occurs via glyceraldehyde-3-phosphate dehydrogenase (GAPD). NADH consumption occurs in two consecutive steps in ethanol production: in ACALD and ALCD2x. The phosphate intake and transport reactions EXpi(e) and PIt2r belong to this community because most of the phosphate consumption takes place at GAPD. Interestingly, the core reaction around which the community forms (GAPD) is not present in the community. It is included in earlier Markov times but when communities start to get larger the role of GAPD becomes more relevant as a part of the glycolysis than its role as a NADH hub. This is a good example of how the graph structure and the clustering method are able to capture two different roles in the same metabolite.

4. M$_{\text{lim}}$: aerobic growth under limiting conditions

This graph has 52 nodes and 228 edges. At Markov time $t = 13$ this graph (Fig. S2D) has a partition into three communities (Fig. 4D):

- Community C1($\mathbf{M}_{\text{lim}}$) contains the glycolysis pathway (detached from the pentose phosphate pathway). This community is involved in precursor formation, ATP production, substrate-level phosphorylation and processing of D-glucose into PEP.

- Community C2($\mathbf{M}_{\text{lim}}$) contains the bioenergetic machinery of the cell; the main difference to the previous scenarios is that the electron transport chain has a smaller role in ATP production (ATPS4r), and substrate-level phosphorylation (PGK, PYK, SUCOAS, ACKr) becomes more important. In $\mathbf{M}_{\text{lim}}$ the electron transport chain is responsible for the 21.8% of the total ATP produced in the cell while in $\mathbf{M}_{\text{glc}}$ it produces 66.5%. The reactions in charge of intake and transport of inorganic ions such as phosphate (EXpi(e) and PIt2r), O$_2$ (EXO$_2$(e) and O$_2$t)and H$_2$O (EXH$_2$O and H$_2$Ot) belong to this community as well. This community includes the reactions in the pentose phosphate pathway that produce precursors for growth: transketolase (TKT2) produces e4p, and ribose-5-phosphate isomerase (RPI) produces r5p.

- Community C3($\mathbf{M}_{\text{lim}}$) is the community that differs the most from those in the other aerobic growth networks ($\mathbf{M}_{\text{glc}}$ and $\mathbf{M}_{\text{etoh}}$). This community gathers reactions that under normal circumstances would not be so strongly related but that the limited availability of ammonium and phosphate have forced together; its members include reactions from the TCA cycle, the pentose phosphate pathway, nitrogen metabolism and by-product secretion. The core feature of the community is carbon secretion as formate and acetate. Reactions PPC, malate dehydrogenase (MDH) reverse and ME2 channel most of the carbon to the secretion routes in the form of formate and acetate. The production of L-glutamine seems to be attached to this subsystem through the production of NADPH in ME2 and its consumption in the glutamate dehydrogenase NAPD (GLUDy).
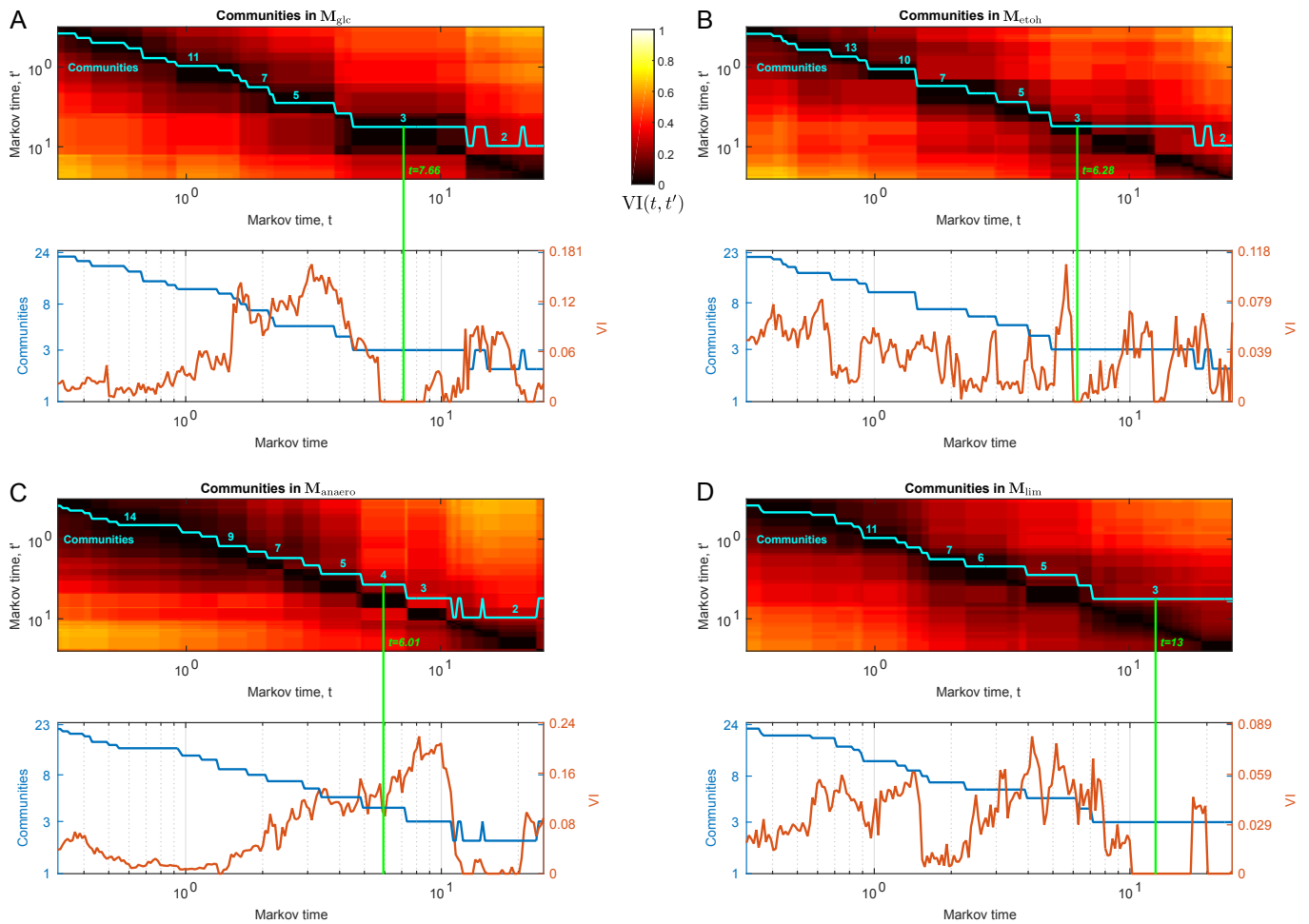
FIG. S2. **Community structure in the MFGs.** Number of communities and VI of the MFGs in four biological scenarios. (A) The graph $\mathbf{M}_{\mathrm{glc}}$ has a robust partition into three communities at $t = 7.66$. (B) $\mathbf{M}_{\mathrm{etoh}}$ has a partition into three communities at $t = 6.28$. (C) $\mathbf{M}_{\mathrm{anaero}}$ has four communities at $t = 6.01$. (D) $\mathbf{M}_{\mathrm{lim}}$ has three communities at $t = 13.0$.

[1] Berg JM, Tymoczko JL, Stryer L. Biochemistry, Fifth Edition. W. H. Freeman; 2002.
[2] Thomas A, Cannings R, Monk NAM, Cannings C. On the structure of protein–protein interaction networks. Biochemical Society Transactions. 2003;31(6):1491–1496.
[3] Alon U. Network motifs: theory and experimental approaches. Nature reviews Genetics. 2007 jun;8(6):450–61. Available from: http://dx.doi.org/10.1038/nrg2102.
[4] Amor B, Yaliraki SN, Woscholski R, Barahona M. Uncovering allosteric pathways in caspase-1 using Markov transient analysis and multiscale community detection. Molecular bioSystems. 2014 aug;10(8):2247–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24947802.
[5] Amor BR, Schaub MT, Yaliraki SN, Barahona M. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. Nature Communications. 2016;7:12477.
[6] Newman M. Networks: An Introduction. New York, NY, USA: Oxford University Press, Inc.; 2010.
[7] Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. Nature. 2000 oct;407(6804):651–4. Available from: http://dx.doi.org/10.1038/35036627.
[8] Wagner A, Fell DA. The small world inside large metabolic networks. Proc R Soc Lond B. 2001 Sep;268(1478):1803–1810. Available from: http://dx.doi.org/10.1098/rspb.2001.1711.
[9] Gleiss PM, Stadler PF, Wagner A, Fell DA. Relevant cycles in chemical reaction networks. Advances in Complex Systems. 2001;04(02n03):207–226. Available from: http://www.worldscientific.com/doi/abs/10.1142/S0219525901000140.
[10] Arita M. The metabolic world of Escherichia coli is not small. Proceedings of the National Academy of Sciences of the

United States of America. 2004 feb;101(6):1543–7. Available from: `http://www.pnas.org/content/101/6/1543.long`.

[11] Guimerá R, Nunes Amaral LA. Functional cartography of complex metabolic networks. Nature. 2005;433(7028):895–900. Available from: `http://www.nature.com/nature/journal/v433/n7028/full/nature03288.html`.

[12] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science (New York, NY). 2002 aug;297(5586):1551–5. Available from: `http://science.sciencemag.org/content/297/5586/1551.abstract`.

[13] Takemoto K. Does habitat variability really promote metabolic network modularity? PloS one. 2013 jan;8(4):e61348. Available from: `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0061348`.

[14] Zhou W, Nakhleh L. Convergent evolution of modularity in metabolic networks through different community structures. BMC evolutionary biology. 2012 jan;12(1):181. Available from: `http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-12-181`.

[15] Cooper K, Barahona M. Role-based similarity in directed networks. arXiv:10122726. 2010 Dec;Available from: `http://arxiv.org/abs/1012.2726`.

[16] Palsson BO. Systems Biology: Properties of Reconstructed Networks. New York, NY, USA: Cambridge University Press; 2006.

[17] Ouzounis CA, Karp P. Global Properties of the Metabolic Map of Escherichia coli. Genome Research. 2000 apr;10(4):568–576. Available from: `http://genome.cshlp.org/content/10/4/568.full`.

[18] Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics (Oxford, England). 2003 jul;19(11):1423–30. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/12874056`.

[19] Ma HW, Zhao XM, Yuan YJ, Zeng AP. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. Bioinformatics (Oxford, England). 2004 aug;20(12):1870–6. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/15037506`.

[20] Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome biology. 2006 jan;7(5):R39. Available from: `http://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-5-r39`.

[21] Samal A, Singh S, Giri V, Krishna S, Raghuram N, Jain S. Low degree metabolites explain essential reactions and enhance modularity in biological networks. BMC bioinformatics. 2006 jan;7:118. Available from: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1434774{\&}tool=pmcentrez{\&}rendertype=abstract`.

[22] Smart AG, Amaral LAN, Ottino JM. Cascading failure and robustness in metabolic networks. Proceedings of the National Academy of Sciences of the United States of America. 2008 sep;105(36):13223–8. Available from: `http://www.pnas.org/content/105/36/13223.full`.

[23] Winterbach W, Mieghem PV, Reinders M, Wang H, de Ridder D. Topology of molecular interaction networks. BMC Systems Biology. 2013;7(1):90. Available from: `http://www.biomedcentral.com/1752-0509/7/90/abstract`.

[24] Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, et al. Metabolic Flux Ratio Analysis of Genetic and Environmental Modulations of Escherichia coli Central Carbon Metabolism. Journal of Bacteriology. 1999;181(21):6679–6688. Available from: `http://jb.asm.org/content/181/21/6679.abstract`.

[25] Orth JD, Thiele I, Palsson B. What is flux balance analysis? Nature Biotechnology. 2010;28(3):245–248. Available from: `http://www.nature.com/doifinder/10.1038/nbt.1614`.

[26] Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics. 2003;19(2):270–277. Available from: `http://bioinformatics.oxfordjournals.org/content/19/2/270.abstract`.

[27] Croes D, Couche F, Wodak SJ, van Helden J. Inferring meaningful pathways in weighted metabolic networks. Journal of Molecular Biology. 2006;356(1):222–236.

[28] Silva MRd, Sun J, Ma H, He F, Zeng AP. Metabolic Networks. In: Junker BH, Schreiber F, editors. Analysis of Biological Networks. John Wiley & Sons, Inc.; 2008. p. 233–253.

[29] Kreimer A, Borenstein E, Gophna U, Ruppin E. The evolution of modularity in bacterial metabolic networks. Proc Natl Acad Sci U S A. 2008 May;105(19):6976–6981. Available from: `http://dx.doi.org/10.1073/pnas.0712149105`.

[30] Samal A, Martin OC. Randomizing Genome-Scale Metabolic Networks. PLoS ONE. 2011;6(7):e22295. Available from: `http://dx.doi.org/10.1371/journal.pone.0022295`.

[31] Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Molecular Systems Biology. 2007;3(1).

[32] Rabinowitz JD, Vastag L. Teaching the design principles of metabolism. Nat Chem Biol. 2012 Jun;8(6):497–501. Available from: `http://dx.doi.org/10.1038/nchembio.969`.

[33] Orth J, Fleming R, Palsson B. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. EcoSal Plus. 2010;Available from: `http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.10.2.1`.

[34] Pagliarini R, Castello R, Napolitano F, Borzone R, Annunziata P, Mandrile G, et al. In Silico Modeling of Liver Metabolism in a Human Disease Reveals a Key Enzyme for Histidine and Histamine Homeostasis. Cell Reports. 2016;15(10):2292 – 2300. Available from: `http://www.sciencedirect.com/science/article/pii/S2211124716305812`.

[35] Heinrich R, Schuster S. The Regulation of Cellular Systems. Springer US; 2012.

[36] Chellaboina V, Bhat SP, Haddad WM, Bernstein DS. Modeling and analysis of mass-action kinetics. IEEE Control Systems. 2009 Aug;29(4):60–78.

[37] King ZA, Drğer A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLoS Comput Biol. 2015 08;11(8):1–13. Available from:

http://dx.doi.org/10.1371%2Fjournal.pcbi.1004321.

[38] Folch-Fortuny A, Tortajada M, Prats-Montalbán JM, Llaneras F, Picó J, Ferrer A. MCR–ALS on metabolic networks: Obtaining more meaningful pathways. Chemometrics and Intelligent Laboratory Systems. 2015;142:293–303.

[39] Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol. 2000 Mar;18(3):326–332. Available from: http://dx.doi.org/10.1038/73786.

[40] Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol. 2000 Apr;203(3):229–248. Available from: http://dx.doi.org/10.1006/jtbi.2000.1073.

[41] Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab; 1999. 1999-66. Previous number = SIDL-WP-1999-0120. Available from: http://ilpubs.stanford.edu:8090/422/.

[42] Gleich DF. PageRank Beyond the Web. SIAM Review. 2015;57(3):321–363. Available from: http://dx.doi.org/10.1137/140976649.

[43] Sankey HR. The Thermal Efficiency of Steam-Engines. Minutes of Proceedings of The Institution of Civil Engineers. 1896;125:182–242.

[44] Rosvall M, Bergstrom CT. Mapping change in large networks. PLoS ONE. 2010;5(1):e8694. Available from: http://dx.doi.org/10.1371/journal.pone.0008694.

[45] Girvan M, Newman MEJ. Community structure in social and biological networks. Proceedings of the National Academy of Sciences. 2002;99(12):7821–7826. Available from: http://www.pnas.org/content/99/12/7821.abstract.

[46] Schaub MT, Delvenne JC, Yaliraki SN, Barahona M. Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. PLoS ONE. 2012;7(2):e32210. Available from: http://arxiv.org/abs/1109.5593.

[47] Lambiotte R, Delvenne J, Barahona M. Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. Network Science and Engineering, IEEE Transactions on. 2014 July;1(2):76–90.

[48] Delvenne JC, Yaliraki SN, Barahona M. Stability of graph communities across time scales. Proc Nat Acad Sci USA. 2010;107(29):12755–12760. Available from: http://www.pnas.org/content/107/29/12755.abstract.

[49] Delvenne JC, Schaub MT, Yaliraki SN, Barahona M. The Stability of a Graph Partition: A Dynamics-Based Framework for Community Detection. In: Mukherjee A, Choudhury M, Peruani F, Ganguly N, Mitra B, editors. Dynamics On and Of Complex Networks, Volume 2. Modeling and Simulation in Science, Engineering and Technology. Springer New York; 2013. p. 221–242.

[50] Beguerisse-Díaz M, Garduño Hernández G, Vangelov B, Yaliraki SN, Barahona M. Interest communities and flow roles in directed networks: the Twitter network of the UK riots. J R Soc Interface. 2014 Dec;11(101). Available from: http://rsif.royalsocietypublishing.org/content/11/101/20140940.

[51] Sawers RG. Formate and its role in hydrogen production in Escherichia coli. Biochemical Society Transactions. 2005;33(1):42–46. Available from: http://www.biochemsoctrans.org/content/33/1/42.

[52] Vemuri GN, Eiteman MA, McEwen JE, Olsson L, Nielsen J. Increasing NADH oxidation reduces overflow metabolism in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2007 Feb;104(7):2402–2407. Available from: http://dx.doi.org/10.1073/pnas.0607469104.

[53] Bacik KA, Schaub MT, Beguerisse-Díaz M, Billeh YN, Barahona M. Flow-Based Network Analysis of the Caenorhabditis elegans Connectome. PLoS Comput Biol. 2016 08;12(8):1–27. Available from: http://dx.doi.org/10.1371%2Fjournal.pcbi.1005055.

[54] Gille C, Bölling C, Hoppe A, Bulik S, Hoffmann S, Hübner K, et al. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. Molecular Systems Biology. 2010;6(1). Available from: http://msb.embopress.org/content/6/1/411.

[55] Csermely P, Ágoston V, Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. Trends in Pharmacological Sciences. 2005;26(4):178 – 182. Available from: http://www.sciencedirect.com/science/article/pii/S0165614705000556.

[56] Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. PLoS Comput Biol. 2010;6(9):e1000938. Available from: http://dx.doi.org/10.1371/journal.pcbi.1000938.

[57] Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. Molecular Systems Biology. 2011;7(1). Available from: http://msb.embopress.org/content/7/1/501.

[58] Vaitheesvaran B, Xu J, Yee J, Q-Y L, Go VL, Xiao GG, et al. The Warburg effect: a balance of flux analysis. Metabolomics. 2015 Aug;11(4):787–796. Available from: http://dx.doi.org/10.1007/s11306-014-0760-9.

[59] Mahadevan R, Edwards JS, Doyle FJ 3rd. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J. 2002 Sep;83(3):1331–1340. Available from: http://dx.doi.org/10.1016/S0006-3495(02)73903-9.

[60] Waldherr S, Oyarzún DA, Bockmayr A. Dynamic optimization of metabolic networks coupled with gene expression. Journal of Theoretical Biology. 2015 nov;365:469–485. Available from: http://www.sciencedirect.com/science/article/pii/S0022519314006250.

[61] Rügen M, Bockmayr A, Steuer R. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. Scientific reports. 2015 jan;5:15247. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4620596{\&}tool=pmcentrez{\&}rendertype=abstract.

[62] Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting Expression Data with Metabolic Flux

Models: Predicting *Mycobacterium tuberculosis* Mycolic Acid Production. PLoS Comput Biol. 2009 08;5(8):e1000489. Available from: `http://dx.doi.org/10.1371%2Fjournal.pcbi.1000489`.

[63] Oyarzún DA. Optimal control of metabolic networks with saturable enzyme kinetics. IET systems biology. 2011 mar;5(2):110–9. Available from: `http://ieeexplore.ieee.org/xpls/abs{\_}all.jsp?arnumber=5734998http://www.ncbi.nlm.nih.gov/pubmed/21405199`.

[64] Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC bioinformatics. 2010 jan;11(1):489. Available from: `http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-489`.

[65] Feist AM, Palsson BO. The biomass objective function. Current Opinion in Microbiology. 2010;13(3):344–349. Available from: `http://www.sciencedirect.com/science/article/pii/S1369527410000512`.

[66] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008;2008(10):P10008. Available from: `http://stacks.iop.org/1742-5468/2008/i=10/a=P10008`.

[67] Meila M. Comparing clusterings: an information based distance. Journal of Multivariate Analysis. 2007;98(5):873 – 895.