Global organization of metabolic fluxes in *E. coli*

Supplementary Material

Contents:

1. Flux balance analysis (FBA)

We implemented the FBA described in Ref. (*1*) starting from a stoichiometric matrix that captures the reconstruction of the K12 derivative MG1655 (*2*) strain of *E. coli*, containing 537 metabolites and 739 reactions. In a steady state the concentrations of all the metabolites are time independent

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} v_j = 0, \qquad (2)$$

where S_{ij} is the stoichiometric coefficient of metabolite A_i in reaction j and v_j is the flux of reaction j. We use the convention that if metabolite A_i is a substrate (product) in reaction j, $S_{ij} < 0$ ($S_{ij} > 0$), and we constrain all fluxes to be positive by dividing each reversible reaction into two "forward" reactions with positive fluxes. Any vector of positive fluxes $\{v_j\}$ which satisfies (2) corresponds to a state of the metabolic network, and hence, a potential state of operation of the cell. We restrict our study to the subspace of solutions for which all components of $\boldsymbol{v}$ satisfy the constraint $v_j > 0$ (*1*). We denote the mass carried by reaction j producing (consuming) metabolite i by $\hat{v}_{ij} = |S_{ij}| v_j$, where S_{ij} is the stoichiometric coefficient of reaction j. An important step in the establishment of the stoichiometric matrix $\boldsymbol{S}$ is to ensure mass conservation, i.e., that all the internal metabolites (metabolites which are not transported through the cell membrane) appear at least once as both a substrate and a product in the reaction system (*1*).

2. Optimal states

2.1 Linear optimization: Using linear programming and adapting constraints for each reaction flux v_i of the form $\beta_i^{\min} \le v_i \le \beta_i^{\max}$, we calculate the flux states optimizing cell growth on various substrates. These constraints can also be used to control the reversibility of the reactions, an irreversible reaction having $\beta_i^{min} = 0$. We used the linear programming code lp_solve (ftp://ftp.ics.ele.tue.nl/pub/lp_solve/) to find an optimized

flux vector for a given set of constraints. The stoichiometric matrix, components of the biomass vector and the choices for the bounds β_i^{min} and β_i^{max} were taken from Segre et al. (*3*). During optimization, we set the minimal uptake basis to have unlimited access to carbon dioxide, potassium, sulfate and phosphate, and limited access to ammonia (maximal uptake rate of 100 mmol/g DW/h) and oxygen (maximal uptake rate of 20 mmol/g DW/h). When we simulate the utilization of additional carbon sources, like glutamate, succinate or glucose, we limit their maximal uptake rate to 20 mmol/g DW/h. In Fig. S1 (a), we compare the flux distributions thus calculated for succinate (black), glutamate (red) and glucose (green) rich conditions, and to Luria-Bertani medium (LB) (blue). The solid line is the best fit from Fig. 1a.

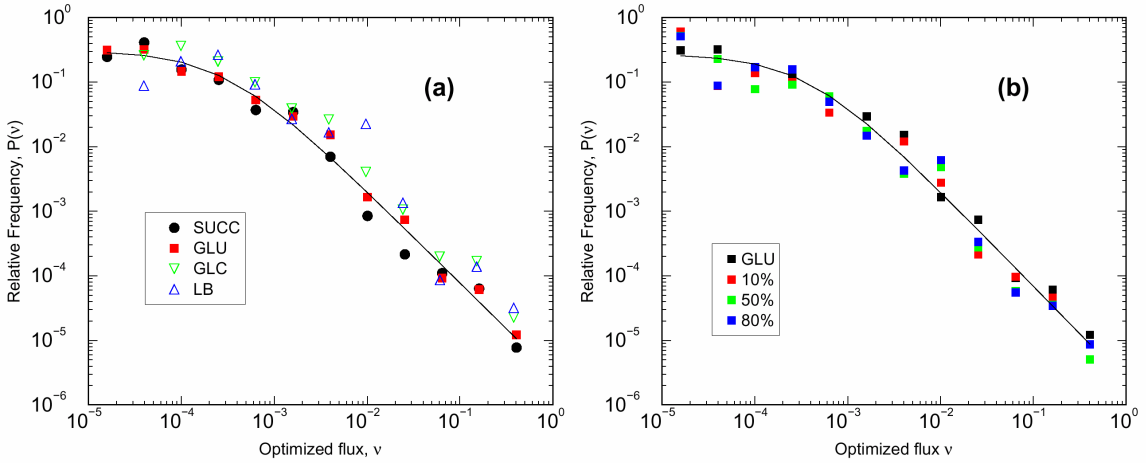

Figure S1. **(a)** Flux distribution for optimization of biomass on succinate (black), glutamate (red), glucose rich media (green) and Luria-Bertani medium (LB) (blue). The solid line is the best fit in Fig. 1a. **(b)** Glutamate (black) substrate with an additional 10% (red), 50% (green) and 80% (blue) randomly chosen input channels. The best fit power law $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$ with $\nu_0 = 0.0004$ and $\alpha = 1.5$ is consistent with that of Fig. 1b.

In the power-law fittings of Figs. 1a, 1b and Fig. S1 we omitted the reactions with fluxes smaller than 10^{-5} for the sake of clarity. To check the validity of our findings for the full system (including the reactions omitted in Fig. S1), we plot the cumulative of the flux distribution obtained by optimization on succinate (black) and glutamate (red) rich substrates in Fig. S2. This figure clearly shows that this plotting procedure is sound, since

fluxes of magnitude less than 10^{-5} are outside the scaling region of the cumulative distribution and are fully in line with the fit used to determine the scaling exponents.

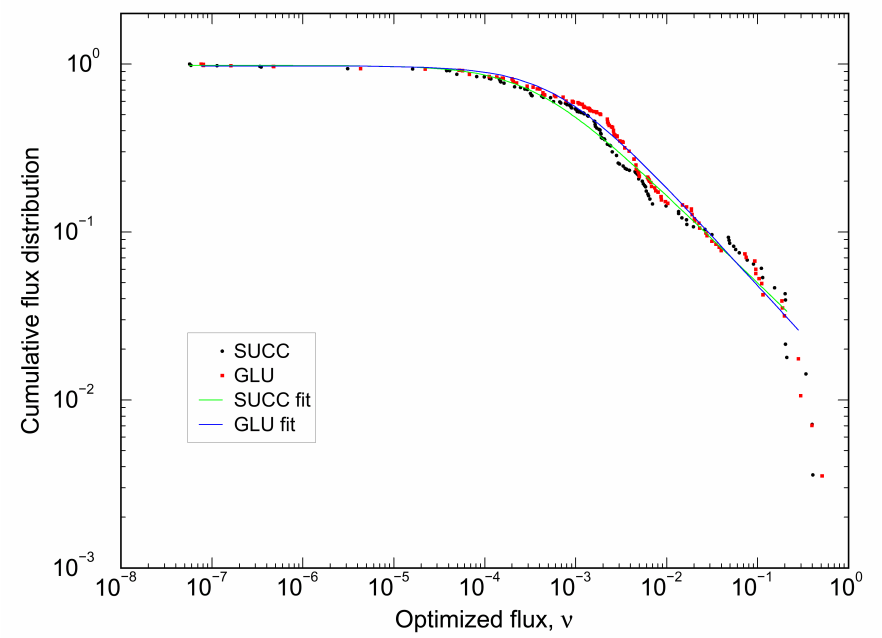

Figure S2. Cumulative distribution of the metabolic fluxes in *E. coli* on a succinate (black) and a glutamate (red) rich environment. The best fit to $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$ yields ν_0=*0.0003* and α=*0.53* for succinate (green), and ν_0=*0.0006* and α=*0.59* for glutamate (blue). The fitting does not include the 3 largest flux values for both glutamate and succinate.

2.2 Random uptake conditions: To investigate the effect of the environment on the flux distribution we choose randomly X%, (where X=10, 50 or 80) of the 89 potential input substrates *E. coli* consumes in addition to the minimal uptake basis of 6 (together with either glutamate or succinate, giving a total of 96 input channels). For each of the chosen transport reactions, we set the uptake rate to 20 mmol/g DW/h before computing the optimal flux distribution. As there is a very large number of possible combinations of the selected input substrates, we repeat this process 5000 times and average over each realization. In Fig. S1 (b), we show the resulting flux distribution for varying degrees of random environments superimposed on a glutamate rich substrate. The best fit power law is identical to that of Fig. 1 (b), for a succinate rich base, showing that the functional form

of the flux distribution is very robust and independent of the growth conditions (uptake metabolites).

2.3 Flux variations under changes in the growth conditions: For Fig. 4 we recorded the flux vector of each independent realization, allowing us to create a flux histogram for each reaction. As this figure shows, the distribution of the individual flux values can vary from Gaussian to multimodal and wide-scale distributions. In Figs. 2d and S3, we compare the calculated absolute value of the fluxes for each reaction on different substrates. The overall deviation from the $y=x$ line (red) is caused by the substrate's differing ability to produce biomass. A glucose rich medium gives higher biomass production than a glutamate rich one. Reactions with zero flux in one of the conditions are shown close to the coordinate axes. The inset shows the absolute relative difference.

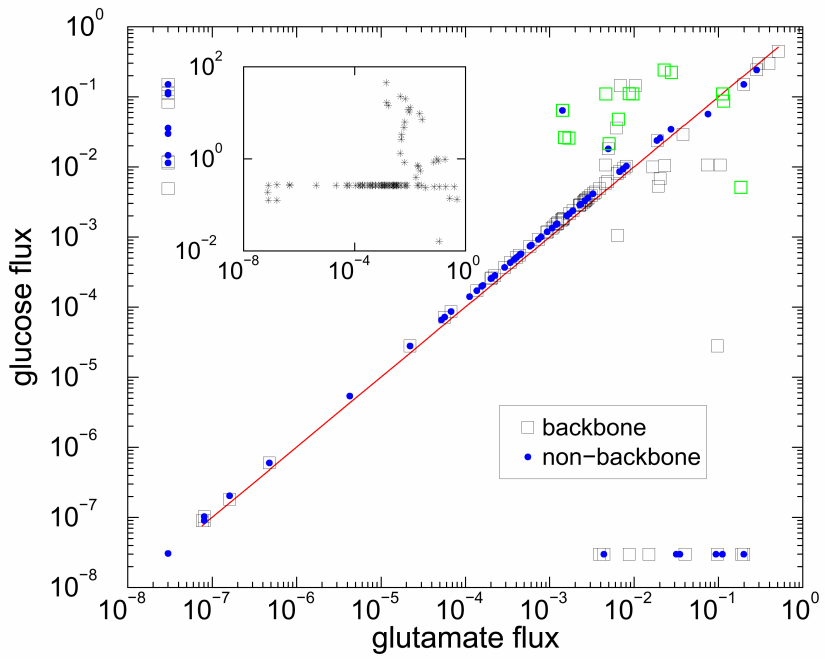

Figure S3. The change in the flux of individual reactions when departing from glutamate to glucose rich conditions. Some reactions are turned on in only one of the conditions (shown close to the coordinate axes). Reactions which are members of the flux backbone for either of the substrates are black squares, the remaining reactions are marked by blue dots and reactions reversing direction are colored green.

Additionally, to systematically quantify the flux fluctuations, we calculated the average flux value and standard deviation for each reaction, not considering the instances where a reaction was rendered inactive. Figure S4 displays our results for a minimal basis with glutamate and (a) 10%, (b) 50% and (c) 80% randomly chosen uptake substrates. For the small fluxes ($< 10^{-3}$) the fluctuation σ is typically linear in the average flux v. For all these reactions, the distribution of flux values is very well fit by a Gaussian distribution (v,σ). The reaction fluxes with a non-linear σ-v behavior have either multimodal or very broad scale flux distributions.

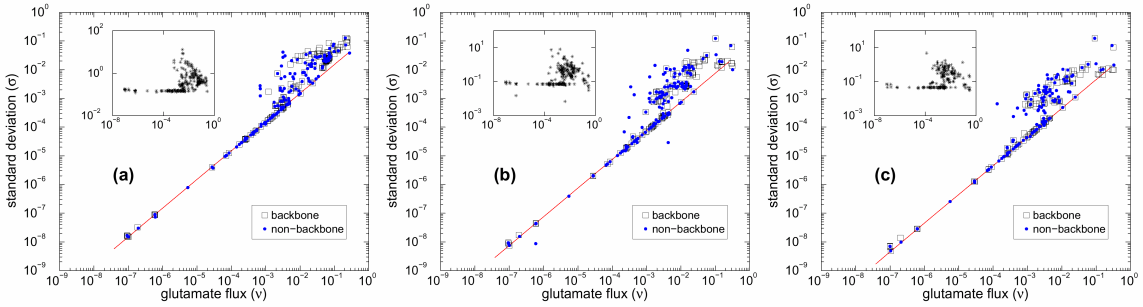

Figure S4. Absolute value of glutamate flux v_i for reaction i averaged over **(a)** 10%, **(b)** 50% and **(c)** 80% randomly chosen inputs, plotted against the standard deviation of that same reaction. The red line is $y=a\,x$ for reference purpose, with **(a)** $a=0.15$, **(b)** $a=0.075$ and **(c)** $a=0.045$. The inset displays the relative flux fluctuation σ_i / v_i per reaction.

3. Non-optimal states

3.1 The "hit-and-run" method: To characterize all the possible flux states of the system using only the constraints imposed by mass conservation and stoichiometry, we sample the solution space by implementing a "hit-and-run" algorithm (*4, 5*). For our database of MG1655 *E. coli* metabolic reactions (*6*) twenty metabolites were given transport reactions either supplying or removing the metabolites in question, in order to ensure mass conservation (see Table S1). We select a set of basis vectors spanning the solution space using singular-value decomposition (*7*). Since the reaction fluxes must be positive,

the "bouncer" is constrained to the part of the solution space intersecting the positive orthant. A schematic illustration of a 2-dimensional solution space embedded in a 3-dimensional flux space is shown in Fig. S5. Reactions which, for different reasons, cannot run are removed from the basis set (section **3.2** and Table S3). In order to render the volume of the solution space finite, we constrain the bouncer within a hypersphere of radius R_{max}. Also, to avoid numerical inaccuracies close to the origin, we constrain the "bouncer" to be outside of a hypersphere of radius $R_{min} < R_{max}$, and we find that the sampling results are independent of the choices of R_{min} and R_{max}. Starting from a random initial point (red) inside the positive flux cone (and between the constraining hyperspheres) in a randomly chosen direction (Fig. S5), the bouncer travels deterministically a distance d between sample points. Each sample point (green), corresponding to a solution vector where the components are the individual fluxes, is normalized by projection onto the unit sphere. After every b^{th} bounce off the internal walls of the flux cone, the direction of the bouncer is randomized.

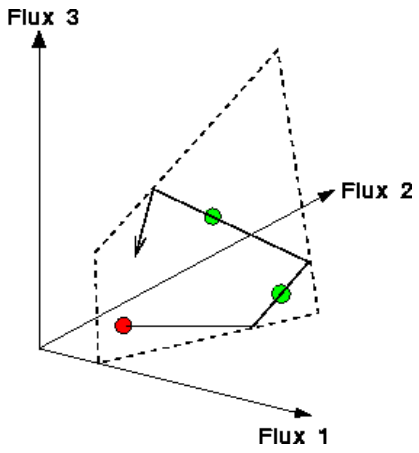

Figure S5. Schematic view of the "hit-and-run" sampling method.

Metabolite Name	Added Transport Reaction
4-Hydroxy-benzyl-alcohol	HBA => HBAxt
Spermidine	SPMD => SPMDxt
Histidine	HIS => HISxt
Heme O	HEMEO => HEMEOxt
Menaquinone	MK => MKxt
Leucine	LEU => LEUxt
N-Acetyl-D-mannosamine	NAMAN => NAMANxt
Peptide	PEPT => PEPTxt
Dipeptide	DIPEP => DIPEPxt
Oligopeptide	OPEP => OPEPxt
Peptidoglycan	PEPTIDO => PEPTIDOxt
Maltose 6-phosphate	MLT6P => MLT6Pxt
Enterochelin	ENTER => ENTERxt
Cadaverine	CADV => CADVxt
Valine	VAL => VALxt
Siroheme	SHEME => SHEMExt
Undecaprenyl pyrophosphate	UDPP => UDPPxt
Lippolysaccharide	LPS => LPSxt
N-Acetylglucosamine	NAGxt => NAG
1-D-Deoxyxylulose-5-phosphate	DX5Pxt => DX5P

Table S1. The list of metabolites either only consumed or produced in the MG1655 *in silico* model (*8*). To ensure mass conservation in the "hit-and-run" sampling, we had to add these transport reactions. Suffix "xt" indicates a metabolite external to the cell, as defined in Refs. 1 and 3.

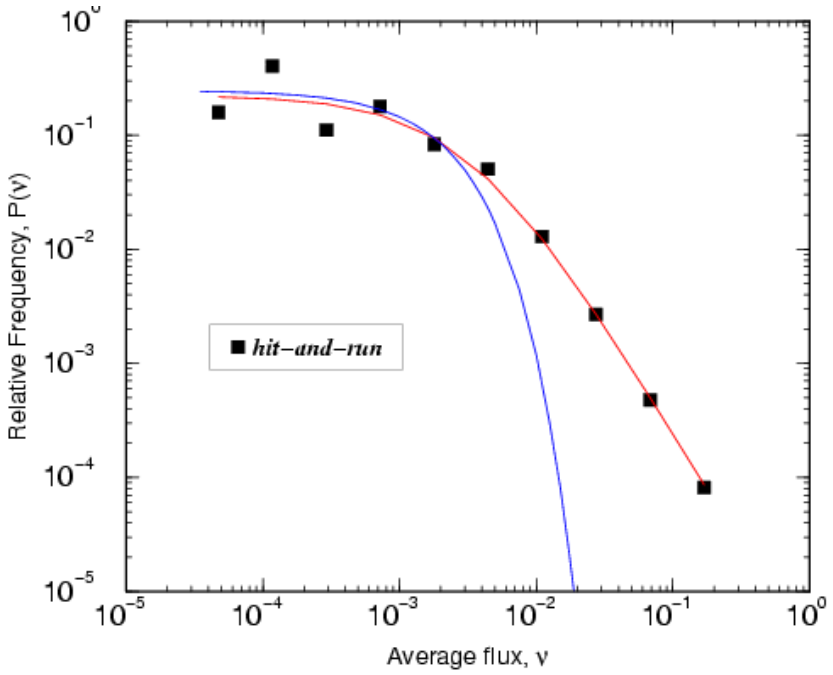

Figure S6. Flux distribution from "hit-and-run" sampling of the *E. coli* solution space. Solid line is the best fit to $P(v) \sim (v + v_0)^{-\alpha}$ with v_0=*0.003* and α=*2*. The best fit to the exponential form $P(v) \sim e^{-v/\xi}$ yields ξ=*0.002* (blue), indicating that such a fit is not appropriate to describe the observed distribution. While the obtained average flux distribution is consistent in shape and flux ranges with those obtained by the optimal FBA, the flux exponent is somewhat larger, and the quality of the scaling is slightly weaker. Interestingly, many individual non-optimal states (Fig. 1c, inset) are consistent with an exponent α=*1*, in accord with the experimental results (Fig. 1d), supporting the prediction[3,9] that these organisms may not have achieved optimality. In some states a power law with an exponential cutoff offered a better fit. These findings imply that the exponent may depend on the organism's position in the solution space, a finding that suggests that further analytical and numerical studies are needed to fully capture the development of the scaling in the optimal and non-optimal states.

3.2 Determination of initial points inside the flux cone: We implemented two different methods for locating possible starting points for the "hit-and-run" method. First, using linear optimization we calculate the flux vector $\boldsymbol{u_i}$ resulting from maximizing the flux

through reaction i with all fluxes constrained $\nu_j \le a$, repeated for all fluxes. Fluxes which can only be zero are removed from the stoichiometric matrix before calculating the vector basis spanning the solution space (see Table S3). The superposition of all the obtained different $\boldsymbol{u_i}$'s is a viable starting point for the "hit-and-run" method. Alternatively, we can determine a possible starting point from selecting a random point, $\boldsymbol{p}$, inside the solution space. We first define the vector orthonormal to a face of the positive orthant, $\boldsymbol{n_i}$, (e.g., the xy-plane) to have only zero or positive components. For all the orthant walls i for which the scalar product $\boldsymbol{n_i} \bullet \boldsymbol{p} = d_i < 0$, we move the orthant wall a distance $d_i + \varepsilon$ along $-\boldsymbol{n_i}$ until that the scalar product changes sign. The point $\boldsymbol{p}$ is now "inside" of the redefined flux cone. For every c^{th} bounce off a flux cone wall, we move the cone walls which are not intersecting the origin as close to the origin as possible while still keeping the bouncer inside the cone. When all cone walls are intersecting the origin again, the bouncer is inside of the original flux cone. It is necessary to remove from the stoichiometric matrix all reactions i corresponding to the null-vector ($\boldsymbol{n_i=0}$) in the orthonormal basis spanning the solution space, and all reactions i and j for which $\boldsymbol{n_i} = -\boldsymbol{n_j}$. These reactions correspond to the zero flux reactions determined by the optimization approach (see Table S3).

4. Fine structure of fluxes, Y(k)

To calculate $Y(k)$, for each metabolite i we determine the mass transport $\hat{\nu}_{ij}$ ($\hat{\nu}_{ij} = |S_{ij}| \nu_i$) for all incoming (outgoing) reactions j before calculating

$$Y(k,i) = \sum_{j=1}^{k} \left(\frac{\hat{\nu}_{ij}}{\sum_{l=1}^{k} \hat{\nu}_{il}} \right)^2$$

for each metabolite. We average over all metabolites which have k incoming (outgoing) topological links, resulting in $Y(k)$. If a reaction producing (consuming) metabolite i has a flux magnitude $\boldsymbol{a}$ (where $0<\boldsymbol{a}<1$, without loss of generality) much larger than the flux of the other reactions, which have comparable magnitudes $\boldsymbol{b}=(1-\boldsymbol{a})/(k-1)$, then

$Y(k,i) = \boldsymbol{a}^2[1+(\boldsymbol{b}/\boldsymbol{a})^2] \approx \boldsymbol{a}^2$. When all reactions have comparable flux values, $\boldsymbol{a}$, we have $Y(k,i) = [k\,\boldsymbol{a}^2/(k\,\boldsymbol{a})^2] = 1/k$. Figure S7 shows a schematic of these two extremes. In Figure S8, we show the calculated $Y(k)$ for (a) a glucose rich substrate and (b) on LB medium. Both cases display a high degree of local heterogeneity in the fluxes.

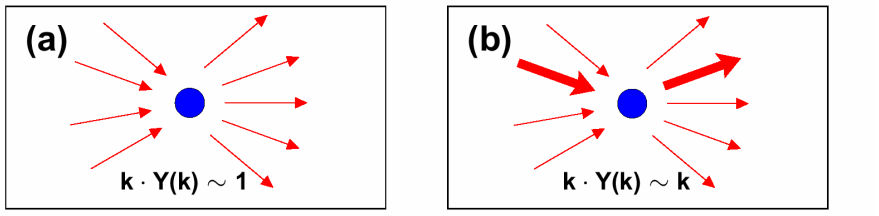

Figure S7. Schematic illustration of the hypothetical scenario in which **(a)** all fluxes have comparable activity, in which case we expect $kY(k) \sim 1$ and **(b)** the majority of the flux is carried by a single incoming or outgoing reaction, for which we should have $kY(k) \sim k$.

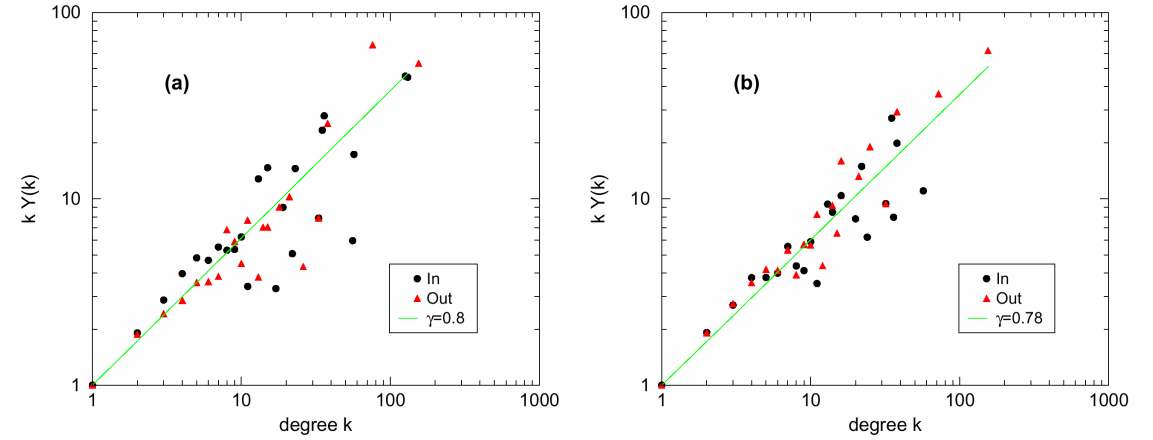

Figure S8. Fragmentation $Y(k)$ for the FBA optimization of *E. coli* on **(a)** glucose and **(b)** on LB for incoming (black) and outgoing (red) fluxes. The best fit (green) to the functional form k^{γ}.

We also investigated the local flux structure when a glutamate uptake basis with additional randomly selected uptake channels were activated (Fig. S9). The trend of strong local heterogeneity is present also for these cases, the power law exponent taking the values (a) 0% $\gamma = 0.75$, (b) 10% $\gamma = 0.67$, (c) 50% $\gamma = 0.71$ and (d) 80% $\gamma = 0.71$.

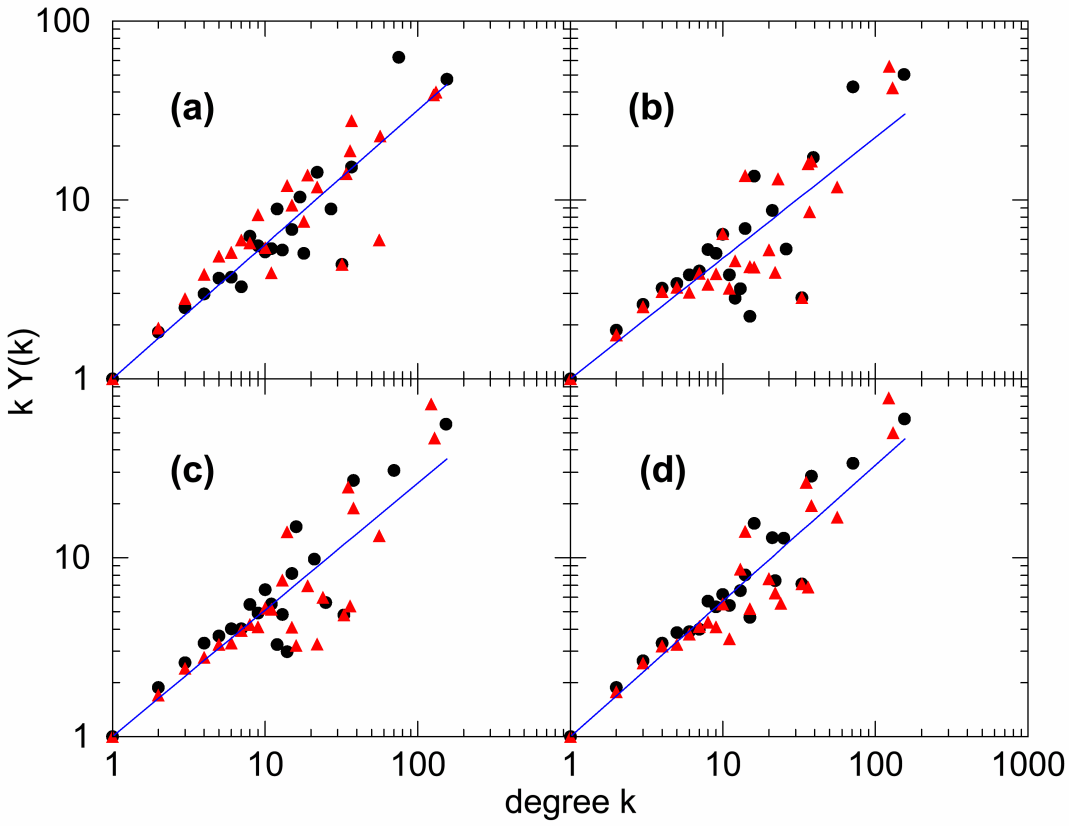

Figure S9. Fragmentation *Y(k)* for the FBA optimization of *E. coli* on **(a)** glutamate ($\gamma = 0.75$), **(b)** glutamate and 10% ($\gamma = 0.67$), **(c)** glutamate and 50% ($\gamma = 0.71$) and **(d)** glutamate and 80% ($\gamma = 0.71$) randomly chosen input channels for incoming (black) and outgoing (red) fluxes. The best fit (blue) to the functional form k^{γ}.

We have calculated *Y(k)* also for the flux distribution calculated by uniformly sampling the interior of the flux cone ("hit-and-run" method). In Fig. S10 (a), we detect that the local heterogeneity is not only limited to the optimized fluxes on *E. coli*.

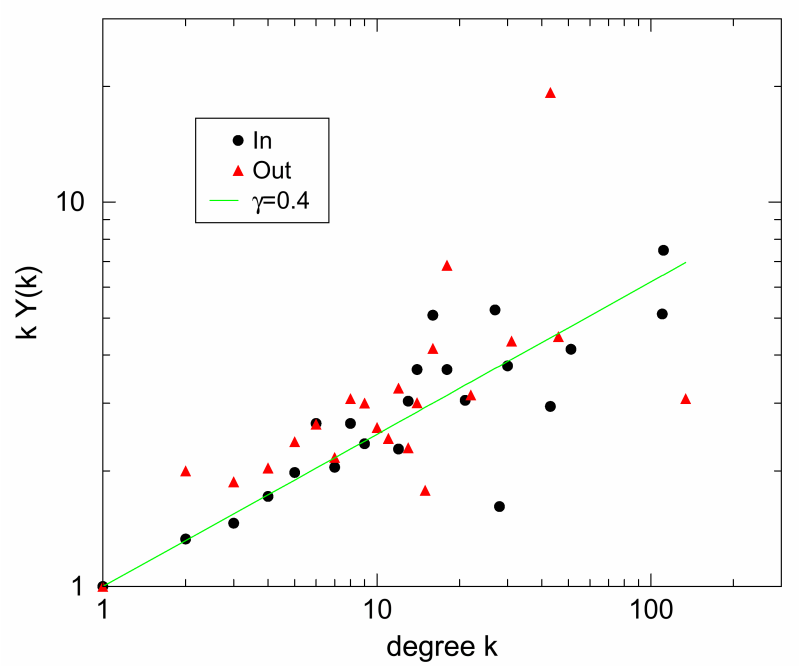

Figure S10. Fragmentation *Y(k)* for the hit-and-run sampling method of the state space of *E. coli* and for incoming (black) and outgoing (red) fluxes. The best fit (green) to the functional form k^{γ} with $\gamma = 0.4$.

5. High flux backbone

The high-flux backbone (HFB) is constructed as follows: For each metabolite we only keep the reactions with the largest flux producing (incoming) and consuming the metabolite (outgoing), discounting reactions with zero flux. Subsequently, a directed link is introduced between two metabolites A and B if (i) A is a substrate of the most active reaction producing B, and (ii) B is a product of the maximal reaction consuming A. We display only metabolites which are connected to at least one other metabolite after steps (i) and (ii). For clarity we removed P_i, PP_i and ADP from Figure 3.

While the metabolites of the HFB participate in numerous other reactions, the magnitude of mass transfer along the side reactions is less than the one along the detected

HFB. This is illustrated in Fig. S11, where we show the distribution of the ratio of the maximal to the next largest flux for each metabolite produced (consumed). The plot indicates that for the vast majority of metabolites there is either a single producing (consuming) reaction (consistent with the network's scale-free nature), or the most active reaction has a significantly larger mass flux ($\hat{v}_{max}$) than the next largest contribution ($\hat{v}_{2nd\text{-}max}$). Indeed, 273 of the 297 HFB reactions (Fig. S1(a)) have a $\hat{v}_{max} / \hat{v}_{2nd\text{-}max}$ ratio larger than two in a glutamate rich medium, similar high ratios being observed for other growth conditions as well.

In Fig. S11 we show the results for (a) glutamate (1st and 3rd columns) and succinate (2nd and 4th columns) and for (b) glucose (1st and 3rd columns) and LB (2nd and 4th columns) conditions. Metabolites with only a single producing (consuming) reaction are labeled "no 2nd".

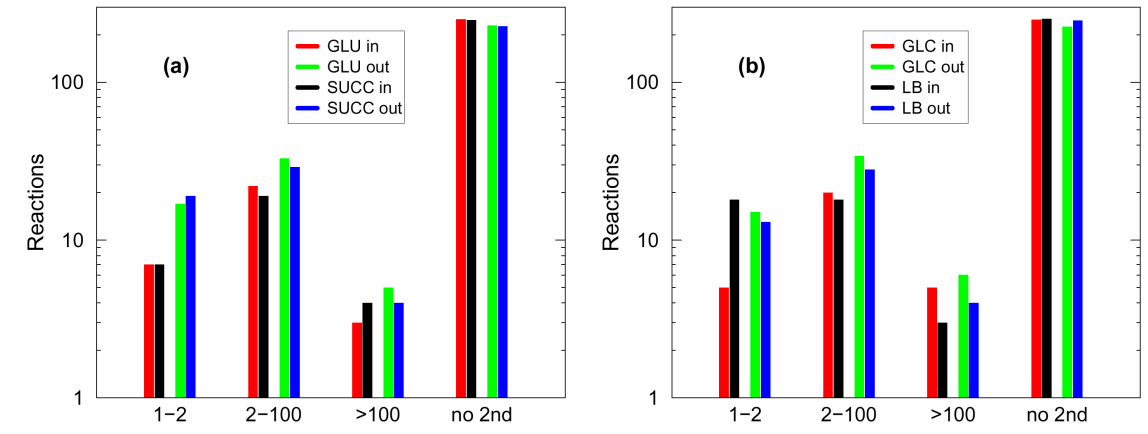

Figure S11. The histogram for the distribution of ratios $\hat{v}_{max} / \hat{v}_{2nd\text{-}max}$ between the largest and the second largest producing (consuming) mass flux for each metabolite on **(a)** glutamate (1st and 3rd columns) and succinate (2nd and 4th columns) and **(b)** glucose (1st and 3rd columns) and LB (2nd and 4th columns) conditions. Metabolites with only a single producing (consuming) reaction are labeled "no 2nd".

We also give a graphical representation of the HFB for two uptake conditions in Fig. S12. Only a few pathways, like Riboflavin and Folate biosynthesis appear disconnected,

indicating that while these pathways are part of the HFB, their end product serves only as the second most important source for some other HFB metabolite. The links of the reaction sequences are directed towards the biomass, which collects the set of metabolites produced by the cell to maintain optimal cell growth. The individual reaction groups largely overlap with the traditional, biochemistry-based partitioning of cellular metabolism: all metabolites of the citric-acid cycle of *E. coli* are recovered, and so are a considerable fraction of other known pathways, such as those being involved in histidine-, murein- and purine biosynthesis, to mention a few. Yet, the HFB represents a significant reduction of the complex network structure, emphasizing the subset of reactions which dominate the activity of the metabolism. As such, it offers a complementary approach to elementary flux mode analyses[10,11], which successfully captures the available modes of operation for smaller networks, but whose application to optimal E. *coli* has not yet been possible.

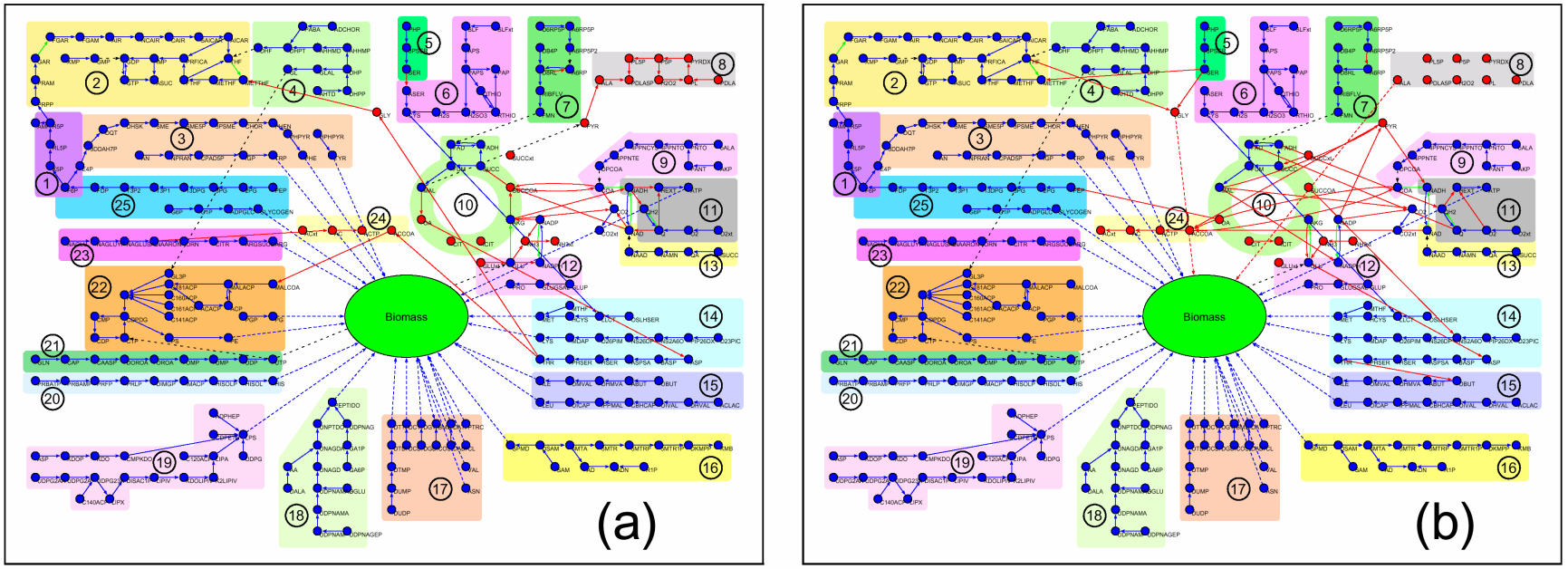

Figure S12. Maximal flow network constructed from the FBA optimized metabolic network of *E. coli* on a (a) glutamate rich and a (b) succinate rich substrate. We connect two metabolites A and B with a directed link pointing from A to B only if the reaction with maximal flux consuming A is the reaction with maximal flux producing B. We show all metabolites which have at least one neighbor after the completion of this procedure. The background colors demarcate different known biochemical pathways. The colors of the metabolites (vertices) and the reactions (edges) help compare (a) with (b): Metabolites colored blue have at least one neighbor in common in (a) and (b), while those colored red have none. Reactions are colored blue if they are identical in (a) and (b), green if a different reaction connects the same neighbor pair and red if this is a new neighbor pair. Black dotted edges indicate where the disconnected pathways (e.g. 4, Folate Biosynthesis) would connect to the cluster via a link that is not part of the HFB. Thus, the red nodes and links highlight changes in the wiring diagram. Dashed edges indicate links to the biomass growth reaction. The numbers

identify the various biochemical pathways correspond to: (1) Pentose Phospate, (2) Purine Biosynthesis, (3) Aromatic Amino Acids, (4) Folate Biosynthesis, (5) Serine Biosynthesis, (6) Cysteine Biosynthesis, (7) Riboflavin Biosynthesis, (8) Vitamin B6 Biosynthesis, (9) Coenzyme A Biosynthesis, (10) TCA Cycle, (11) Respiration, (12) Glutamate Biosynthesis, (13) NAD Biosynthesis, (14) Threonine, Lysine and Methionine Biosynthesis, (15) Branched Chain Amino Acid Biosynthesis, (16) Spermidine Biosynthesis, (17) Salvage Pathways, (18) Murein Biosynthesis, (19) Cell Envelope Biosynthesis, (20) Histidine Biosynthesis, (21) Pyrimidine Biosynthesis, (22) Membrane Lipid Biosynthesis, (23) Arginine Biosynthesis, (24) Pyruvate Metabolism and (25) Glycolysis.

6 Uptake Metabolites.

We give a list of the 96 possible substrates the *in silico E.coli* cell can assimilate, and from which we randomly select X% out of the 90 that are not in the minimal uptake basis. We have highlighted (yellow) the glutamate minimal uptake basis.

Abbrev.	Metabolite Name	Abbrev.	Metabolite Name
AC	Acetate	HIS	Histidine
ACAL	Acetaldehyde	HYXN	Hypoxanthine
AD	Adenine	ILE	Isoleucine
ADN	Adenosine	INS	Inosine
AKG	a-Ketoglutarate	K	Potassium
ALA	Alanine	LAC	D-Lactate
AMP	Adenosine monophosphate	LEU	Leucine
ARAB	Arabinose	LYS	L-Lysine
ARG	Arginine	MAL	Malate
ASN	Asparagine	MAN	Mannose
ASP	Aspartate	MDAP	Meso-diaminopimelate
BCAA	Branched chain amino acid	MELI	Melibiose
C140	Myristic acid	MET	Methionine
C160	Palmitic acid	MLT	Maltose
C180	Stearic acid	MNT	Mannitol
CO2	Carbon dioxide	NA	Sodium
CYS	Cysteine	NAD	Nicotinamide adenine dinucleotide
CYTD	Cytidine	NH3	Ammonia
CYTS	Cytosine	NMN	Nicotinamide mononucleotide
DA	Deoxyadenosine	O2	Oxygen
DALA	D-Alanine	OPEP	Oligopeptide
DC	Deoxycytidine	ORN	Ornithine
DG	Deoxyguanosine	PEPT	Peptide
DHA	Dihydroxyacetone	PHE	Phenylalanine
DIN	Deoxyinosine	PI	Phosphate (inorganic)
DIPEP	Dipeptide	PNTO	Pantothenate
DSER	D-Serine	PRO	Proline
DT	Thymidine	PTRC	Putrescine
DU	Deoxyuridine	PYR	Pyruvate
ETH	Ethanol	RIB	Ribose
FOR	Formate	RMN	Rhamnose
FRU	Fructose	SER	Serine
FUC	Fucose	SLA	Sialic acid
FUM	Fumarate	SLF	Sulfate
GABA	4-Aminobutanoate	SPMD	Spermidine
GL	Glycerol	SUC	Sucrose
GL3P	Glycerol 3-phosphate	SUCC	Succinate
GLAC	Galactose	THR	Threonine
GLAL	D-Glyceraldehyde	TRE	Trehalose
GLC	a-D-Glucose	TRP	Tryptophan
GLCN	Gluconate	TYR	Tyrosine
GLN	Glutamine	URA	Uracil
GLT	Glucitol	UREA	Urea
GLTL	Galacitol	URI	Uridine
GLU	Glutamate	VAL	Valine
GLY	Glycine	XAN	Xanthine
GN	Guanine	XTSN	Xanthosine
GSN	Guanosine	XYL	D-Xylose

Table S2. Uptake metabolites.

Pathway name	Enzyme	Gene name	Reaction	EC number
Ubiquinone				
	Chorismate pyruvate-lyase	ubiC	CHOR => 4HBZ + PYR	4.1.3.27
	Hydroxybenzoate octaprenyltransferase	ubiA	4HBZ + OPP => O4HBZ + PPI	2.5.1.-
	Octaprenyl-hydroxybenzoate decarboxylase	ubiD, ubiX	O4HBZ => CO2 + 2OPPP	4.1.1.-
	2-Octaprenylphenol hydroxylase	ubiB	2OPPP + O2 => 2O6H	1.13.14.-
	Methylation reaction		2O6H + SAM => 2OPMP + SAH	2.1.1.-
	2-Octaprenyl-6-methoxyphenol hydroxylase	ubiH	2OPMP + O2 => 2OPMB	1.14.3.-
	2-Octaprenyl-6-methoxy-1,4-benzoquinone methylase	ubiE	2OPMB + SAM => 2OPMMB + SAH	2.1.1.-
	2-Octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase	ubiF	2OPMMB + O2 => 2OMHMB	1.14.13.-
	3-Dimethylubiquinone 3-methyltransferase	ubiG	2OMHMB + SAM => QH2 + SAH	2.1.1.64
Coenzyme A Biosynthesis				
	ACP Synthase	acpS	COA => PAP + ACP	2.7.8.7
Glutathione Biosynthesis				
	Glutamate-cysteine ligase	gshA	CYS + GLU + ATP => GC + PI + ADP	6.3.2.2
	Glutathione synthase	gshB	GLY + GC + ATP => RGT + PI + ADP	6.3.2.3
Thiamin (Vitamin B1) Biosynthesis				
	ThiC protein	thiC	AIR => AHM	
	HMP kinase	thiN	AHM + ATP => AHMP + ADP	2.7.1.49
	HMP-phosphate kinase	thiD	AHMP + ATP => AHMPP + ADP	2.7.4.7
	Hypothetical		T3P1 + PYR => DTP	
	ThiG protein	thiG	DTP + TYR + CYS => THZ + HBA + CO2	
	ThiE protein	thiE	DTP + TYR + CYS => THZ + HBA + CO2	
	ThiF protein	thiF	DTP + TYR + CYS => THZ + HBA + CO2	
	ThiH protein	thiH	DTP + TYR + CYS => THZ + HBA + CO2	
	THZ kinase	thiM	THZ + ATP => THZP + ADP	2.7.1.50
	Thiamin phosphate synthase	thiB	THZP + AHMPP => THMP + PPI	2.5.1.3
Vitamin B6 (Pyridoxine) Biosynthesis				
	Pyridoxal-phosphate biosynthetic proteins pdxJ-pdxA	pdxAJ	PHT + DX5P => P5P + CO2	4.2.99.2
	Threonine synthase	thrC	PHT => 4HLT + PI	4.2.99.2
	Hypothetical enzyme		4HLT => PYRDX	4.1.1.-
	Added reaction (mass conservation)		HBA => HBAxt	
	Added reaction (mass conservation)		DX5Pxt => DX5P	

Table S3. List of reactions from the *in silico* MG1655 model (*8*) which can only have zero flux.

7. Calculating analytically the flux exponents for model systems

In the following, we show that the distribution of fluxes on a network with scale-free topology displays power-law decay as one goes from small to larger fluxes. There are two ways to demonstrate this on various constructs with scale-free edge-distributions: a) exact calculations on deterministic trees and b) theoretical and numerical calculations of the distribution on stochastic scale-free networks. Here we discuss both approaches.

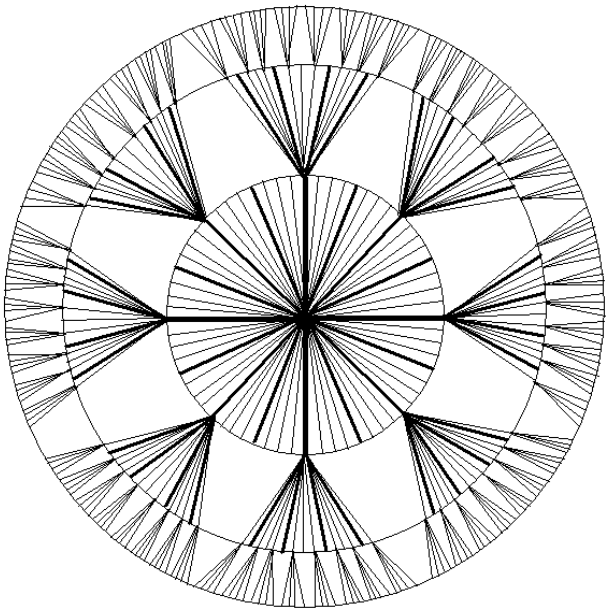

Figure S13. Deterministic scale-free tree with flux flow. The circles are only guides to the eye and not a part of the network.

a) First we construct a deterministic scale-free tree. Then, we assume that a given pattern for the inflow and outflow the flux distribution can be obtained exactly from a rather simple calculation. We define, by giving the corresponding rules of construction, the following family of scale-free trees:

Step 0: Start with n edges going out radially from a centre.

Step 1: Substitute each edge with n new edges "starting" from the center and "ending" on a circle (naturally, this circle is used only for visualization purposes).

Step 2: From every m^{th} node out of the the n^2 nodes on the circle, draw n new edges, so that each new edge ends on a new concentric circle.

Step 3: Repeat Step 1 and Step 2, so that at each time, an edge is substituted with n edges starting from an edge's inner end. Step 2 is carried out only for edges in the outermost layer.

Figure S13 shows a schematic visualization of these rules. Note that circles are shown only to guide the eye. The width of the lines decreases as the number of iterations increases. After q iterations, there are

Number of nodes	Number of edges
j^0	n^{q+1}
j^1	n^q
.....	.....
j^q	n

where $j = n^2 / m$ and q is the number of iterations carried out. We did not count the single edges leading "back" from a node since they are negligible for large graphs (if n^2 / m is not an integer, $j = \lfloor n^2 / m \rfloor + 1$ can be used). In the example, the parameter choices are $n = 4$, $m = 2$ and $j = 8$. Thus, using these numbers and denoting the ratio of nodes having k edges as $P'(k)$, we have

$$P'(4k) = P'(k)/8,$$

(there are 8 times less nodes having 4 times more edges). Assuming, that $P'(k)$ scales as $k^{-\delta}$, we see that this form for $P'(k)$ indeed satisfies the above equation with $\delta = \ln 8 / \ln 4 = 3/2$. This $P'(k)$ is defined only for a discrete set of k values (powers of 4). In order to obtain the corresponding "smooth" distribution (a distribution fitted in such a way that its cumulative counterpart has the same scaling behavior as that of the discrete one, but has values for all k's, we have to correct for the scaling of the "binning size" (distance between the discrete k values occurring in the construction) and obtain:

$$P'(k) \sim k^{-\gamma} \sim k^{-(3/2+1)}$$

such that $\gamma = \delta + 1 = 2.5$. It directly follows from the above that in general, $\gamma = \ln j / \ln n + 1$.

Next, we calculate the flux distribution assuming that an amount of flux $\nu = 1$ enters the network at its "outmost" edges and that the outflow takes place at the central node. Thus, the flow is inward, and since at every level going from outside to inside, the fluxes from the 4 incoming nodes add up. It is easy to see, in complete analogy with the above, that for a construction of q steps, there are

Number of edges	Amount of flux
j^0	$n\, j^q$
j^1	$n\, j^{q-1}$
.....	.....
$j^q\, n$	*1*

In the above example, $n = 4$, $m = 2$ and $j = 8$. For these numbers, denoting the ratio of nodes having flux ν as $P'(\nu)$, we have

$$P'(8\nu) = P'(\nu)/8$$

(there are 8 times less edges having 8 times more flux). Assuming, that $P'(\nu)$ scales as $\nu^{-\delta}$, we see that this form for $P'(\nu)$ indeed satisfies the above equation with $\delta = 1$. Interestingly enough, for the above construction, δ is equal to 1 for all possible choices of *n* and *m* as can be directly seen from the fact that the flux intensity and the number of corresponding edges scale inversely for all *m*. Again, we are interested in the corresponding continuous distribution and as above, we get,

$$P(\nu) \sim k^{-\alpha},$$

such that $\alpha = 2$.

Here we have to make a few relevant remarks. i) We obtained for the flux distribution an exponent $\alpha = 2$ independent of the two parameters of the construction. This result is in agreement with that of Goh *et al* obtained for a related problem (PNAS 99 (2002) p. 12583). ii) The assumption of unit fluxes entering the network does not affect the scaling behavior: as a large number of fluxes are added along the paths even if we assume fluctuating values for

the entry fluxes, the differences coming from their fluctuations average out and do not modify scaling. iii) We have also investigated the case when influx takes place at each "free" node of the construction. The corresponding calculation is less straightforward, but leads to exactly the same result $\alpha = 2$. iv) We considered a tree above, while a metabolic network contains loops. On one hand, the role of loops has been shown in our paper to be relatively insignificant; on the other hand, the inclusion of loops is not expected to change the scaling behavior (see next section).

b) As for further theoretical arguments supporting the scaling of the flux distribution in stochastic scale-free networks, we should make two points: i) It has recently been shown (PNAS 99 (2002) p. 12583) that the distribution of a quantity called "betweenness-centrality" (BC) obeys a universal power law decay in all of the theoretically and numerically investigated scale-free networks. On the other hand, BC is closely related to a quantity called "load" of a node which, in turn is closely related to the flux through the incoming edges (see the PNAS reference for further details). ii) Finally, we have numerically investigated the effects of perturbing a scale-free network on the flux distribution. This was carried out by randomly adding and removing edges to an originally scale-free graph and calculating the resulting modified flux distribution. Our preliminary results show that the flux distribution on scale-free networks is robust against such perturbations, of course, up to a point, where the underlying topology qualitatively changes.

8. Distribution of experimentally determined fluxes

Using the experimentally determined fluxes for the central metabolism of *E. coli* (see Ref. 12 for details), we calculated the flux distribution as follows. In Ref. 12, the authors give the values for a total of 189 individual flux measurements, of which 31 are given for *E. coli* strain JM101 under 3 different external conditions (providing 93 fluxes) and 32 are given for *E. coli* strain PB25 under the same 3 external conditions (96 fluxes). Since the number of fluxes is low for each separate combination of experimental condition and cell strain, we calculated the distribution of all fluxes (Fig. 1d in the Manuscript). We reason that this procedure is sensible, since our theoretical and numerical calculations indicate that the resulting power-law flux distribution is insensitive to the details of the extra-cellular environment (see Figs. 1a-c and S1).

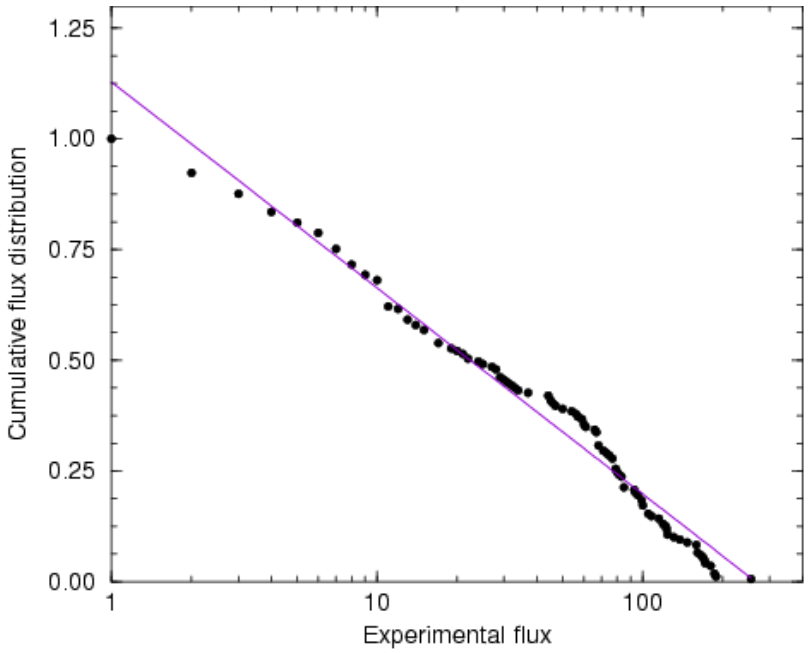

Figure S14. Cumulative distribution of experimental fluxes. The solid line is the best fit logarithmic curve $C(\nu) = \lambda \ln(\nu / \kappa)$ with $\lambda = -0.2$ and $\kappa = 2.7 \cdot 10^2$.

The best power-law fit to the flux distribution $P(\nu)$ gives an exponent of $\alpha \cong 1$. This suggests that $C(\nu)$, the cumulative flux distribution, is logarithmic, since it is related to the

histogram through $C(\nu) = \int_{\nu}^{\kappa} P(\nu')d\nu'$, where κ is the upper flux cut-off. In Fig. S14, we show the cumulative flux distribution together with the best fit logarithmic curve, substantiating our claim.

References

1. C. H. Schilling, B. O. Palsson, *Proc Natl Acad Sci U S A* **95**, 4193-8 (1998).
2. K. F. Jensen, *J Bacteriol* **175**, 3401-7. (1993).
3. D. Segre, D. Vitkup, G. M. Church, *Proc Natl Acad Sci U S A* **99**, 15112-7 (2002).
4. R. L. Smith, *Operations Res.* **32**, 1296-1308 (1984).
5. L. Lovász, *Math Programming* **86**, 443-461 (1999).
6. J. S. Edwards, B. O. Palsson, *Proc Natl Acad Sci U S A* **97**, 5528-33 (2000).
7. N. D. Price, J. L. Reed, J. A. Papin, I. Famili, B. O. Palsson, *Biophys J* **84**, 794-804 (2003).
8. J. S. Edwards, R. U. Ibarra, B. O. Palsson, *Nature Biotech* **19**, 125-30. (2001).
9. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Nature* **420**, 186-9 (2002).
10. Schuster, S., Fell, D. A. & Dandekar, T. *Nat Biotechnol* **18**, 326-32 (2000).
11. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. *Nature* **420**, 190-3 (2002).
12. Emmerling, M. et al. *J Bacteriol* **184**, 152-64 (2002).