Metabolic network structure determines key aspects of functionality and regulation

Jörg Stelling*, Steffen Klamt*, Katja Bettenbrock*, Stefan Schuster† & Ernst Dieter Gilles*

** Max Planck Institute for Dynamics of Complex Technical Systems, D-39106 Magdeburg, Germany*
† Max Delbrück Center for Molecular Medicine, D-13092 Berlin, Germany

The relationship between structure, function and regulation in complex cellular networks is a still largely open question[1–3]. Systems biology aims to explain this relationship by combining experimental and theoretical approaches[4]. Current theories have various strengths and shortcomings in providing an integrated, predictive description of cellular networks. Specifically, dynamic mathematical modelling of large-scale networks meets difficulties because the necessary mechanistic detail and kinetic parameters are rarely available. In contrast, structure-oriented analyses only require network topology, which is well known in many cases. Previous approaches of this type focus on network robustness[5] or metabolic phenotype[2,6], but do not give predictions on cellular regulation. Here, we devise a theoretical method for simultaneously predicting key aspects of network functionality, robustness and gene regulation from network structure alone. This is achieved by determining and analysing the non-decomposable pathways able to operate coherently at steady state (elementary flux modes). We use the example of *Escherichia coli* central metabolism to illustrate the method.

Elementary-mode analysis establishes a link between structural analysis and metabolic flux analysis (MFA). Elementary flux modes can be defined as the smallest sub-networks enabling the metabolic system to operate in steady state[7]. For example, in a hypothetical network (Fig. 1), five elementary modes exist, which cannot further be decomposed. By linear combination of e_1 to e_5 all thermodynamically and stoichiometrically feasible stationary flux distributions can be obtained. In each elementary mode, the enzymes are weighted by the relative fluxes they carry. Up to the non-negative scaling factors for each mode, the set of elementary modes is unique for a given network structure[8]. Hence, it enables us to investigate the space of all physiological states that are meaningful for the cell in the long-term perspective. Flux balance analysis (FBA), in contrast, uses linear programming to obtain a single (not necessarily unique, see Fig. 1) solution to an optimization problem, which is in most cases maximal growth per substrate uptake[2]. Accordingly, FBA focuses on a specific behaviour. It can thus not cope with cellular regulation without additional constraints.

Elementary-mode analysis has mainly been applied to biochemical networks of moderate complexity[1,7–9]. To explore the utility of the approach for a system of realistic complexity, we chose the central metabolism of the bacterium *Escherichia coli* as an example. In analogy to other network analyses[2,10], central carbon metabolism was modelled in (partially extended) detail, whereas in the anabolic part of the model, combining predominantly linear pathways into single assembly reactions served to reduce model complexity. The growth rate is approximated by the production rate of macromolecular cellular constituents such as DNA and protein. We model growth as one reaction converting a fixed ratio of precursors into biomass. Altogether, the network contains 89 substances and 110 reactions, of which 68 reactions can be attributed to single gene products or to multi-enzyme complexes cooperating in a single reaction (see Supplementary Information).

We determined elementary flux modes for the utilization of representative substrates feeding into different parts of metabolism (Table 1). The total number of elementary modes for given conditions is here used as a quantitative measure of the degrees of freedom[11], that is, of flexibility. Glucose, for example, can be used in approximately 45 times more different ways than acetate, corresponding to biological intuition. Flux mode number thus directly relates network structure to function. An empty set implies that no steady-state flux distribution fulfilling the specifications exists, hence predicting an inviable phenotype. For instance, anaerobic use of any of the four substrates except glucose is impossible without additional terminal electron acceptors.

In particular, we analyse the ability to grow or not to grow of mutants carrying deletions in single genes. For this purpose, the number of flux modes for a mutant Δi using substrate S_k is determined by (additionally) selecting for those flux modes that do not require gene i. We denote by $N(\mu,\Delta i)$ the number of flux modes showing a positive growth rate μ for this mutant. The relative number of flux modes for mutant strains (Supplementary Information) allows a correct prediction of the experimentally determined growth phenotype in the overwhelming majority of cases (Fig. 2a). Most situations with an empty (non-empty) set of flux modes correspond to inviable (viable) mutants. The only two false negatives are phosphogluco-isomerase (*pgi*) mutants, because in the model, growth depends on glucose-6-phosphate production, whereas *in vivo* this precursor is substitutable[12]. Erroneous positive predictions may be caused by insufficient pathway capacities (kinetic constraints) *in vivo*. A statistically significant classification of growth behaviour ($P < 10^{-5}$) results from the analysis. Altogether 90% of the predictions (81 out of 90 cases) were correct, which justifies usage of the relative number of elementary modes as a reliable indicator of network function.

As fault-tolerance is an essential feature of living systems, we investigated the structure–function relationship with respect to network robustness. We define robustness as insensitivity of network function, that is, the ability to sustain bacterial growth, towards internal disturbances like mutations[13]. The number of elementary modes qualitatively indicates whether a mutant is viable or not, but it does not necessarily describe to what extent a mutation affects growth quantitatively. We therefore additionally calculated the maximal biomass yield Y^{max} for each combination of mutant and substrate as a quantitative measure of network performance. Central metabolism of *E. coli* behaves in a highly robust manner, because mutants with significantly reduced metabolic flexibility

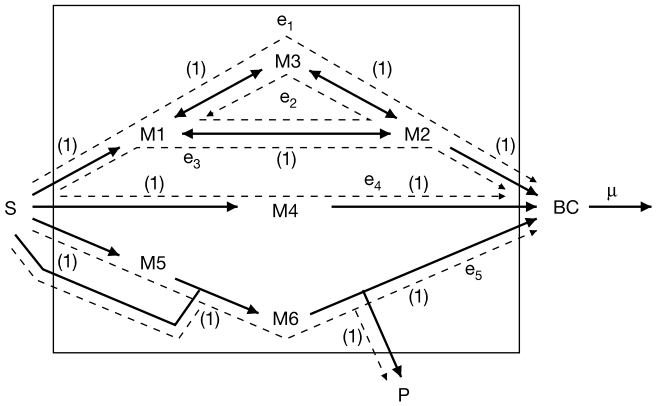

Figure 1 Example network. Reactions (solid arrows, 1:1 stoichiometry for substrates and products) convert substrate S into a biomass component BC and a secreted by-product P via internal metabolites M1–M6. Cellular growth rate μ is approximated by the production of BC. The hypothetical network comprises five elementary flux modes e_i (dashed arrows, relative flux in parentheses). The modes e_1, e_3 and e_4 give the same BC:S yield of 1:1, while e_5 gives a yield of 1:2.

show a growth yield similar to wild type (Fig. 2b). Robustness relies, at least in part, on pathway redundancy. Analysis of the set of elementary modes in wild type reveals the existence of multiple, alternative pathways with identical biomass yield (Fig. 2c). Only when the number of elementary modes is severely cut down by a mutation is functionality affected. Hence, elementary-mode analysis points to a coexistence of robustness and fragility, as already shown for cellular regulation[14–16].

Graph-theoretical methods are widely used to analyse complex networks[5,17]. In particular, the network diameter D, defined as the average minimal path length between any two nodes (substances), has been shown to be relatively invariant upon random removal of nodes in metabolic networks. This has been suggested to reflect network robustness, as an increasing diameter would indicate network disintegration[5]. However, for the network studied here, a constant diameter does not necessarily imply identical functionality (Fig. 2b). Robustness and fragility, hence, are not predicted by a pure graph-theoretical measure of network topology. In contrast to the network diameter, elementary modes reflect specific characteristics of metabolism such as molar yields. We therefore tackled the question whether the number of elementary modes N directly relates to network robustness. As a measure for robustness, we used the maximal growth yield for each mutant as already shown in Fig. 2b. Counting the number of cases for which $Y^{\max}(\Delta i) > 0$ gave the number of viable single-gene mutants for each substrate regime, that is, a measure of the probability to tolerate random deletion mutations. For different substrate uptake regimes, the organism's resistance to arbitrary gene deletions correlates well ($r^2 = 0.93$) with N for the corresponding wild type (Fig. 2d). The number of elementary modes thus provides an estimate for fault-tolerance.

Finally, we address whether regulation in complex metabolic networks could be predicted by elementary-mode analysis. A direct, quantitative correlation between metabolic fluxes and transcriptome or proteome patterns has not been observed[10,18,19]. However, the existence of a more indirect link seems likely. We start from the assumption that optimization during biological evolution can be characterized by the two objectives of flexibility—associated with robustness—and efficiency[11,13]. This is, for example, supported by evidence from the evolution of energy transduction[20].

Flexibility means the ability of cellular systems to adapt to a wide range of environmental conditions, that is, to realise a maximal bandwidth of thermodynamically feasible flux distributions, hence of elementary flux modes. Efficiency, as the second objective, could be defined as the fulfilment of cellular demands with an optimal outcome such as maximal cell growth[2,6], using a minimum of constitutive elements (such as genes and proteins)[11]. Since these two criteria impose contradictory challenges, optimal cellular regulation needs to find a trade-off. Our analysis will therefore rely on a parameter characterizing flexibility and efficiency derived from metabolic network structure, for which we introduce the term 'control-effective flux'.

Control-effective fluxes are determined directly from the set of elementary modes, and do not require optimization. The analysis begins by assigning an efficiency to each elementary mode. These efficiencies relate the mode's output (growth or ATP production) to the investment required to establish the mode, that is, to produce the enzymes. This investment is approximated by the sum of all (absolute) fluxes, because for comparable metabolite concentrations, the flux through an enzymatic reaction scales linearly with enzyme concentration. For the hypothetical network (Fig. 1), elementary mode e_4 would be favoured over e_1 and e_3 involving more enzymatic steps for the same growth yield (see Appendix A, Supplementary Information). Subsequently, we determine control-effective fluxes for a specific reaction as the (normalized) average flux through this reaction in all elementary modes, whereby for each mode the actual flux is weighted by the mode's efficiency (see Methods). In general, control-effective fluxes represent the importance of each reaction for efficient and flexible operation of the entire network. In contrast to FBA, this approach takes network flexibility directly into account because all optimal and sub-optimal modes are considered.

As cellular control on longer timescales is predominantly achieved by genetic regulation, the control-effective fluxes should correlate with messenger RNA levels. Theoretical transcript ratios $\Theta(S_1,S_2)$ for growth on two alternative substrates S_1 and S_2 were therefore calculated as ratios of control-effective fluxes and compared to previously published complementary DNA-microarray data for *E. coli* growing exponentially on glucose, glycerol and acetate[10,19]. The structure-derived prediction of the differential expression of 50 genes for acetate versus glucose shows good agreement with experiment (Fig. 3a). A test for systematic errors was subsequently performed by comparing the distribution of residuals to the normal distribution resulting from completely

Table 1 **Number and distribution of elementary flux modes.**

Selection*		Glucose	Acetate	Glycerol	Succinate	Sum
-	N	27,099	598	11,332	4,249	43,279
Growth only	$N(\mu, \neq ATP)$	73.1%	58.7%	78.6%	76.3%	74.6%
ATP only	$N(\neq \mu, ATP)$	3.2%	5.0%	2.4%	2.4%	3.0%
Growth and ATP	$N(\mu, ATP)$	6.6%	2.0%	5.1%	4.2%	5.9%
No growth/ATP	$N(\neq \mu, \neq ATP)$	17.1%	34.3%	13.9%	17.1%	16.5%
Aerobic growth	$N(\mu, O_2)$	73.1%	60.7%	83.6%	80.5%	76.4%
Anaerobic growth	$N(\mu, \neq O_2)$	6.6%	0.0%	0.0%	0.0%	4.1%

*We denote the number of elementary flux modes simultaneously meeting a set of conditions, $C_1,\ldots,C_n$, by $N(C_1,\ldots,C_n)$. These conditions include, for example, the situation where cells can grow, which is abbreviated by μ. Excess energy production in the form of ATP (ATP), the substrate metabolized (S_k for the k-th substrate) and oxygen uptake (O_2) are specified accordingly. The operator ' $\neq$ ' indicates that certain fluxes must not occur. The total number of modes includes one futile cycle without substrate uptake.

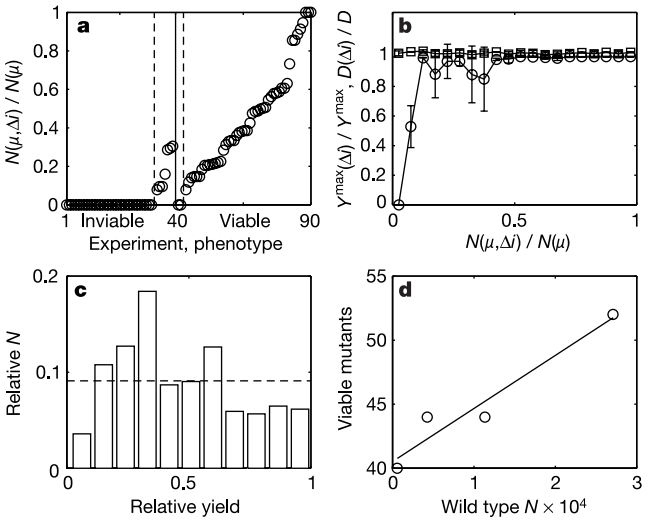

Figure 2 Metabolic network topology and phenotype. **a**, Relative number of elementary modes N enabling deletion mutants in gene i (Δi) of *E. coli* to grow (abbreviated by μ) for 90 different combinations of mutation and carbon source. The solid line separates experimentally determined mutant phenotypes, namely inviability (experiments 1–40) from viability (experiments 41–90). Dashed lines delimit the situations with erroneous predictions. **b**, Dependency of the mutants' maximal growth yield $Y^{max}(\Delta i)$ (open circles) and the network diameter $D(\Delta i)$ (open squares) on the share of elementary modes operational in the mutants. Data were binned to reduce noise. **c**, Distribution of growth-supporting elementary modes in wild type (rather than in the mutants), that is, share of modes having a specific biomass yield (the dotted line indicates equal distribution). **d**, Effect of arbitrary single-gene deletions on viability for single substrate uptake (open circles) assessed by the mutants' maximal growth yields as in **b**, but considering the numbers of cases in which $Y^{max}(\Delta i) > 0$, and relating them to the total number of modes for the four substrates in wild type (Table 1).

random deviations[21]. It leads to identification of three presumable outliers (Fig. 3b). Two overestimated transcript ratios are linked to genes involved in acetate metabolism (*pta*, *ackA*) and, hence were expected to be upregulated on acetate. This can be explained by the fact that the gene *acs* codes for an enzyme operating in parallel; for this gene our theoretical prediction is in agreement with experimental observation[19]. The transcript ratio of *aspA* encoding for aspartase was underestimated. Elementary-mode analysis suggests that aspartase is required for an effective conversion of excess NADPH generated by the TCA cycle to NADH (Supplementary Information). Residual analysis thus sustained or generated hypotheses amenable to further experimental investigation. Removal of the three outliers from a total of 50 data sets leads to to a high correlation between prediction and experiment ($r^2 = 0.60$) with a linear regression close to perfect match. For instance, average expression ratios from independent experimental studies[10,19] correlate with $r^2 = 0.84$ (not shown).

Additional evidence supporting our approach is provided by three observations. First, predictions based solely on efficiency, namely on the two flux modes with optimal ATP and biomass yield, Θ^{opt} (Ac, Glc), displayed a weak correlation (Fig. 3c). Only 28 data points appear because many fluxes are zero and, hence, yield zero or undefined predictions. Such an efficiency analysis corresponds to the approach followed by FBA. Neglecting flexibility may explain why FBA—even when supplied with information on regulatory circuits—only provides qualitative predictions for a subset of genes[22]. Secondly, in view of the fact that experimental errors are large compared to the effects of changes in the medium, the predictions for other combinations of substrates also agree reasonably well with experiment (Fig. 3d). Finally, prediction quality was poor when for elementary-mode analysis a simultaneous uptake of all substrates was enabled ($N = 507{,}633$, not shown). The combination of these findings points to a multi-level, hierarchical organisation of metabolic regulation[23]. Transcriptional regulation for growth on a specific substrate seems to rely on selection of this substrate regime by the cell, for instance by catabolite repression. According to the substrate regime, gene expression levels are, at an intermediate level of control, adjusted to provide a general set-up for metabolic efficiency and flexibility. At a lower level, short-term regulation of fluxes for a specific situation, such as for one defined substrate concentration, can then be achieved, for example, by allosteric control of metabolic enzymes[6].

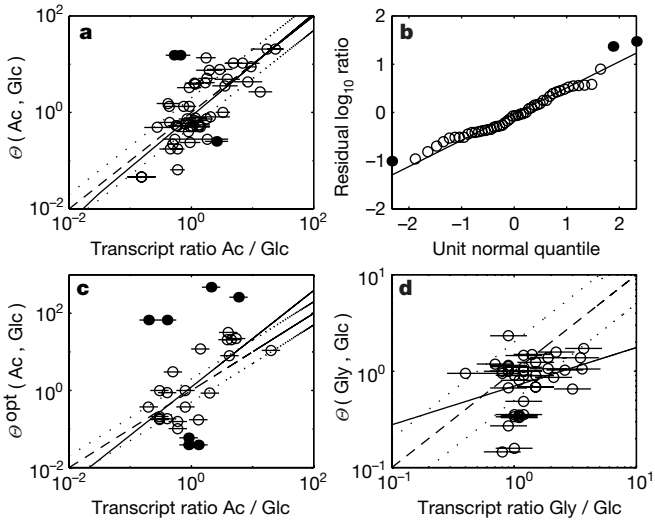

Figure 3 Prediction of gene expression patterns. **a**, Calculated ratios between gene expression levels during exponential growth on acetate and exponential growth on glucose (filled circles indicate outliers) based on all elementary modes versus experimentally determined transcript ratios[19]. Lines indicate 95% confidence intervals for experimental data (horizontal lines), linear regression (solid line), perfect match (dashed line) and two-fold deviation (dotted line). **b**, Distribution of residuals, that is, differences between measurement and prediction, compared to a normal distribution of prediction errors. Outliers (filled circles) were identified by their deviation from the linear regression (solid line). **c**, Predicted transcript ratios for acetate versus glucose for which, in contrast to **a**, only the two elementary modes with highest biomass and ATP yield (optimal modes) were considered. **d**, Gene expression ratios for growth on glycerol versus glucose[10] in analogy to **a**.

Elementary-mode analysis decomposes complex metabolic networks into simpler units performing a coherent function. The integrative analysis of elementary modes presented here can be used to reconstruct key aspects of cellular behaviour from metabolic network topology, namely to reliably classify mutant phenotypes, to analyse network robustness, and to quantitatively predict functional features of genetic regulation. Including additional knowledge on, for example, newly annotated genes is straightforward[8]. A refined approximation of bacterial growth could serve to improve our method. The concept of extreme pathways[24] bears strong similarity with elementary modes[7,8,11]. For our model of *E. coli* central metabolism, both approaches yield equivalent sets of functional entities and, thus, identical results (not shown). Whereas these approaches characterize the spectrum of different, potential functionalities of the metabolic system, FBA focuses on a single flux distribution. FBA provides similarly good predictions of mutant phenotypes[2], but it fails whenever network flexibility—for instance, in the analysis of pathway redundancy or in quantitative prediction of gene expression—has to be taken into account. Elementary-mode analysis, in contrast, helps us understand the huge amount of mRNA expression data provided nowadays. Subsequent studies will apply the analysis developed here to organisms other than *E. coli* and further validate it with upcoming transcriptome and mutant data. More generally, we conclude that robustness of metabolic networks is linked to redundancy, and that hierarchical genetic control supports this robustness by finding a trade-off between network efficiency and flexibility. Like recent studies demonstrating the power of combining data and methods from different origins[25–27], the systems biology analysis presented here can thus contribute to elucidation of the fundamental design principles of living cells. □

Methods

Network model

For the catabolic part of the model, substrate uptake reactions, glycolysis, pentose phosphate pathway, the TCA cycle with its glyoxylate bypass and the excretion of by-products (acetate, formate, lactate and ethanol) were included. Previous networks[2] were extended by—among others—the anaplerotic reactions via malic enzyme and pyruvate oxidase as well as by parallel pathways for initial acetate metabolism. The anabolic part of the model covers the conversion of precursors into building blocks like amino acids, to macromolecules and, by assuming a fixed macromolecular composition of the cell, to biomass.

Computation of elementary modes

Elementary flux modes, e_i, as specific vectors r of reaction rates were determined using: (1) the steady-state condition $Sr = 0$ (S is the stoichiometric matrix); and (2) the non-decomposability property that there must not be any steady-state flux vector r that has zero components wherever e_i does and at least one additional zero component[7,11]. Computation of elementary modes was performed using the FluxAnalyzer[28], a program with graphical user interface for the analysis of metabolic networks based on Matlab (Mathworks, Inc.). The software is available upon request from S.K. In particular, a core algorithm described previously[8] was implemented and optimized by additional preprocessing steps. Unless stated otherwise, flux modes were computed separately for each substrate before being combined.

Robustness analysis

Maximal biomass yield $Y^{\max}$ was defined as the optimum of $Y_{i,X/S_k} = e_i^{\mu}/e_i^{S_k}$. The superscripts to e specify single reaction rates—here, growth and substrate uptake—in elementary flux mode i selected for utilization of substrate S_k. Analysis of network connectivity was performed as described in ref. 5. Results were checked for consistency with topological characteristics of large-scale networks: the network studied here is scale-free, that is, the probability $P(k)$ for a substance to participate in k reactions decays according to $P(k) \approx k^{-\gamma}$, with $\gamma = 1.4$ (Supplementary Information). The diameter of the

entire network ($D = 2.9$) corresponds well to network diameters previously reported[5,29]. We assessed the effect of random mutations on viability by independently determining the number of flux modes, the maximal growth yield, and the network diameter after deletion of each single reaction for growth on glucose, acetate, glycerol or succinate, respectively.

Prediction of gene expression

To calculate control-effective fluxes for each reaction l, we determine the efficiency of any e_i by relating the system's output Ω to the substrate uptake and to the sum of all absolute fluxes. With flux modes normalized by the total substrate uptake, efficiencies $\varepsilon_i(S_k, \Omega)$ for the targets for optimization Ω—growth and ATP generation—are defined as:

$$\varepsilon_i(S_k,\mu) = \frac{e_i^{\mu}}{\sum_l |e_i^l|} \quad \text{and} \quad \varepsilon_i(S_k,\mathrm{ATP}) = \frac{e_i^{\mathrm{ATP}}}{\sum_l |e_i^l|} \tag{1}$$

Control-effective fluxes $v_l(S_k)$ are obtained by averaged weighting of the product of reaction-specific fluxes and mode-specific efficiencies over the inventory of elementary modes using the substrate under consideration:

$$v_l(S_k) = \frac{1}{Y^{\max}_{X/S_k}} \cdot \frac{\sum_i \varepsilon_i(S_k,\mu)|e_i^l|}{\sum_i \varepsilon_i(S_k,\mu)} + \frac{1}{Y^{\max}_{A/S_k}} \cdot \frac{\sum_i \varepsilon_i(S_k,\mathrm{ATP})|e_i^l|}{\sum_i \varepsilon_i(S_k,\mathrm{ATP})} \tag{2}$$

Here, $Y^{\max}_{X/S_k}$ and $Y^{\max}_{A/S_k}$ denote optimal yields for biomass production and for ATP generation for cellular maintenance, respectively (experimentally determined yield parameters can, however, easily be incorporated into the approach). Subscripts specify elementary-mode number (i) and reaction number (l). Owing to the normalization of modes, the effect of system input (substrate requirement) is implied in equation (2). Theoretical transcript ratios for growth on two alternative substrates S_1 and S_2 are calculated as

$$\Theta_l(S_1,S_2) = \frac{v_l(S_1)}{v_l(S_2)} \tag{3}$$

Received 7 May; accepted 20 September 2002; doi:10.1038/nature01166.

1. Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. Pathway alignment: Application to the comparative analysis of glycolytic enzymes. *Biochem. J.* **343,** 115–124 (1999).
2. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic phenotype: its definition, characteristics and capabilities. *Proc. Natl Acad. Sci. USA* **97,** 5528–5533 (2000).
3. Cornish-Bowden, A. & Cardenas, M. L. Complex networks of interactions connect genes to phenotypes. *Trends Biochem. Sci.* **26,** 463–465 (2001).
4. Kitano, H. Systems Biology: A brief overview. *Science* **295,** 1662–1664 (2002).
5. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
6. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19,** 125–130 (2001).
7. Schuster, S., Dandekar, T. & Fell, D. A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17,** 53–60 (1999).
8. Schuster, S., Fell, D. A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol.* **18,** 326–332 (2000).
9. Van Dien, S. J. & Lidstrom, M. E. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C_3 and C_4 metabolism. *Biotechnol. Bioeng.* **78,** 296–312 (2002).
10. Oh, M.-K. & Liao, J. C. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol. Prog.* **16,** 278–286 (2000).
11. Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Chapman & Hall, New York, 1996).
12. Canonaco, F. *et al.* Metabolic flux response to phosphoglucose isomerase knock-out in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiol. Lett.* **204,** 247–252 (2001).
13. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387,** 913–917 (1997).
14. von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. The segment polarity network is a robust developmental module. *Nature* **406,** 188–192 (2000).
15. Stelling, J. & Gilles, E. D. in *Proc. 2nd Intl Conf. Syst. Biol.* (eds Yi, T. M., Hucka, M., Morohashi, M. & Kitano, H.) 181–190 (Omnipress, Madison, WI, 2001).
16. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295,** 1664–1669 (2002).
17. Strogatz, S. H. Exploring complex networks. *Nature* **410,** 268–276 (2001).
18. ter Kuile, B. H. & Westerhoff, H. V. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* **500,** 169–171 (2001).
19. Oh, M.-K., Rohlin, L., Kao, K. C. & Liao, J. C. Global expression profiling of acetate grown *Escherichia coli*. *J. Biol. Chem.* **277,** 13175–13183 (2002).
20. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292,** 504–507 (2001).
21. Cleveland, W. S. *Visualizing Data* (AT&T Bell Laboratories, Murray Hill, NJ, 1993).
22. Covert, M. W. & Palsson, B. O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277,** 28058–28064 (2002).
23. Stelling, J., Kremling, A., Ginkel, M., Bettenbrock, K. & Gilles, E. D. *Foundations of Systems Biology* (ed. Kitano, H.) 189–212 (MIT Press, Cambridge, MA, 2001).
24. Schilling, C. H., Letscher, D. & Palsson, B. O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* **203,** 229–248 (2000).
25. Marcotte, E. M. *et al.* A combined algorithm for genome-wide prediction of protein function. *Nature* **402,** 83–86 (1999).
26. Holter, N. S. *et al.* Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc. Natl Acad. Sci. USA* **97,** 8409–8414 (2000).
27. Ge, H., Liu, Z., Church, G. M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genet.* **29,** 482–486 (2001).
28. Klamt, S., Stelling, J., Ginkel, M. & Gilles, E. D. FluxAnalyzer: Exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* (in the press).
29. Bilke, S. & Peterson, C. Topological properties of citation and metabolic networks. *Phys. Rev. E* **64,** 036106 (2001).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements We thank M. Ginkel for software optimization, J. Liao for providing us with data before publication, U. Sauer, S. Bonhoeffer and A. Cornish-Bowden for critical reading of the manuscript and suggestions. S.S. gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (e-mail: stelling@mpi-magdeburg.mpg.de).

The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry

Haining Zhong*, Laurie L. Molday† & Robert S. Molday† King-Wai Yau*‡§

** Departments of Neuroscience and ‡ Ophthalmology, and § Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*
† Department of Biochemistry and Molecular Biology, and Department of Ophthalmology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

Cyclic nucleotide-gated (CNG) channels are crucial for visual and olfactory transductions[1–4]. These channels are tetramers and in their native forms are composed of A and B subunits[5], with a stoichiometry thought to be 2A:2B (refs 6, 7). Here we report the identification of a leucine-zipper[8]-homology domain named CLZ (for carboxy-terminal leucine zipper). This domain is present in the distal C terminus of CNG channel A subunits but is absent from B subunits, and mediates an inter-subunit interaction. With cross-linking, non-denaturing gel electrophoresis and analytical centrifugation, this CLZ domain was found to mediate a trimeric interaction. In addition, a mutant cone CNG channel A subunit with its CLZ domain replaced by a generic trimeric leucine zipper produced channels that behaved much like the wild type, but less so if replaced by a dimeric or tetrameric leucine zipper. This A-subunit-only, trimeric interaction suggests that heteromeric CNG channels actually adopt a 3A:1B stoichiometry. Biochemical analysis of the purified bovine rod CNG channel confirmed this conclusion. This revised stoichiometry provides a new foundation for understanding the structure and function of the CNG channel family.

Distinct A subunits (or α-subunits) of vertebrate CNG channels, named CNGA1–4 according to the latest nomenclature[5], mediate rod phototransduction (A1), cone phototransduction (A3) and olfactory transduction (A2 and A4). Two B subunits (or β-subunits) are known, named CNGB1 and CNGB3 (CNGB2 does not exist[5]), with CNGB1 being part of native rod and olfactory channels and CNGB3 part of cone channels. When expressed heterologously, most A subunits (CNGA1–3) form functional homomeric chan-