IRis

Northeastern University

Physics Faculty Publications

Department of Physics

February 01, 2008

Predicting synthetic rescues in metabolic networks

Adilson E. Motter
Northwestern University

Natali Gulbahce
Los Alamos National Laboratory, Northeastern University, and Dana Farber Cancer Institute

Eivind Almaas
Lawrence Livermore National Laboratory

Albert-László Barabási
University of Notre Dame and Northeastern University

Recommended Citation

Motter, Adilson E.; Gulbahce, Natali; Almaas, Eivind; and Barabási, Albert-László, "Predicting synthetic rescues in metabolic networks" (2008). *Physics Faculty Publications.* Paper 108. http://hdl.handle.net/2047/d20000681

This work is available open access, hosted by Northeastern University.

Molecular Systems Biology 4; Article number 168; doi:10.1038/msb.2008.1
Citation: *Molecular Systems Biology* 4:168
© 2008 EMBO and Nature Publishing Group All rights reserved 1744-4292/08
www.molecularsystemsbiology.com

molecular systems biology

Predicting synthetic rescues in metabolic networks

Adilson E Motter[1,*], Natali Gulbahce[2,3,4], Eivind Almaas[5] and Albert-László Barabási[6,7]

[1] Department of Physics and Astronomy and Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA, [2] Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA, [3] Department of Physics and Center for Complex Network Research, Northeastern University, Boston, MA, USA, [4] Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, MA, USA, [5] Network Biology and Microbial Systems Group, Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA, USA, [6] Departments of Physics and Computer Science and Center for Complex Network Research, University of Notre Dame, Notre Dame, IN, USA and [7] Departments of Physics, Biology and Computer Science and Center for Complex Network Research, Northeastern University, Boston, MA, USA
* Corresponding author. Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.
Tel.: +1 847 491 4611; Fax: +1 847 491 9982; E-mail: motter@northwestern.edu

Received 18.9.07; accepted 16.12.07

An important goal of medical research is to develop methods to recover the loss of cellular function due to mutations and other defects. Many approaches based on gene therapy aim to repair the defective gene or to insert genes with compensatory function. Here, we propose an alternative, network-based strategy that aims to restore biological function by forcing the cell to either bypass the functions affected by the defective gene, or to compensate for the lost function. Focusing on the metabolism of single-cell organisms, we computationally study mutants that lack an essential enzyme, and thus are unable to grow or have a significantly reduced growth rate. We show that several of these mutants can be turned into viable organisms through additional gene deletions that restore their growth rate. In a rather counterintuitive fashion, this is achieved via additional damage to the metabolic network. Using flux balance-based approaches, we identify a number of synthetically viable gene pairs, in which the removal of one enzyme-encoding gene results in a non-viable phenotype, while the deletion of a second enzyme-encoding gene rescues the organism. The systematic network-based identification of compensatory rescue effects may open new avenues for genetic interventions.
Molecular Systems Biology 12 February 2008; doi:10.1038/msb.2008.1
Subject Categories: cellular metabolism; computational methods
Keywords: complex networks; genetic interactions; genetic recovery; metabolism; systems biology

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. Creation of derivative works is permitted but the resulting work may be distributed only under the same or similar licence to this one. This licence does not permit commercial exploitation without specific permission.

Introduction

Recent advances in systems and network biology indicate that specific cellular functions are rarely carried out by single genes, but rather by groups of cellular components, including genes, proteins, and metabolites (Elena and Lenski, 1997; Hartwell *et al*, 1999; Vogelstein *et al*, 2000; Barabási and Oltvai, 2004; Bonhoeffer *et al*, 2004; Albert, 2005; Segrè *et al*, 2005). Such a network-based view changes the way we think about the impact of mutations and other genetic defects: the damage caused by a malfunctioning protein or gene is often not localized, but spreads through the cellular network, leading to a loss of cellular function by incapacitating one or several functional modules (Barabási, 2007; Goh *et al*, 2007). The increasingly sophisticated experimental tools that help us systematically map various cellular interactions offer hope that in the future we will be able to focus not only on the individual components but also monitor and explore the global changes in the cellular network induced by the defective gene or protein. Such network-based approaches indicate that the loss of proteins involved in a large number of protein–protein interactions often results in the death of the organism, a finding that may be useful for the design of antibiotics or cancer drugs. Yet for most genetic diseases, particularly those caused by germ-line mutations, the goal is not to kill the cell, but to recover the lost cellular function or limit the existing damage. This raises an important question: can we develop network-based strategies to predict how to recover function that may have been lost due to defective genes?

In single-cell organisms, the frequently observed reduction in an organism's growth rate following a gene deletion often represents only a transient effect, reflecting the fact that the metabolic network of the mutant operates in a suboptimal regime until appropriate regulatory changes and mutations accumulate to bring the metabolic system to a new optimal steady state (Fong and Palsson, 2004; Herring *et al*, 2006). Experiments in fixed nutrient environments show that after many generations mutants typically increase their growth rate,

converging through adaptation to a new optimal value predicted by flux balance analysis (FBA) (Edwards and Palsson, 2000). If the growth rate in this optimal state is zero, then the organism cannot grow, indicating that the deleted gene is essential. We will refer to these genes as *optimally essential* (Box 1).

Often experiments observe no growth for mutants missing a metabolic enzyme that are predicted to be viable by FBA, prompting us to classify the deleted gene as essential. One potential explanation for the observed discrepancy is that the gene may have an unknown function, regulatory or other, whose absence inhibits growth. Yet for some enzymes an equally compelling explanation is the following: an important challenge of each mutant is to reproduce until the evolutionary tuning of its regulatory system approaches the new optimal growth state. Thus, even if FBA predicts a non-zero optimal growth rate, some mutants may not survive due to their inability to grow in the suboptimal state right after the gene deletion (Figure 1A, red line). The growth rate of the organisms shortly after a gene deletion can be effectively calculated using the minimization of metabolic flux adjustment (MOMA) method (Segrè *et al*, 2002), a variant of FBA. In the following, we will call a gene *suboptimally essential* if the optimal (FBA predicted) growth rate is non-zero in its absence, while the MOMA-predicted growth rate is zero. Therefore, experiments will probably classify organisms missing a suboptimally essential gene as unable to grow. However, in contrast with the optimally essential genes, a mutant missing a suboptimally essential gene would be determined as viable if its metabolism and regulatory system had the chance to re-adjust to its environment.

Here, we show that the growth rate of an organism lacking a suboptimally essential gene may be restored via the removal of other enzyme-encoding genes. We will refer to this as the *Lazarus effect*, as it restores the growth of mutants initially classified as non-viable by experiments since they displayed zero growth rates. We also discuss *suboptimal recovery*, a weaker manifestation of the proposed mechanism, which forces viable mutants to increase their growth rate following additional gene deletions. Our approach is inspired by a method proposed in Motter (2004) to control cascading failures in complex networks and by microbial optimization methods for the targeted production of metabolites (Burgard *et al*, 2003; Pharkya and Maranas, 2006).

Box 1 Summary of new terminology and effects associated with the identification of genetic rescue interactions

Terminology	Definition	Computational method
Synthetically viable gene pair[a]	Removal of one gene is lethal but deletion of a second gene rescues the cell	MOMA and FBA
Optimally essential gene[b]	Gene deletion leads to zero growth rate in growth-maximizing states	FBA
Suboptimally essential gene[c]	Gene deletion leads to zero growth rate but growth is possible in optimal states	MOMA and FBA
Lazarus effect	Gene deletion restores the growth of otherwise non-viable mutants	MOMA and FBA
Suboptimal recovery	Gene deletion increases the growth of already growing strains	MOMA and FBA

[a]Synthetically viable gene *sets* are defined analogously for interactions involving more genes.
[b]These genes are *essential* for growth regardless of the state of the other genes.
[c]The deletion of these genes is *lethal* but the genes themselves are not *essential*.

Results

The principle underlying the proposed rescue effect is illustrated schematically in Figure 1B–E. Consider the situation where, in the wild-type organism, the optimal growth state corresponds to the utilization of the $M_1 \rightarrow M_2 \rightarrow M_4$ pathway, i.e. the flux of reactions involving the M_3 metabolite is either zero or close to zero. In the early state after the deletion of the enzyme catalyzing the $M_2 \rightarrow M_4$ reaction, metabolism operates suboptimally (Figure 1C) by minimizing the necessary flux rearrangement compared to the optimal wild-type flux state (Figure 1B). The optimal postdeletion state, however, requires more drastic flux reorganization, sending most of the flux through the $M_1 \rightarrow M_3 \rightarrow M_4$ pathway (Figure 1D). It would take additional regulatory and metabolic adjustments to reach this new optimal state. This process can be facilitated by deleting the enzyme catalyzing the $M_1 \rightarrow M_2$ reaction, forcing the cell to use the optimal $M_1 \rightarrow M_3 \rightarrow M_4$ pathway (Figure 1E). Therefore, by suppressing stoichiometrically inefficient pathways, we can force the cell to enhance the activity of a more efficient set of reactions, resulting in an increased growth rate. Our goal is to show that such additional deletions, whose role is to enhance the activity of the most efficient pathways, can be predicted by systematically comparing the suboptimal and the optimal fluxes under the same conditions.

To implement the approach described in Figure 1, we developed an algorithm to identify rescue deletions for all mutants missing an enzyme-encoding gene. For this, we use MOMA to determine the suboptimal fluxes $\boldsymbol{\nu}^1_{\mathrm{MOMA}}$ characterizing the mutant shortly after a gene deletion (Figure 1C) and FBA to predict a flux state $\boldsymbol{\nu}^1_{\mathrm{FBA}}$ compatible with optimal growth for the mutant (Figure 1D). If the mutant's metabolism operates suboptimally after the gene deletion, the FBA-predicted growth rate for the mutant is larger than the MOMA-predicted growth rate, and thus, we have a chance to intervene and increase the suboptimal growth rate. In this case, based on the difference in flux pattern between $\boldsymbol{\nu}^1_{\mathrm{FBA}}$ and $\boldsymbol{\nu}^1_{\mathrm{MOMA}}$ (see Materials and methods), we test a set of secondary rescue gene deletions that aim to reduce the difference between the suboptimal and the optimal growth rates (Figure 1E) by using MOMA to determine the new metabolic flux state $\boldsymbol{\nu}^2_{\mathrm{MOMA}}$. If appropriate rescue gene deletions are identified, the obtained growth rate G^2_{MOMA} is higher than the growth rate of the original mutant, concluding our procedure. If the MOMA-predicted growth rate for the original mutant is zero, the rescue deletions can bring along the *Lazarus effect*, inducing a non-zero growth rate; if it is non-zero, the rescue deletions may induce a *suboptimal recovery*, increasing the mutant's growth rate toward its optimal FBA-predicted value.

© 2008 EMBO and Nature Publishing Group

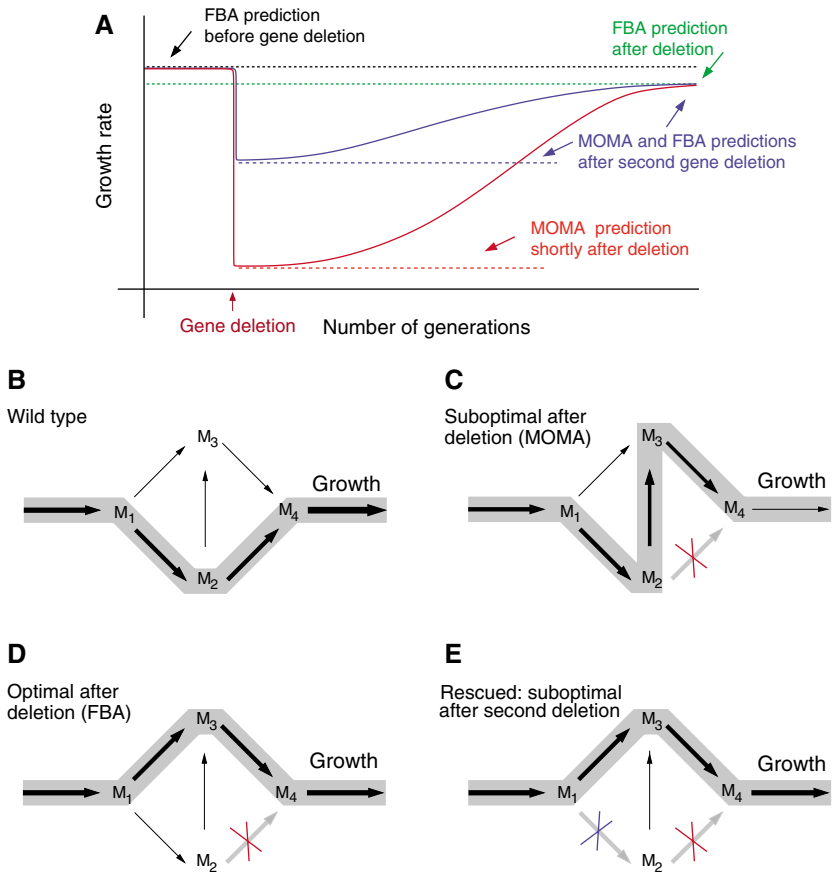

Figure 1 Schematic illustration of the consequences of gene deletion on the organism's growth rate. (**A**) The growth rate following the deletion of an enzyme-encoding gene often drops, but after many generations may recover to a new optimal value not very different from the original one (red line). The optimal growth rate before and after the deletion is predicted by FBA (black and green dotted lines). The blue line indicates the predicted buffering effect of additional gene deletions: by deleting appropriately selected additional genes, the suboptimal growth rate shortly after gene deletions is higher than without the rescue deletions. (**B**–**E**) The effect of rescue deletions on the fluxes of a metabolic network, where M_1 … M_4 represent metabolites and the width of the arrows represents the strength of individual fluxes.

Note that the identified rescue deletions do not change the optimal growth rate, but affect only the suboptimal growth rate (see Materials and methods). The new terminology related to this recovery mechanism is summarized in Box 1.

We illustrate the proposed procedure in Figure 2 for the TCA cycle of *Escherichia coli* MG1655 fed arabinose as the sole carbon source (see Materials and methods and Supplementary Information). MOMA predicts that the deletion of the *fbaA* gene rearranges the fluxes throughout the whole cycle and inhibits the production of phenylalanine, tyrosine, and L-lysine (dotted reactions in Figure 2B), which represent necessary building blocks of the biomass (cf. Figure 2A). Thus, the suboptimal growth rate of this mutant is zero, a prediction supported by experiments in arabinose media (Fraenkel, 1987). In contrast, FBA indicates that a non-zero growth rate can be achieved by a global rearrangement of the flux states (Figure 2C), resulting in changes in flux magnitudes and directions (e.g. the *sucCD* reaction). Consequently, the organism *could* grow if it could get past its suboptimal state when, soon after the gene deletion, its growth rate is zero. We can force the organism to approach the new optimal state by deleting, for example, the genes *aceA* and *sucAB*, which catalyze reactions that are active in the suboptimal state (Figure 2B) but are not active in the optimal state (Figure 2C). These two rescue deletions will activate the production of *all* biomass components after rerouting the fluxes through the pentose phosphate pathway (Figure 2D), and result in a non-zero growth rate, rescuing the otherwise non-viable mutant.

The growth rate of the *fbaA* mutant can be further enhanced by deleting additional genes that catalyze reactions that are inactive in the optimal state (see Materials and methods). We illustrate this in Figure 3A, which shows the predicted suboptimal growth rate of the *fbaA*-deficient *E. coli* mutant after the concurrent removal of several genes in addition to *aceA* and *sucAB*. While the rescue deletion of *aceA* is sufficient to recover growth, the additional deletion of *sucAB*, *tnaB*, *xapB*, and *prr* further enhances the growth rate, with a large enhancement predicted after the removal of *tnaB*. The biomass production reaches a plateau of about 67% of the wild-type biomass production rate after the deletion of 40 genes. The situation is similar for suboptimal recovery: as we show in Figure 3B for the case of the *nuoA* mutant with glucose as the carbon source, additional gene deletions can increase the growth rate of the mutant, eventually approaching 59% of its wild-type optimal value.

Systematically applying our method to the *E. coli* metabolism in glucose minimal medium, we identified 6 suboptimally essential genes, which represent candidates for

© 2008 EMBO and Nature Publishing Group

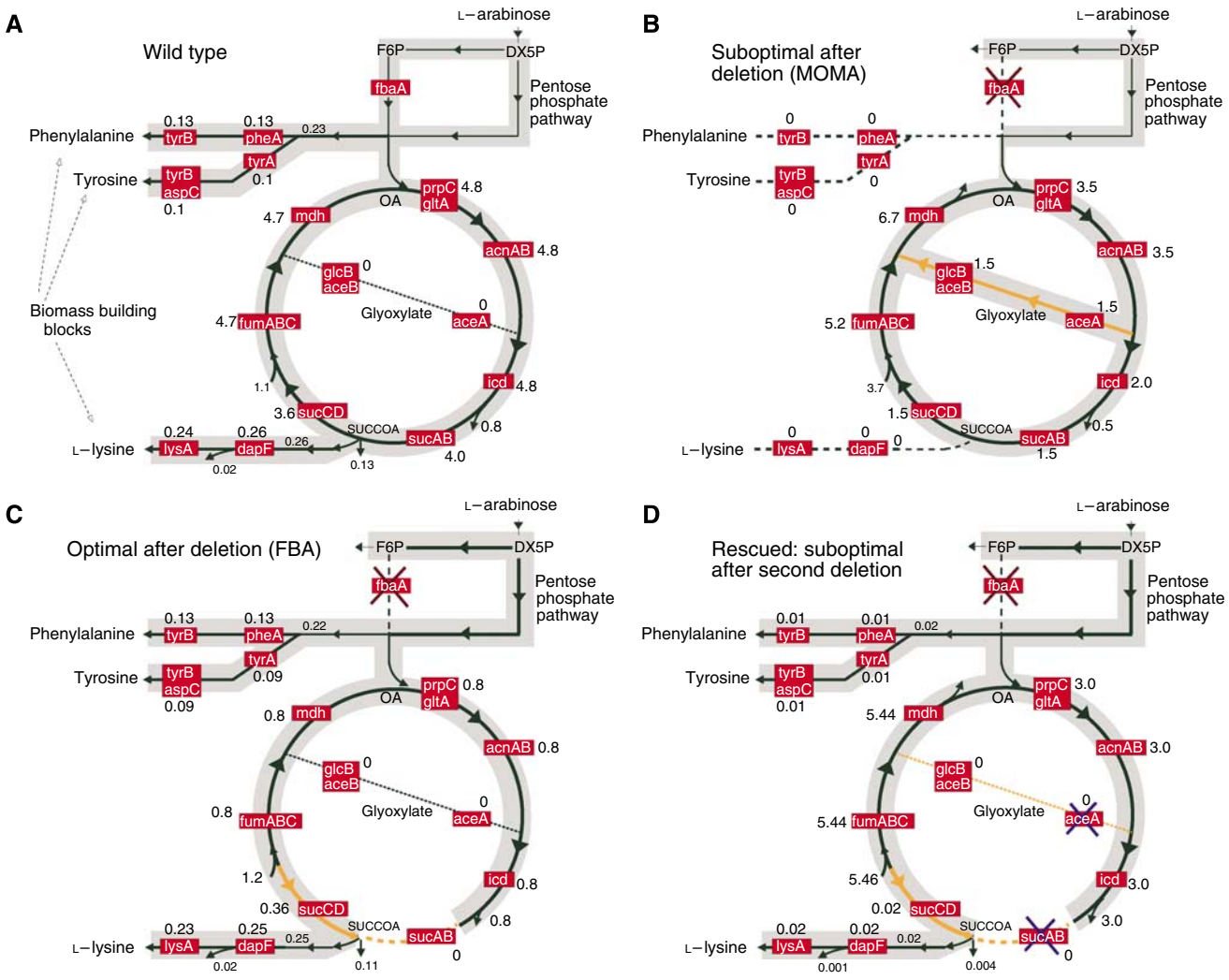

Figure 2 Distribution of metabolic fluxes in the *E. coli* 's TCA cycle in arabinose minimal medium for (**A**) wild-type organism predicted by FBA, (**B**) *fbaA* mutant predicted by MOMA, (**C**) optimal state of *fbaA* mutant predicted by FBA, and (**D**) *fbaA* mutant with the rescue deletions of genes *aceA* and *sucAB*, predicted by MOMA. Key flux changes are highlighted in orange. Note that the metabolic flux pattern predicted by MOMA after the *fbaA* deletion (**B**) is similar to the wild-type fluxes (**A**). With the rescue deletions, however, MOMA-predicted fluxes (**D**) are brought closer to the FBA-predicted fluxes (**C**), restoring the organisms' ability to produce biomass. While we show a double deletion for its pedagogical value, we note that the deletion of *aceA* alone is sufficient to rescue the mutant (see Figure 3A) and that the mutant can also be rescued with other single-gene deletions (see Figure 4B and Supplementary Information).

the Lazarus effect, and 17 candidates for suboptimal recovery (see Figure 4A). Most of the mutants miss genes involved in the central metabolism, while a few miss genes that participate in amino-acid metabolism and transport processes. Of particular interest are mutants with the genes *pfk*, *fbaA*, or *tpiA* deleted, whose essentiality has been tested and is supported by experiments (Fraenkel, 1987). As we show in Supplementary Table SI and Figure 4A, the growth rate of these mutants is restored by additional targeted gene deletions that increase the suboptimal growth rate from zero to more than 45% of the wild-type growth rate.

In Figure 4B we show that, for various media, the increase in the biomass production rate obtained after the deletion of a *single* rescue gene can be more than 10% of the wild-type rate. In other cases, however, we need to simultaneously delete several genes to rescue growth. This is illustrated in Figure 3B, where we show that the growth performance of non-viable *tpiA*-deficient mutants in a glucose medium can be restored only through the concurrent deletion of six genes, *aceA*, *gadA*, *gadB*, *lpdA*, *tynA*, and *gpt*, representing a six-viable set, which is the converse of the *k*-robust set necessary to suppress cellular growth (Deutscher *et al*, 2006). The suboptimal *tpiA* mutant uses the glyoxylate pathway, which is shut down by these rescue deletions. Our prediction, that the glyoxylate pathway is not needed in the optimal state, is supported by a recent experimental observation (Fong *et al*, 2006). This observation indicates that the flux of the glyoxylate pathway in viable but not fully evolved *tpiA* mutants is initially non-zero. However, over the course of a few weeks of adaptive evolution in glucose media, the glyoxylate flux converges to zero (Fong *et al*, 2006). Once the six genes are absent, the concurrent deletion of additional genes can further increase the organism's growth rate (Figure 3B).

© 2008 EMBO and Nature Publishing Group

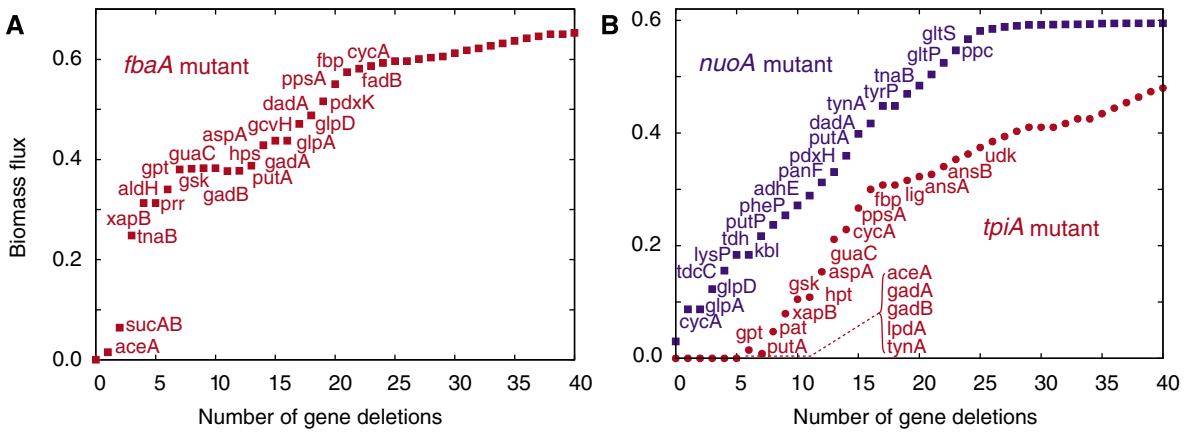

Figure 3 The impact of rescue deletions. (**A**) Predicted biomass production for the *fbaA* mutant of *E. coli* in arabinose minimal medium as a function of the number of rescue deletions when starting with *aceA* and *sucAB*. Deleted rescue genes are indicated in the figure. (**B**) Biomass production of *tpiA*- and *nuoA*-deficient mutants in glucose minimal medium as a function of the number of individual rescue deletions. Deleted genes are indicated in the figure. The optimal biomass flux remains unchanged with the addition of rescue deletions. The biomass fluxes are normalized by the wild-type flux G^{wt}_{FBA}=0.745 mmol/g DW-h in (**A**) and 0.908 mmol/g DW-h in (**B**).

Note that, while the proposed rescue procedure works in all media, the list of mutants that can be rescued by additional deletions as well as the necessary rescue deletions depends on the tested medium. Indeed, we find that the number of *E. coli* mutants whose growth rate increases by more than 10% of the wild-type growth rate after rescue deletions is 8, 21, and 25 in minimal acetate, minimal glucose, and rich media, respectively. Therefore, the rescue effect is more frequent in richer media, where the increased availability of substrates in the environment increases the number of non-essential metabolic genes that can be deleted to improve performance. Furthermore, the proposed rescue mechanism is expected to work for all organisms, allowing us to predict rescue deletions each time an accurate metabolic reconstruction is available. To show this, we determined all single-gene rescues that can recover the growth rate by more than 1% of the wild-type rate in glucose media for deletion mutants of three reconstructed organisms with very different genomes: *Helicobacter pylori* (341 enzyme-encoding genes), *E. coli* (660), and *Saccharomyces cerevisiae* (750). Interestingly, the obtained number of mutant-rescue combinations for these organisms, 58, 94 and 58, respectively, is consistently large and to some extent comparable despite the significant differences in their metabolism.

In our analysis of the most significant cases for the eukaryote *S. cerevisiae*, we predict the Lazarus effect for 3 mutants and suboptimal recovery for 11 other gene deletions in a glucose minimal medium (Figure 4C). It is interesting to note that several of these genes are human orthologs (Steinmetz *et al*, 2002; BiGG, 2007), including genes *pfk*, *tpi1*, *lpd1*, and *mir1*. Of the three mutants predicted to exhibit the Lazarus effect, two of them have been experimentally verified to be non-viable, while positive growth has been observed in experiments for the third one (SGD, 2007). The observed small disagreement with our predictions is probably due to the incompleteness of the reconstructed model or the fact that the organisms in the experiments were not fully adapted to the medium modeled in our computations. As in the case of *E. coli*, the intensity of the recovery generally increases with the number of genes in the rescue set and best recovery may involve up to 50 genes in the examples shown in Figure 4C. However, we predict that a comparable recovery can be obtained with significantly fewer deletions (Supplementary Information). In particular, we also find numerous examples in several media of single-gene rescue deletions resulting in a significant increase of biomass production in *S. cerevisiae* mutants (Figure 4D).

The focus so far in this work has been on developing an approach to computationally predict synthetic rescue in metabolic networks. Importantly, the founding hypothesis of this approach, that the suboptimal growth rate of an organism can be improved by the removal of properly selected genes, is consistent with experiments. To demonstrate this, in Figure 5 we reanalyze experimental results (Fong and Palsson, 2004; Fischer and Sauer, 2005) for the growth rate of several mutants in their suboptimal state, before and after a gene deletion. The compiled data in Figure 5A indicate that the suboptimal growth rates of *E. coli* MG1655 can indeed improve considerably after the deletion of selected enzyme-encoding genes, an effect observed in multiple environments. In Figure 5B, we show similar results for *B. subtilis* 168, following the removal of genes involved in various cellular functions. Note that the metabolism of the wild-type strains in these experiments is not fully adapted to the media and operates in a suboptimal regime. This experimental evidence, together with the power of FBA (Edwards *et al*, 2001; Ibarra *et al*, 2002) and MOMA (Segrè *et al*, 2002; Shlomi *et al*, 2005) to predict the optimal and suboptimal growth rates of an organism in agreement with experimental data, supports our hypothesis that properly selected gene deletions can improve the growth rate of an organism that has not yet adapted to its environment. To further substantiate this claim, we calculated the reaction fluxes determined by experimental uptake and growth rates (Fong and Palsson, 2004) as well as the corresponding optimal reaction fluxes. We used these flux distributions to test our assumption that gene deletions that increase (not increasing) growth tend to be associated with reactions whose fluxes are much larger (smaller) than the optimal fluxes. As shown in

© 2008 EMBO and Nature Publishing Group

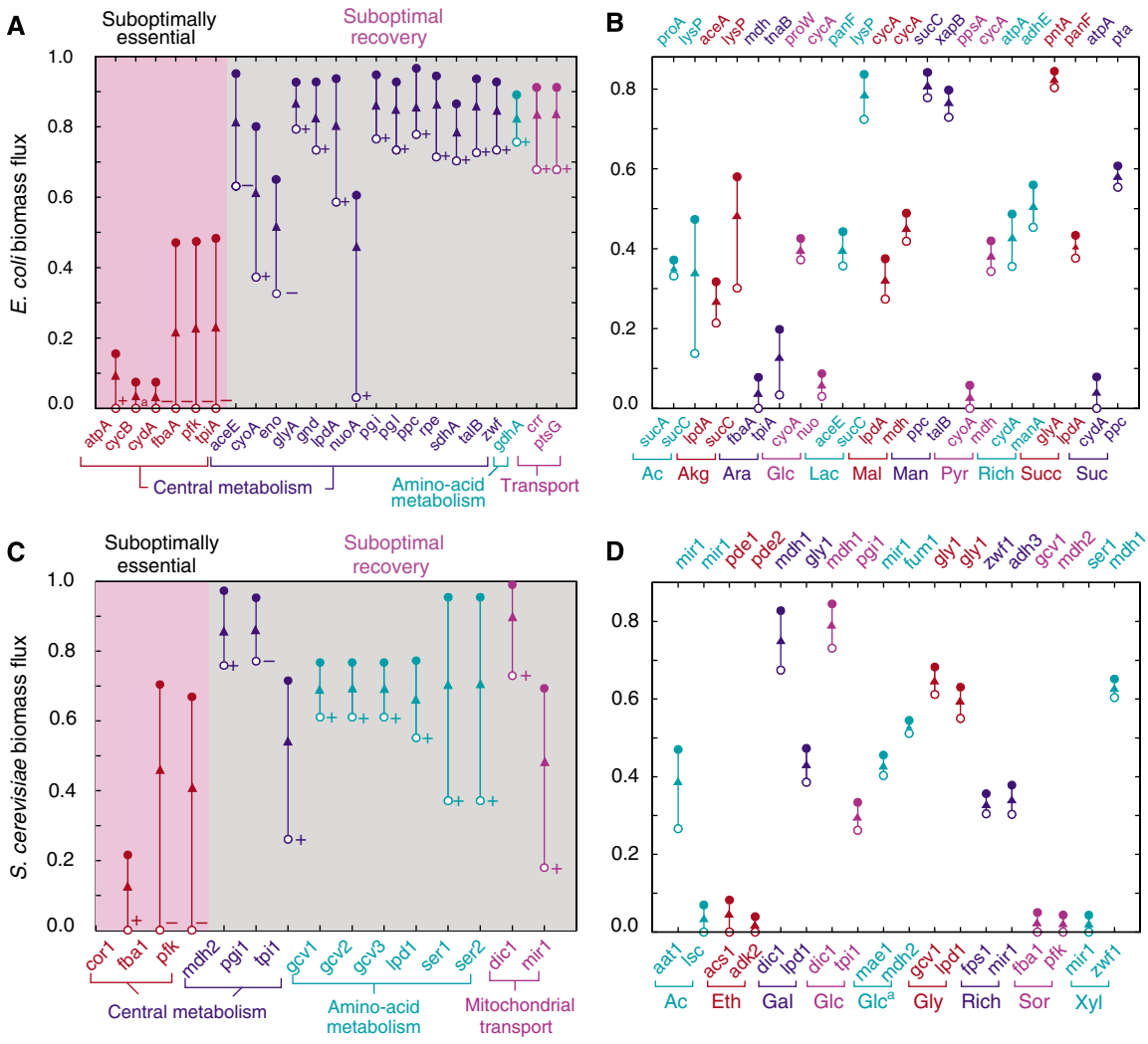

Figure 4 The impact of rescue deletions for *E. coli* (**A**, **B**) and *S. cerevisiae* (**C**, **D**) gene-deficient mutants. (**A**, **C**) Predicted biomass production before (○) and after (●) rescue deletions in glucose minimal media. The mutants are generated through the deletion of the genes shown at the *x*-axis. We show the results for all mutants with $G^1_{\mathrm{MOMA}} < G^1_{\mathrm{FBA}}$ such that $G^1_{\mathrm{MOMA}} \leq 0.8\ G^{\mathrm{wt}}_{\mathrm{FBA}}$ and $G^1_{\mathrm{FBA}} \geq 0.2\ G^{\mathrm{wt}}_{\mathrm{FBA}}$. If the rescue deletion changes the growth rate from zero to some positive value, we observe the Lazarus effect, applying to suboptimally essential genes (left). If the rescue deletion only enhances the growth rate, we observe a suboptimal recovery (right). The experimental information on the lethality of the original *E. coli* (Edwards and Palsson, 2000; Gerdes *et al*, 2003; Baba *et al*, 2006; PEC, 2007) and *S. cerevisiae* (Giaever *et al*, 2002; Steinmetz *et al*, 2002; SDG, 2007) gene-deficient mutants is indicated with (+) for viable mutants, (−) for non-viable mutants, and ([a]) for a gene absent in the databases. (**B**, **D**) Same as in (**A**, **C**) for single-gene rescue deletions in various media. We show selected mutants with significant biomass improvements after the rescue deletion of a single gene. The rescue deletion is indicated at the top, and the tested media are indicated at the bottom. The abbreviations stand for acetate (Ac), α-ketoglutarate (Akg), arabinose (Ara), ethanol (Eth), galactose (Gal), glucose (Glc), glucose anaerobic (Glc[a]), glycerol (Gly), lactate (Lac), malate (Mal), mannose (Man), pyruvate (Pyr), rich medium (Rich) (see Supplementary Information), sorbitol (Sor), succinate (Succ), sucrose (Suc), and xylose (Xyl). The biomass fluxes are normalized by the wild-type flux $G^{\mathrm{wt}}_{\mathrm{FBA}}$ in all panels. In units of mmol/g DW-h, the wild-type fluxes for *E. coli* are 0.187 (Ac), 0.535 (Akg), 0.745 (Ara), 0.908 (Glc), 0.367 (Lac), 0.388 (Mal), 0.908 (Man), 0.303 (Pyr), 2.87 (Rich), 0.418 (Succ), and 1.37 (Suc), while for *S. cerevisiae* they are 0.189 (Ac), 0.311 (Eth), 0.703 (Gal), 0.819 (Glc), 0.180 (Glc[a]), 0.532 (Gly), 1.34 (Rich), 0.798 (Sor), and 0.742 (Xyl). All the genes involved in the rescues of (**A**, **C**) are listed in Supplementary Information, while the minimum rescue sets are listed in Supplementary Tables SII and SIII, respectively. The alternative rescue genes for each media in (**B**, **D**) are listed along with the corresponding recoveries in Supplementary Information.

Supplementary Table SIV, a total of 20 out of 22 *E. coli* mutants analyzed are correctly predicted with this assumption, in support of the proposed rescue mechanism.

Discussion

The mechanism behind the rescue effect introduced above does not depend on the specific details of MOMA or FBA; in fact, any computational or experimental methodology that can help us estimate the metabolic fluxes can be used to identify candidates for rescue deletions. For example, one could use ^{13}C-tracer techniques (Sauer, 2004) to experimentally determine the reaction fluxes of the suboptimal gene-deficient strain and an optimal, or close to optimal, version of the same strain. Candidates for rescue deletions typically correspond to genes catalyzing reactions that are active in the suboptimal state but inactive in the optimal state. By identifying these reactions experimentally, one could minimize biases due to inaccurate modeling in the identification of the candidate

© 2008 EMBO and Nature Publishing Group

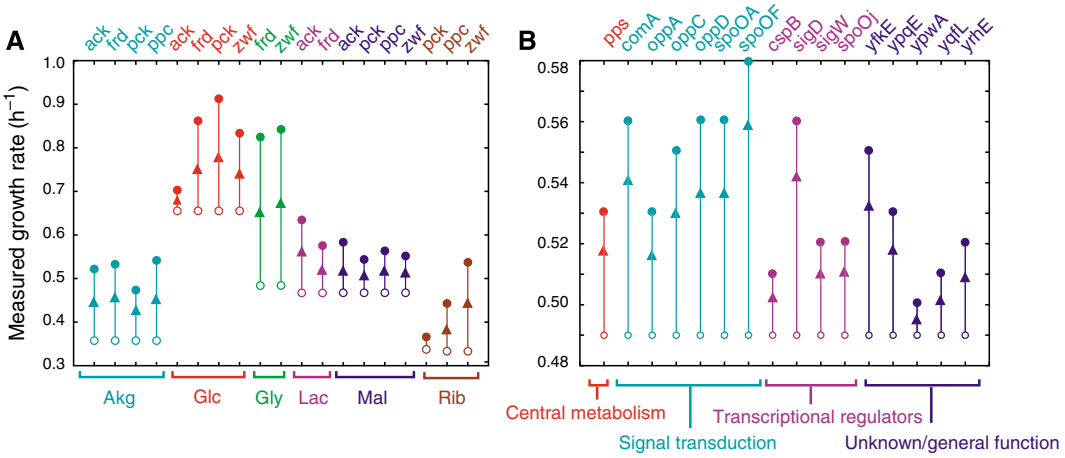

Figure 5 Experimental evidence that gene deletions can enhance suboptimal growth rates: growth rate before (○) and after (●) gene deletions for (**A**) *E. coli* MG1655 (Fong and Palsson, 2004) and (**B**) *B. subtilis* 168 (Fischer and Sauer, 2005). The deleted genes are indicated at the top. All genes in (**A**) are involved in the catalysis of central metabolic reactions, and growth is measured after 10 days in α-ketoglutarate (Akg), glucose (Glc), glycerol (Gly), lactose (Lac), malate (Mal), and ribose (Rib) media. The carbon source in (**B**) is glucose.

rescue deletions. However, we find that our *in silico* predictions are robust to parameter choices and do not rely on the fine tuning of metabolic fluxes or environmental conditions (see Supplementary Information). Furthermore, we predict that the rescue set in an impaired cell is not unique, and the number of rescue combinations that lead to the same effect generally increases with the number of genes in the set (Supplementary Information). These observations corroborate the feasibility of systematic experimental implementation of synthetic rescues. Indeed, the main difficulties expected in verifying our predictions, namely the inaccuracies in matching real genetic and environmental conditions as well as potential side effects of rescue deletions due to, e.g. unknown function, are substantially alleviated by the robustness and flexibility of the rescue interactions. This generality, which transcends particular computational methods, could serve as a bridge to implementations of our approach in multi-cellular organisms, as it facilitates the control of undesirable effects in the recovery of specific cellular functions.

The possibility of rescuing a mutant using additional gene deletions is a general mechanism not limited to metabolism. For example, the removal of *comA* and *sigD* genes enhances the growth rate of *Bacillus subtilis* (see Figure 5), despite the fact that they have no known enzymatic functions. Additionally, it has been observed in *E. coli* that *edd*-deficient mutants grow at a reduced rate and *eda*-deficient mutants do not grow at all in a gluconate medium, while the double *edd*/*eda* mutant is viable. In this case, the mechanism for the rescue effect is different from the one discussed above: upon the deletion of *eda*, the cells accumulate toxic compounds; this accumulation stops when *edd* is also deleted (Fraenkel, 1987). Gene deletion-induced rescue processes have been observed previously in mammalian cells as well. For example, Irs2 knockout mice develop diabetes in 6–8 weeks (Kushner *et al*, 2004; Hahnfeldt and Hlatky, 2005). Yet the additional knockout of Ptp1b partially compensates for the lack of Irs2, doubling the survival time. Similar effects were documented for mutations in HK1.ros, HK1.fas, and HK1.TGK-α in tandem with the loss of *p53* gene (Wang *et al*, 2000; Hahnfeldt and Hlatky, 2005). The mechanisms behind these examples involve mostly local gene–gene interactions, as opposed to the global effect we have systematically unveiled here. They indicate, however, that organisms could be characterized in general by potentially extensive sets of *synthetically viable* double knockouts, representing gene pairs for which the double mutant is viable while one of the single mutants is not. High-throughput techniques, increasingly used to identify synthetically lethal pairs (Tong *et al*, 2001; Ooi *et al*, 2003), could be used to uncover such synthetically viable gene pairs as well. Other techniques may be developed to identify similar interactions between gene *sets*, as proposed above. The results could be used to detect new genetic compensatory mechanisms and would offer a better understanding of cellular functions, just as synthetically lethal gene pairs have deepened our understanding of genetic interactions (Wong *et al*, 2004; Boone *et al*, 2007). Furthermore, our results force us to adjust the current paradigm of gene essentiality: even if the deletion of a gene is lethal, the gene is not necessarily essential to support life (Kobayashi *et al*, 2003; Glass *et al*, 2006; Hashimoto *et al*, 2005; Pál *et al*, 2006) because the organism's ability to metabolize biomass may be restored by additional gene deletions.

Finally, our findings may also offer a new alternative to restore the loss of cellular function caused by specific mutations. Indeed, current approaches based on gene therapy (Ho and Commins, 2001; Kaiser, 2005; Kimmelman, 2005) may trigger abnormal activity associated with the vector and insertion site, such as oncogenesis, or reinforce the activity of pathways encoding malfunctioning products of the faulty gene, such as misfolding proteins. From a drug design and therapy perspective, it may be more advantageous to block the activity of selected pathways rather than trying to restore the activity of a faulty gene or protein. Specific previous experimental studies that can be related to the recovery mechanism reported here corroborate the feasibility of such an approach. It has been observed, for example, that *Myc* deletions rescue the *Apc* deficiency in murine small intestine. This presumably takes place because *Myc* is required for gene activation involved in cancer development often following *Apc*

© 2008 EMBO and Nature Publishing Group

inactivation (Sansom *et al*, 2007). In a different study, the combination of antibiotics exhibiting hyper-antagonistic interactions, where the combined effect of two antibiotics is weaker than at least one alone, has been shown to select against resistant strains (Chait *et al*, 2007). In the context of our work this means that, in the two-drug sublethal medium, the 'deficient' bacterial cells (non-resistant strain) prevail. Another example is found in studies of *E. coli* mutants unable to grow anaerobically on glucose and other hexoses when gene *adh* (ethanol production) or gene *pta* (acetic acid production) is inactivated, but the mutant with both genes deactivated will grow through the production of lactic acid as the major fermentation product (Gupta and Clark, 1989). On the other hand, a non-fermenting mutant of *E. coli*, NZN111, is rescued to ferment glucose through the inactivation of the *ptsG* gene, resulting in the production of succinate, acetate, and ethanol by rerouting fluxes that would go through the partially blocked pathways of pyruvate in NZN111 (Chatterjee *et al*, 2001). In addition, it has been shown that the concurrent deletion of genes *zwf*, *sfcA*, *maeB*, *ndh*, *ldhA*, and *frdA* maximizes the biomass yield in wild-type *E. coli* MG1655 by eliminating the elementary metabolic modes associated with low biomass yield (Trinh *et al*, 2006). The latter study also demonstrates the feasibility of creating mutants with several targeted rescue deletions (Causey *et al*, 2003), in support of our suggestions. These examples are not limited to a single organism and can be interpreted as different manifestations of a common rescue mechanism. Therefore, a combination of experimental and computational studies aimed at systematically uncovering synthetically viable gene pairs and gene sets, as well as the underlying rescue effects, may open new avenues for the next generation of therapeutic strategies.

Materials and methods

Constraint-based approach

For a network with m metabolites and n reactions, the stoichiometric constraints are represented by $\sum_j S_{ij}v_j = 0$, where $S=(S_{ij})$ is the $m \times n$ matrix of stoichiometric coefficients and $\boldsymbol{\nu}=(\nu_j)$ is the vector of fluxes. The individual fluxes are limited by thermodynamic constraints, substrate availability, and the maximum reaction rates supported by the catalyzing enzymes and transporting proteins, as

$$\alpha_j \leqslant v_j \leqslant \beta_j \qquad (1)$$

where $\alpha_j=\beta_j=0$ for uptake reactions of substrates not available in the medium. The biomass production is incorporated as an additional reaction $\sum_i c_i x_i \xrightarrow{G} 1$ biomass unit, where the stoichiometric coefficient c_i corresponds to the experimentally measured biomass composition of metabolite x_i (Edwards and Palsson, 2000). FBA consists of finding a metabolic state that satisfies these constraints while maximizing the biomass flux G. The deletion of genes responsible for the production of the enzymes involved in reaction j corresponds to imposing the bounds $\alpha_j=\beta_j=0$ in Equation (1). MOMA aims to find a solution $\boldsymbol{\nu}_{\mathrm{MOMA}}$, compatible with the constraints imposed to the mutant, while being closest to the original metabolic state $\boldsymbol{\nu}_{\mathrm{FBA}}$ in terms of Euclidean distance in the space of fluxes. Our implementations of FBA and MOMA are based on the optimization softwares GNU Linear Programming Kit (Makhorin, 2001) and Object-Oriented Quadratic Programming Package (Gertz and Wright, 2001), respectively, and have been tested using independent implementations of the CPLEX solver (ILOG CPLEX, Version 10.2.0, www.ilog.com).

Identifying rescue gene deletions

Consider a strain generated by the deletion of a metabolic gene that constrains at least one of the non-zero metabolic fluxes of the wild-type organism and such that the biomass flux after this deletion is $G^1_{\mathrm{MOMA}} < G^1_{\mathrm{FBA}}$. To increase G^1_{MOMA}, we compute the vector of all metabolic fluxes $\boldsymbol{\nu}^1_{\mathrm{FBA}}=(\nu^1_{j\mathrm{FBA}})$ predicted by FBA and use them to define a second gene deletion. This deletion is defined by identifying the minimum number of metabolic genes that deactivate most or all reactions j with $\nu^1_{j\mathrm{FBA}}=0$. These gene deletions force the metabolic system to operate closer to the optimal regime predicted by FBA, while they do not change the FBA fluxes and the predicted steady-state biomass production (i.e. $G^2_{\mathrm{FBA}}=G^1_{\mathrm{FBA}}$). The corresponding changes in the MOMA-predicted fluxes are expected to increase the biomass flux from G^1_{MOMA} to $G^2_{\mathrm{MOMA}} > G^1_{\mathrm{MOMA}}$. We recursively discard the deletions that have no impact on the MOMA-predicted biomass production. The number of gene deletions is further reduced through recursively activating genes from the rescue set that contribute the least to the increase in G^2_{MOMA}. Note that this approach increases the biomass production itself, which is not necessarily related to the biomass yield considered in metabolic engineering studies (Causey *et al*, 2003; Trinh *et al*, 2006).

Algorithmically, we start with a mutant strain defined by the deletion of one or more metabolic genes, and identify the sets of rescue deletions by adhering to the following procedure:

(1) Calculate the FBA optimal flux vector $\boldsymbol{\nu}^{\mathrm{wt}}_{\mathrm{FBA}}$ for the wild-type strain.
(2) Calculate the FBA optimal flux vector $\boldsymbol{\nu}^1_{\mathrm{FBA}}$ for the mutant strain.
(3) Calculate the MOMA flux vector $\boldsymbol{\nu}^1_{\mathrm{MOMA}}$ for the mutant strain, using $\boldsymbol{\nu}^{\mathrm{wt}}_{\mathrm{FBA}}$ as a reference flux vector.
(4) Continue if biomass flux $G^1_{\mathrm{MOMA}} < G^1_{\mathrm{FBA}}$, as a set of rescue deletions may exist.
(5) Identify reaction set K consisting of all reactions j such that $\nu^1_{j\mathrm{MOMA}} \neq 0$ and $\nu^1_{j\mathrm{FBA}}=0$. Potential recovery is implemented by setting $(\alpha_j, \beta_j)=(0, 0)$ for every $j \in K$.
(6) Identify K^* as self-consistent subset of K by obeying the gene–enzyme relationships.
(7) Incrementally reduce K^* by identifying and activating recursively the gene from the rescue set whose deletion contributes the least to the increase of the biomass flux. Gene activation is implemented by restoring the original (α_j, β_j) of the corresponding reactions.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

This study was partially supported by NSF-MRSEC program (DMR-0520513) at the Materials Research Center of Northwestern University (AEM), DOE under contract DE-AC52-06NA25396 (NG), LLNL-LDRD office (06-ERD-061) and DOE under contract W-7405-Eng-48 (EA), and NIH U01 A1070499-01 and 1P20 CA11300-01 (ALB).

References

Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* **118:** 4947–4957

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 8

Barabási A-L (2007) Network medicine—from obesity to the 'diseasome'. *N Engl J Med* **357:** 404–407

Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5:** 101–113

© 2008 EMBO and Nature Publishing Group

BiGG (2007) Database of Biochemically, Genetically and Genomically (BiGG) Structured Genome-Scale Metabolic Network Reconstruction. Available online at: http://bigg.ucsd.edu/

Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ (2004) Evidence for positive epistasis in HIV-1. *Science* **306:** 1547–1550

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. *Nat Rev Genet* **8:** 437–449

Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotech Bioeng* **84:** 647–657

Causey TB, Zhou S, Shanmugam KT, Ingram LO (2003) Engineering the metabolism of *Escherichia coli* W3110 for the conversion of sugar to redox-neutral and oxidized products: homoacetate production. *Proc Natl Acad Sci USA* **100:** 825–832

Chait R, Craney A, Kishony R (2007) Antibiotic interactions that select against resistance. *Nature* **446:** 668–671

Chatterjee R, Millard CS, Champion K, Clark DP, Donnelly MI (2001) Mutation of the *ptsG* gene results in increased production of succinate in fermentation of glucose by *Escherichia coli*. *Appl Environ Microbiol* **67:** 148–154

Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat Genet* **38:** 993–998

Duarte NC, Herrgard MJ, Palsson BO (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **14:** 1298–1309

Edwards JS, Ibarra RU, Palsson BO (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotech* **19:** 125–130

Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97:** 5528–5533

Elena SF, Lenski RE (1997) Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390:** 395–398

Fischer E, Sauer U (2005) Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat Genet* **37:** 636–640

Fong SS, Nanchen A, Palsson BO, Sauer U (2006) Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* **281:** 8024–8033

Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* **36:** 1056–1058

Fraenkel DG (1987) Glycolysis, penthose phosphate pathway, and Entner–Doudoroff pathway. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Neidhardt FC (ed), Vol. 1, pp 142–150. Washington: Am Soc Microbiol

Gerdes SY, Gerdes SY, Scholle MD, Campbell JW, Balázsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabási A-L *et al* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185:** 5673–5684

Gertz EM, Wright SJ (2001) Object-Oriented Software for Quadratic Programming (OOQP). Available online at: www.cs.wisc.edu/~swright/ooqp/

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A *et al* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418:** 387–391

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison III CA, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103:** 425–430

Goh K-I, Cusick JE, Valle D, Childs E, Vidal M, Barabási A-L (2007) The human disease network. *Proc Natl Acad Sci USA* **104:** 86850–86890

Gupta S, Clark PD (1989) *Escherichia coli* derivatives lacking both alcohol dehydrogenase and phosphotransacetylase grow anaerobically by lactate fermentation. *J Bacteriol* **171:** 3650–3655

Hahnfeldt P, Hlatky L (2005) in preparation

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* **402:** C47–C52

Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato J (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol* **55:** 137–149

Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* **38:** 1406–1412

Ho M-W, Commins J (2001) Common gene therapy vector causes cancer as well as toxic shock. ISIS Report, 20 September (The Institute of Science in Society, London)

Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420:** 186–189

Kaiser J (2005) Gene therapy: putting the fingers on gene repair. *Science* **310:** 1894–1896

Kimmelman J (2005) Recent developments in gene transfer: risk and ethics. *BMJ* **330:** 79–82

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Débarbouille M, Dervyn E, Deuerling E *et al* (2003) Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* **100:** 4678–4683

Kushner JA, Haj FG, Klaman LD, Dow MA, Kahn BB, Neel BG, White MF (2004) Islet-sparing effects of protein tyrosine phosphatase-1b deficiency delays onset of diabetes in IRS2 knockout mice. *Diabetes* **53:** 61–66

Makhorin A (2001) GNU Linear Programming Kit (GLPK). Available online at: www.gnu.org/software/glpk/glpk.html

Motter AE (2004) Cascade control and defense in complex networks. *Phys Rev Lett* **93:** 098701

Ooi SL, Shoemaker DD, Boeke JD (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat Genet* **35:** 277–286

Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440:** 667–670

PEC (2007) Profiling of *Escherichia coli* chromosome (PEC) database. Available online at: http://www.shigen.nig.ac.jp/ecoli/pec/

Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* **8:** 1–13

Sansom OJ, Meniel VS, Muncan V, Phesse TJ, Wilkins JA, Reed KR, Vass JK, Athineos D, Clevers H, Clarke AR (2007) *Myc* deletion rescues *Apc* deficiency in the small intestine. *Nature* **446:** 676–679

Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* **15:** 58–63

SDG (2007) *Saccharomyces* Genome Database (SGD). Available online at: http://www.yeastgenome.org/

Segrè D, DeLuna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nat Genet* **37:** 77–83

Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99:** 15112–15117

© 2008 EMBO and Nature Publishing Group

Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* **102:** 7695–7700

Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T, Chu AM, Giaever G, Prokisch H, Oefner PJ, Davis RW (2002) Systematic screen for human disease genes in yeast. *Nat Genet* **31:** 400–404

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364–2368

Trinh CT, Carlson R, Wlaschin A, Srienc F (2006) Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. *Metabol Eng* **8:** 628–638

Vogelstein B, Lane D, Levine AJ (2000) Surfing the *p53* network. *Nature* **408:** 307–310

Wang XJ, Greenhalgh DA, Donehower LA, Roop DR (2000) Cooperation between Ha-ras and fos or transforming growth factor alpha overcomes a paradoxic tumor-inhibitory effect of *p53* loss in transgenic mouse epidermis. *Mol Carcinog* **29:** 67–75

Wong SL, Zhang LV, Tong AHY, Li Z, Goldberg DS, King OD, Lesage G, Vidal M, Andrews B, Bussey H, Boone C, Roth PF (2004) Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci USA* **101:** 15682–15687

SOME RIGHTS RESERVED

Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Licence.

© 2008 EMBO and Nature Publishing Group