Genome-scale gene/reaction essentiality and synthetic lethality analysis

Patrick F Suthers[1], Alireza Zomorrodi[1] and Costas D Maranas*

Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
[1] These authors contributed equally to this work
* Corresponding author. Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA. Tel.: + 1 814 863 9958;
Fax: + 1 814 865 7846; E-mail: costas@psu.edu

Synthetic lethals are to pairs of non-essential genes whose simultaneous deletion prohibits growth. One can extend the concept of synthetic lethality by considering gene groups of increasing size where only the simultaneous elimination of all genes is lethal, whereas individual gene deletions are not. We developed optimization-based procedures for the exhaustive and targeted enumeration of multi-gene (and by extension multi-reaction) lethals for genome-scale metabolic models. Specifically, these approaches are applied to *i*AF1260, the latest model of *Escherichia coli*, leading to the complete identification of all double and triple gene and reaction synthetic lethals as well as the targeted identification of quadruples and some higher-order ones. Graph representations of these synthetic lethals reveal a variety of motifs ranging from hub-like to highly connected subgraphs providing a birds-eye view of the avenues available for redirecting metabolism and uncovering complex patterns of gene utilization and interdependence. The procedure also enables the use of falsely predicted synthetic lethals for metabolic model curation. By analyzing the functional classifications of the genes involved in synthetic lethals, we reveal surprising connections within and across clusters of orthologous group functional classifications.
Molecular Systems Biology **5**: 301; published online 18 August 2009; doi:10.1038/msb.2009.56
Subject Categories: simulation and data analysis; cellular metabolism
Keywords: bilevel optimization; gene essentiality; genome-scale metabolic models; network robustness; synthetic lethality

Introduction

Robustness is an inherent property of metabolic networks enabling living systems to maintain their cellular functions in response to genetic and environmental perturbations (Kim *et al*, 2007). The study of metabolic robustness in response to genetic perturbations is usually associated with the concepts of gene essentiality and lethality alluding to whether an organism can survive single- or multiple-gene deletions. Essential genes consist of genes whose individual deletion is lethal (i.e. no biomass formation) under a specific environmental condition (e.g. glucose minimal medium). By analogy, synthetic lethals (SLs) refer to pairs of non-essential genes whose simultaneous deletion is lethal (Novick *et al*, 1989; Guarente, 1993). Here, we extend the concepts of essentiality and synthetic lethality to reactions that preclude biomass formation upon the elimination of a single or a pair of reactions, respectively.

Synthetic gene lethality can arise for a variety of reasons. For example, two gene protein products can be interchangeable with respect to an essential function (isozymes), act in the same essential pathway (with each mutation decreasing the flux through that pathway), or operate in two separate pathways with redundant or complementary essential functions (Guarente, 1993; Tucker and Fields, 2003; Kaelin, 2005). The study of synthetic lethality plays a pivotal role in elucidating functional associations between genes and gene function predictions (Ooi *et al*, 2006). For example, SL screens have been used to identify new genes involved in morphogenesis (Bender and Pringle, 1991; Wang and Bretscher, 1997), vacuolar protein transport (Chen and Graham, 1998), DNA damage (Mullen *et al*, 2001), spindle migration (Schoner *et al*, 2008) and in many other studies (Kuepfer *et al*, 2005; Ye *et al*, 2005). In the context of human genetics, gene lethality studies have been implicated in cancer therapies and the development

of new pharmaceuticals (Hartman *et al*, 2001; Dolma *et al*, 2003; Kamb, 2003; Kaelin, 2005).

The traditional method for identifying SL interactions relies on mutant screens (Forsburg, 2001); however, in recent years we have witnessed rapid progress in the development of high-throughput SL screens. In one of the first efforts, Tong *et al* (2001) developed a genome-scale method for the construction of double mutants termed synthetic genetic array (SGA) analysis and applied it to the yeast genome. Later, Ooi *et al* (2003) introduced a systematic technique called synthetic lethality analysis by microarray (SLAM), which takes advantage of molecular bar codes to detect lethality. Other efforts in this direction include development of an improved technology called diploid-based synthetic lethality analysis on microarrays (dSLAM) that exploits heterozygous diploid yeast knockouts (YKOs) to detect genome-wide lethality (Pan *et al*, 2004) and more recently a technique termed GIANT-coli for high-throughput generation of double mutants in *Escherichia coli* based on F factor-driven conjugation (Typas *et al*, 2008). Despite these advances in large-scale screening techniques, the comprehensive mapping of all SL pairs remains a labor-intensive task. In particular, for the well-studied genetic system of *Saccharomyces cerevisiae*, even using SGA (Tong *et al*, 2001, 2004) only about 4% of the total estimated interactions under a single-growth condition have been queried. This task becomes even more taxing when considering multiple growth conditions (Wong *et al*, 2004).

The availability of genome-scale metabolic models of organisms has provided the foundation for the development of computational frameworks to rapidly predict the effect of multiple genetic manipulations on the strain growth phenotype under different media. For example, by applying flux balance analysis (FBA) to the *i*FF708 metabolic network model of *S. cerevisiae* (Forster *et al*, 2003), Segre *et al* (2005) calculated the maximal rates of biomass production of all single- and double-gene knockouts in comparison to the wild-type strain to assess the spectrum of epistatic interactions. Plaimas *et al* (2008) proposed using a machine learning strategy to distinguish between essential and non-essential reactions in *E. coli* by characterizing an enzyme based on its local network topology, gene homologies, co-expression and FBA. Although most studies focused so far on *S. cerevisiae* and *E. coli* (Wong *et al*, 2004; Kim *et al*, 2007; Le Meur and Gentleman, 2008) there are nonetheless reports studying the *in silico* lethality for other organisms. For example, the reconstructed metabolic network of *Helicobacter pylori* was used to carry out single- and double-mutation studies based on FBA (Thiele *et al*, 2005). Alternatively, Wunderlich and Mirny (2006) introduced a network topological measure termed synthetic accessibility and showed that just the topology of the metabolic network of both *E. coli* and *S. cerevisiae* is sufficient to predict the viability of knockout strains with an accuracy comparable to FBA. Similarly, other studies explore metabolic network essentiality and lethality using the topological concept of missing alternatives in reaching one or more nodes in the network (Palumbo *et al*, 2005, 2007).

The majority of *in vivo* and *in silico* studies have concentrated on perturbing/deleting a single gene or a gene pair at a time. Thus, these analyses might fail to assess the full range of robustness and functional organization of the metabolic networks afforded by higher-order interactions and redundancies. Extending the concept of lethality for not just pairs but triples, quadruples, etc. can capture multi-gene/reaction interdependencies. The challenge in exhaustively identifying higher-order SLs lies in the combinatorial complexity of the underlying mathematical problem. Efforts towards addressing this challenge include the work of Deutscher *et al* (2006) who conducted an *in silico* multiple knockout investigation of the *i*FF708 (Forster *et al*, 2003) yeast metabolic network. They cataloged gene sets that provide mutual functional backup of up to eight interacting genes. In a subsequent study, Deutscher *et al* (2008) developed a computational approach based on ideas from game theory for multiple knockout analysis in *S. cerevisiae* to elucidate insights into the localization of metabolic functions. Alternatively, Behre *et al* (2008) extended their previous study on single knockouts (Wilhelm *et al*, 2004) by introducing a generalized framework for analyzing structural robustness of metabolic networks based on the concept of elementary flux modes. They applied this framework to metabolic networks describing amino acid metabolism in both *E. coli* and human hepatocytes, and for the central metabolism in human erythrocytes. Yeast is the preferred system for the analysis of genetic interactions (Ooi *et al*, 2006) due to its short non-coding regions, a genome containing <7% introns (Grate and Ares, 2002) and its existence in both haploid and diploid states. Consequently, most research focused on investigating lethality in *S. cerevisiae*, rather than on other model microorganisms such as *E. coli*. All studies using *E. coli* are limited to only a sample of pairwise interactions.

In this paper, we present a comprehensive map of SL gene and reaction pairs for genome-scale models. We move beyond SL pairs to exhaustively identify SL triples and some higher-order interactions among genes or reactions. Limited by the absence of customized algorithms, most existing *in silico* multiple knockout studies use brute-force searches (Deutscher *et al*, 2006) or focus on limited parts of metabolism (Behre *et al*, 2008). We overcome these challenges by introducing a bilevel optimization framework that uses FBA to completely identify all multi-reaction/gene lethals for genome-scale models. This framework is applied to the *i*AF1260 model of *E. coli* K12 (Feist *et al*, 2007) for aerobic growth on minimal glucose medium. We contrast the predicted SLs against experimental data and provide a number of model refinement possibilities. We elucidate all SL gene and reaction triples and also introduce the concept of degree of essentiality to unravel the contribution of each reaction in 'buffering' cellular functionalities. This study provides a complete analysis of gene and reaction essentiality and lethality for the latest *E. coli* model *i*AF1260 and ushers the computational means for performing similar analyses for other genome-scale models. Furthermore, by exhaustively elucidating all model growth predictions in response to multiple gene knockouts, it provides a many-fold increase in the number of genetic perturbations that can be used to assess the performance of *in silico* metabolic models.

Results

SL pairs

By testing the impact of the removal of one gene at a time on the feasibility of biomass formation, we identified a total of 188 essential genes (15% of total metabolic genes) and five essential non-gene associated reactions (out of a total of 155 present in the model) for the *E. coli* *i*AF1260 model (Feist *et al*, 2007) when incorporating the list of reactions suppressed under aerobic glucose medium conditions. These results are in agreement with the list reported by Feist *et al* (2007). The relatively small fraction of genes that are essential alludes to the built-in robustness of *E. coli* metabolism to single-gene deletions, implying that a higher-order gene essentiality analysis is indeed needed to adequately assess metabolic network redundancy. By using the exhaustive enumeration procedure described in the Materials and methods section, we identified 83 genes and four non-gene associated reactions involved in 86 SL pairs (∼0.01% of total possible pairs) as shown in Figure 1. All these SL pairs are next analyzed in detail in terms of their phenotypic, topological and functional impact.

Phenotypic classification

The identified SL pairs are phenotypically classified into two types. The first type includes the ones that yield auxotrophic strains that can be rescued through the supply of missing nutrients (i.e. amino acids or other compounds), whereas the second type includes those that lack essential functionalities that cannot be restored by adding extra components to the growth medium. Of the 86 predicted SL pairs, 53 (∼62%) of them were found to yield auxotrophic strains *in silico* that can be restored through supplementation. For example, an *E. coli* strain lacking both *asnA* (b3744) and *asnB* (b0674) can be rescued *in silico* through the supplementation of the growth medium by asn-L (L-asparagine). Note that the names and abbreviations of all metabolites and reactions follow those in *i*AF1260 (Feist *et al*, 2007). On the other hand, disruption of *modA* (b0763) and *cysA* (b2422) results in a strain that cannot be rescued through the addition of the missing compound mobd (molybdate) as the gene disruptions eliminate MOB-Dabcpp (molybdate periplasm transport through ABC system). As another example, a double-mutant strain lacking the gene pair *folA* (b0048) and *folM* (b1606) is unable to grow on supplemented medium, as it can neither produce nor uptake the precursor metabolite thf (5,6,7,8-tetrahydrofolate).

Topological classification

By representing all genes forming SL pairs as nodes connected by an edge, a variety of different topological motifs emerge (see Figure 1). These include disjoint pairs, stars and highly connected subgraphs. Disjoint pairs are motifs representing
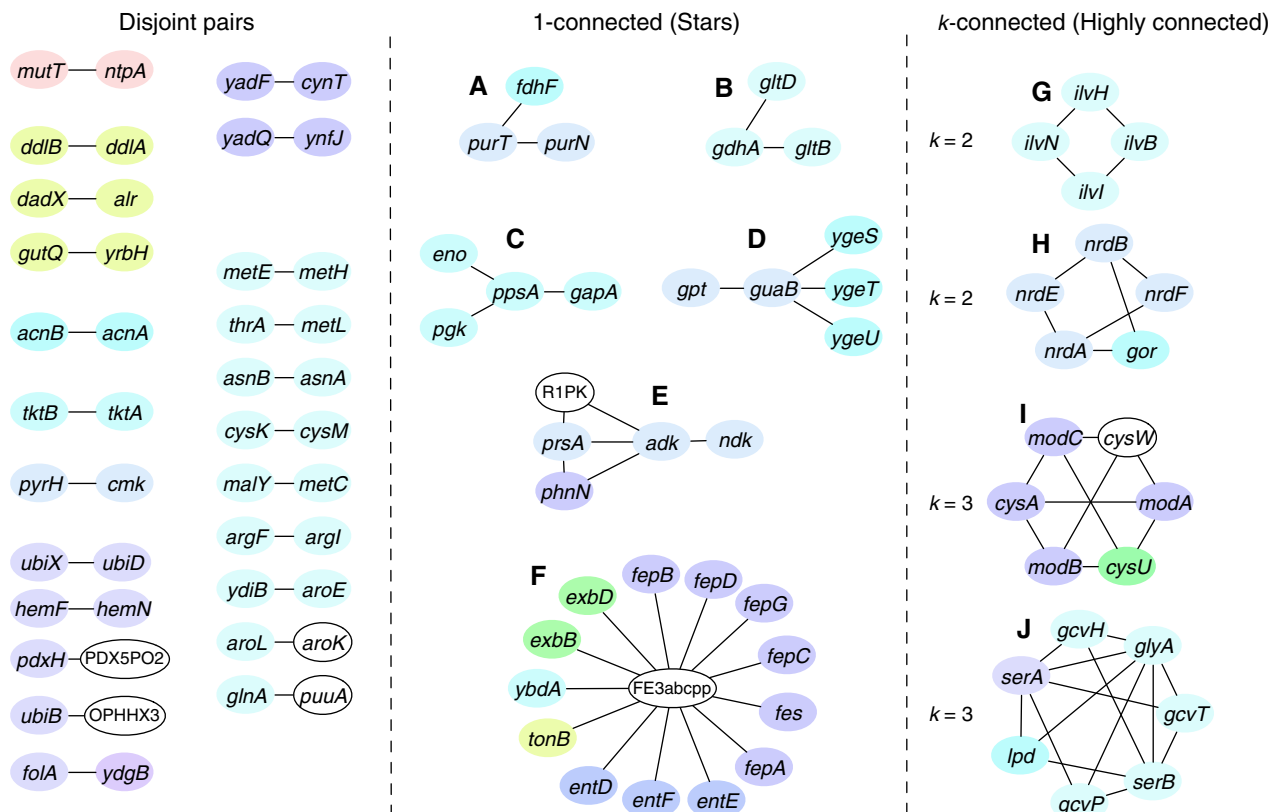

Figure 1 Topological and functional classification of clusters of SL gene pairs. Three types of network motifs are present: disjoint pairs (left); stars, or 1-connected motifs (center); and highly connected subgraphs, or *k*-connected motifs (right). Genes are color-coded in accordance to the COG (Tatusov *et al*, 2003) functional categorization. Names of genes are set in italics and the names of non-gene associated reactions are set in roman. Note that all the reaction abbreviations follow those in *i*AF1260 (Feist *et al*, 2007).

gene pairs that either code for isozymes of an essential reaction or for two separate reactions that form an SL reaction pair. We found a total of nineteen disjoint gene pairs and two disjoint pairs containing non-gene associated reactions (see Figure 1) of which seventeen encode isozyme pairs. For example, both *cysK* (b2414) and *cysM* (b2421) enable the same reaction CYSS (cysteine synthase) that is required for cysteine anabolism.

Star motifs are clusters with a single gene (i.e. hub gene) connected to all other genes. An important implication of these clusters is that unless the hub gene is present, all 'satellite' genes need to be functional for biomass formation feasibility. Star motifs are 1-connected graphs as biomass formation is preserved by simply retaining the functionality of a single gene (i.e. the hub-gene). We identified a total of six star clusters involving 32 genes and two non-gene associated reactions organized in 30 SL pairs. For example, in cluster F (see Figure 1), the hub is a non-gene associated reaction FE3abcpp (Fe(III) transport through the ABC system [periplasm to cytoplasm]). All members of this cluster are directly or indirectly involved with iron III transport from the extra-cellular environment to the periplasm or from the periplasm to the cytoplasm.

Highly connected subgraphs, formally known as *k*-connected motifs with $k > 1$ (Diestel, 2005), describe clusters that unlike star clusters ($k=1$) require the functionality of more than one gene for biomass formation to be feasible. We identified four such *k*-connected clusters that contained a total of 22 genes participating in 33 SL pairs. In all four clusters many multi-protein enzymes/isoenzymes were present. The largest cluster of this type (i.e. J) consists of seven nodes and fourteen edges with genes coding for four reactions involved in serine, glycine and folate metabolism (see Figure 2B). The underlying reasons for the complicated connectivity can be deduced by redrawing this cluster using reactions instead of genes (see Figure 2). As shown in Figure 2A, both GLYCL and GHMT2r form SLs with two other reactions. When this figure is expanded to show the gene–reaction associations (see Figure 2B), the reason for the essentiality connections between the corresponding genes becomes more clearly discernible.

Functional classification

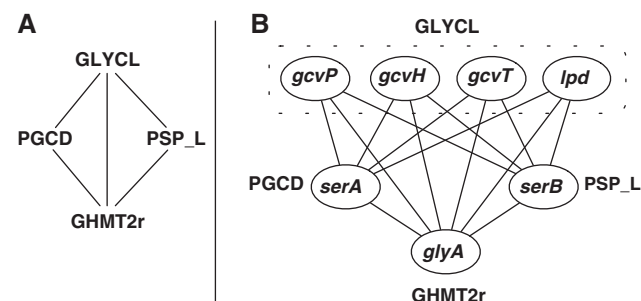We investigated the membership of genes to clusters of orthologous groups (COGs) ontology (Tatusov *et al*, 2003),

as illustrated using different colors in Figure 1. Table I lists the number of genes involved in each category. As shown in Figure 1, SL genes participate in diverse parts of cellular function, though predominantly in amino acid, nucleotide and inorganic transport and metabolism. A comparison with the COG functional classification of essential genes (Table I) reveals that a large number of essential genes are also involved in amino acid and nucleotide transport and metabolism as a consequence of the pivotal role of these pathways in contributing biomass components. However, unlike SLs, only a small portion of the essential genes are involved in inorganic ion transport and metabolism. In contrast, only a few genes in SL pairs belong to coenzyme transport and metabolism.

When analyzing the COG functional classifications, shown in Figure 1, a number of trends are revealed. We find that most lethal pairs involve genes that belong to the same COG. Notably, all genes in categories G (carbohydrate transport and metabolism), M (cell wall/membrane/envelope biogenesis) and L (replication, recombination and repair) follow the pattern of intra-category lethality with no exceptions. Using the gene–reaction–protein (GPR) associations, we deduce that these gene pairs almost always encode isozymes catalyzing essential reactions. Conversely, most lethal pairs whose genes belong to different functional groups form highly connected clusters. It has been noted earlier that two functionally distant genes can cause synthetic lethality because a gene deletion not only causes the loss of function of the primary function but also creates a cascade of compensatory cellular responses possibly affecting many pathways (Schoner *et al*, 2008). These inter-category connections are thus indicative of the need to bring to bear different parts of metabolism to enable the production of all biomass precursors. This is quite apparent for category C (energy production and conversion) for which all but two of the genes form inter-category SL gene pairs. Interestingly, the majority of the genes in this category form SLs with genes from category F (nucleotide transport and metabolism) alluding to the interdependence of nucleotide and energy (such as ATP and GTP) metabolism in supporting crucial aspects of metabolism.

In vivo comparisons of the predicted results

We searched for experimental evidence to examine the validity of the *in silico* predicted SL pairs. Direct experimental evidence was found in the literature (see Table II) for eleven such SLs. All of these SLs could be rescued by nutrient supplementation: five with amino acids alone, five with other metabolites and one with a combination of amino acids and other nutrients. One such auxotrophic example is the predicted SL (*aroK*, *aroL*). Lobner-Olesen and Marinus (1992) reported that an *E. coli* strain deficient in *aroK* (b3390) and *aroL* (b0388) requires aromatic amino acid supplementation to grow. The conversion of shikimic acid to its phosphorylated derivative, shikimate 3-phosphate is an essential step in the synthesis of aromatic amino acids in *E. coli* and is catalyzed by the two isozymes shikimic acid kinase (SK) I and II, encoded by *aroK* (b3390) and *aroL* (b0388), respectively. In another example, the cysteine supplementation requirement of an *E. coli* strain lacking both cysteine synthase genes *cysK* (b2414) and *cysM* (b2421), was observed experimentally by Saito *et al* (1993).

Figure 2 (**A**) Reaction centric view of cluster J in Figure 1. The reaction abbreviations follow those in *i*AF1260 (Feist *et al*, 2007). GLYCL, glycine cleavage system; PGCD, phosphoglycerate dehydrogenase; PSP_L, phospho-serine phosphotase (PSP_L); GHMT2r, glycine hydroxymethyl). (**B**) Arranged gene/reaction associations of cluster J.

Table I Number of essential genes and genes involved in SL gene pairs for different COG (Tatusov *et al*, 2003) functional classes

COG functional class	COG Abbreviation	# of essential genes	# of genes involved in SL pairs
Amino acid transport and metabolism	E	55	28
Nucleotide transport and metabolism	F	22	13
Inorganic ion transport and metabolism	P	7	12
Energy production and conversion	C	7	8
Cell wall/membrane/envelope biogenesis	M	20	6
Carbohydrate transport and metabolism	G	3	6
Coenzyme transport and metabolism	H	59	6
Replication, recombination and repair	L	—	2
Lipid transport and metabolism	I	18	1
Post-translational modification, protein turnover, chaperons	O	—	1
Secondary metabolites biosynthesis, transport and catabolism	Q	3	—
General function prediction only	R	2	—
Translation, ribosomal structure and biogenesis	J	1	—
Signal transduction mechanism	T	1	—
Defense mechanisms	V	1	—

Table II Direct and indirect experimental evidence for predicted SL gene pairs and their auxotrophic characteristics

SL gene pair	Topology	Experimental evidence	Growth supplementation of mutant strain
SLs with direct evidence			
ddlA (b0381), *ddlB* (b0092)	Disjoint pair	McCoy and Maurelli (2005)	D-alanyl-D-alanine (D-Ala-D-Ala) dipeptide
dadX (b1190), *alr* (b4053)	Disjoint pair	Wild *et al* (1985)	D-Ala
gutQ (b2708), *yrbH* (b3197)	Disjoint pair	Meredith and Woodard (2005)	D-arabinose 5-phosphate
acnA (b1276), *acnB* (b0118)	Disjoint pair	Gruer *et al* (1997)	Glutamate
tktA (b2935), *tktB* (b2465)	Disjoint pair	Zhao and Winkler (1994)	Aromatic amino acids and Vitamin B_6
cynT (b0339), *YadF* (b0126)	Disjoint pair	Hashimoto and Kato (2003)	High CO_2 concentration
metE (b3829), *metH* (b4019)	Disjoint pair	Ahmed (1973) and Urbanowski *et al* (1987)	Methionine
cysK (b2414), *cysM* (b2421)	Disjoint pair	Saito *et al* (1993)	Cysteine
argF (b0273), *argI* (b4254)	Disjoint pair	Lee and Cho (2006)	Arginine
aroL (b0388), *aroK* (b3390)	Disjoint pair	Lobner-Olesen *et al* (1992)	Aromatic amino acids
purT (b1849), *purN* (b2500)	Cluster A	Nygaard and Smith (1993)	Purine
SLs with indirect evidence			
ubiX (b2311), *ubiD* (b3843)	Disjoint pair	Meganathan (2001) and Gulmezian *et al* (2001)	—
metC (b3008), *malY* (b1622)	Disjoint pair	Zdych *et al* (1995)	—
aroE (b1692), *YdiB* (b3281)	Disjoint pair	Lindner *et al* (2005) and Michel *et al* (2003)	—
adk (b0474), *ndk* (b2518)	Cluster E	Willemoes and Kilstrup (2005)	—
nrdA/B (b2234/5), n*rdE/F* (b2675/6)	Cluster H	Jordan *et al* (1994b) and Jordan *et al* (1994a)	—

Another five of the experimentally verified SLs yielded strains that were auxotroph for compounds other than amino acids (see Table II). For example, Wild *et al* (1985) demonstrated that alanine racamase activity in *E. coli* is due to two distinct genes, *alr* (b4053) and *dadX* (b1190) and found that the double *alr* and *dadX* mutant (a predicted SL) is dependent on external D-Ala for growth. Similarly, McCoy and Maurelli (2005) reported the dependence of an *E. coli* strain, deficient in both *ddlA* (b0381) and *ddlB* (b0092) on an exogenous supply of D-alanyl-D-alanine (D-Ala-D-Ala) dipeptide for growth. Finally, Zhao and Winkler (1994) observed that the *tktA* (b2935) and *tktB* (b2465) double mutant, predicted to be an SL, is devoid of two transketolase isoenzymes and requires pyridoxine (vitamin B_6) as well as all aromatic amino acids and vitamins for growth.

In addition, we also uncovered five other cases for which experimental evidence indirectly supports the lethality of the identified SL pairs (see Table II). An example of this type is the predicted SL involving *ndk* (b2518) and *adk* (b0474). These two genes code for the nucleoside diphosphate kinase (Ndk) and adenylate kinase (Adk) activities, respectively, that catalyze two reactions involved in ADP synthesis (Willemoes and Kilstrup, 2005). It has been reported that the presence of Adk alone is able to restore the normal growth rates of mutant strains of *E. coli* lacking Ndk (Willemoes and Kilstrup, 2005), implying that the simultaneous disruption of both *adk* and *ndk* would be lethal for *E. coli*. Similarly, it has been shown that MalY, encoded by *malY* (b1622) is able to compensate for the methionine requirement of *metC* mutants for growth (Zdych *et al*, 1995), in agreement with the lethality of disrupting both of these two genes. Interestingly, all but one of the SLs that can be rescued by supplementation form disjoint pairs (see Table II; Figure 1). One possible reason for this is that disjoint pairs (unlike stars and *k*-connected motifs) tend to correspond to isozymes which are much more likely to have been experimentally characterized. Overall, the presence of direct

Table III Mismatches between the predicted SL pairs and experimental data for single gene knockouts

Gene (Blattner no.)[a]	Topology	Experimental condition for which is essential[b]
pyrH (b0171)	Disjoint pair	Always
ubiD (b3843)[c]	Disjoint pair	Always
pdxH (b1638)	Disjoint pair	Shared
ubiB (b3835)	Disjoint pair	Always
folA (b0048)	Disjoint pair	Always
yadF (b0126)	Disjoint pair	Always
metE (b3829)	Disjoint pair	Glucose
metL (b3940)	Disjoint pair	Shared
thrA (b0002)	Disjoint pair	Shared
metC (b3008)[c]	Disjoint pair	Shared
aroE (b3281)[c]	Disjoint pair	Shared
glnA (b3870)	Disjoint pair	Shared
eno (b2779)	Cluster C	Always
gapA (b1779)	Cluster C	Always
pgk (b2926)[d]	Cluster C	Always
guaB (b2508)[c]	Cluster D	Shared
prsA (b1207)	Cluster E	Always
adk (b0474)	Cluster E	Always
entD (b0583)[d]	Cluster F	Always
nrdA (b2234)[c]	Cluster H	Always
nrdB (b2235)[c]	Cluster H	Always
glyA (b2551)	Cluster J	Shared
lpd (b0116)[d]	Cluster J	Glucose
serA (b2913)	Cluster J	Shared
serB (b4388)	Cluster J	Shared

[a]All listed genes are reported as essential based on experimental data on glucose MOPS medium (Baba *et al*, 2006) and analyzed by (Feist *et al*, 2007). Glycerol minimal medium data were derived and analyzed by (Joyce *et al*, 2006). All conditions were aerobic.
[b]*Always*, essential under rich medium; *Glucose*, essential on glucose minimal medium conditions only; *Shared*, essential on both glucose and glycerol minimal media.
[c]At least one of the genes forming a pair with these genes is not expressed under aerobic glucose conditions based on data from (Covert *et al*, 2004), for expression level cutoff of 300.
[d]Classified as non-essential based on the analysis of the glucose minimal medium data of (Baba *et al*, 2006) by Kumar and Maranas (2009).

and indirect experimental evidence for some of the predicted SLs alludes to the reliability of the *i*AF1260 model and SL predictions.

Model refinement suggestions

Comparisons of *in silico* predictions and *in vivo* observations for single gene essentiality data (Becker and Palsson, 2008) were used before to drive the process of metabolic model refinement (Kumar and Maranas, 2009). We believe that extending this workflow to include SL pairs, triplets, etc. will provide additional layers of model validation and opportunities for correction. We identified 27 *in silico* SLs that are inconsistent with *in vivo* SL data. They fall into two different groups. The first one includes predicted SLs that contain one or more *in vivo* essential genes, whereas the latter contains predicted SL that are in agreement with *in vivo* SL data but imply incorrect supplementation rescue (i.e. auxotrophy) scenarios.

The majority of the inconsistent SL predictions (i.e. first group) involve at least one member reported as essential *in vivo* (Baba *et al*, 2006; Joyce *et al*, 2006; Feist *et al*, 2007). As indicated in Table III, there are 25 *in vivo* essential genes involved in 44 of the predicted SL pairs. Kumar and Maranas (2009) recently showed that three of these essential genes (see Table III) are most likely misclassified as alluded by their marginal essentiality scores. Of the remaining 22 genes, we find that six of them form SL pairs with genes that are not expressed under aerobic glucose minimal conditions (Covert *et al*, 2004). Therefore, the essentiality prediction for these six genes *in vivo* can be recapitulated by appending appropriate regulatory constraints to the model that restricts gene expression for seven genes under aerobic glucose conditions (see Table IV). In support of this observation, knockout

Table IV Model refinements for *i*AF1260 suggested by SL gene pair analysis

Modification[a]	Comments
Suppress *ubiX* (b2311)	Cannot complement *ubiD* (b3843); not expressed[b]
Suppress *malY* (b1622)	Cannot complement *metC* (b3008); not expressed
Suppress *ydiB* (1692)	Cannot complement *aroE* (b3281); not expressed
Suppress *ygeS* (b2866)	Cannot complement *guaB* (b2508); not expressed
Suppress *ygeT* (b2867)	Cannot complement *guaB* (b2508); not expressed
Suppress *ygeU* (b2868)	Cannot complement *guaB* (b2508); not expressed
Suppress *nrdE*[c] (b2675) and *nrdF* (b2676)	Cannot complement *nrdA* (b2234) and *nrdB* (b2235) ; not expressed
Suppress PDX5PO2	Cannot complement *pdxH* (b1638)
Suppress OPHHX3 under aerobic conditions	Cannot complement *ubiB* (b3835)
Suppress R1PK under aerobic conditions	Cannot complement *prsA* (b1207)
Suppress *cmk* (b0910)	Cannot complement *pyrH* (b0171); Hypothesis
Suppress *ydgB* (b1606)	Cannot complement *folA* (b0048); Hypothesis
Suppress *cynT* (b0339)	Cannot complement *yadF* (b0126); Hypothesis
Suppress *metH* (b4019)	Cannot complement *metE* (b3829); Hypothesis
Change HSDy GPR[a] relationship from OR to AND	*thrA* (b0002) and *metL* (b3940) cannot complement each other; both essential; Hypothesis
Suppress *puuA* (b1297) cannot complement *glnA* (b3870)	Hypothesis
Suppress *ppsA* (b1702)	Cannot complement *eno* (b2779) or *gapA* (b1779); Hypothesis
Suppress *gcvP* (b2903), *gcvH* (b2904), *gcvT* (b2905)	*gcvP* (b2903), *gcvH* (b2904), *gcvT* (b2905) cannot complement *serA* (b2913), *serB* (b4388), *glyA* (b2551); Hypothesis

[a]All modifications are for aerobic glucose conditions unless specified otherwise. GPR, gene–protein–reaction association.
[b]Not expressed under aerobic glucose conditions based on data from (Covert *et al*, 2004), for an expression level cutoff of 300.
[c]The expression level of *nrdE* from data in (Covert *et al*, 2004) was only slightly above the expression level cutoff (300).

mutants for some of these six genes have been rescued through overexpressing their SL partner(s) in *E. coli*. For example, there is genetic evidence regarding complementation of *nrdA* (b2334) or *nrdB* (b2335) mutants of *E. coli* with *nrdE* (b2675) or *nrdF* (b2676) overexpressed on a plasmid (Jordan *et al*, 1994a, b). In another example, *ydiB* (b1692) and *aroE* (b3281) encode YdiB and its paralog AroE, respectively, which are members of the quinate/shikimate 5-dehydrogenase family that functions in the essential shikimate pathway (Lindner *et al*, 2005). This relationship implies that only the simultaneous disruption of both of these genes would be lethal. However, *aroE* (b3281) is reported to be essential (Table III). This outcome may arise from inability of the low specific activity of YdiB to compensate for the deletion of AroE unless amplified, as reported by Michel *et al* (2003). We note that no regulation rules were introduced in Covert *et al* (2004) for *nrdE*, *nrdF* or *ydiB*.

Four of the 22 genes that are reported to be essential (Table III), (i.e. *pdxH* (b1638), *ubiB* (b3835), *adk* (b0474), *prsA* (b1207)) form lethal pairs with four non-gene associated reactions. In addition, *ubiB*, *adk* and *prsA* are always essential under rich medium conditions, whereas *pdxH* is essential in glucose minimal medium but not in rich medium. These results imply that the model is missing regulatory restrictions for the non-gene associated reactions listed above. We suggest that OPHHX3, R1PK and PDX5PO2 should be suppressed under the examined experimental conditions (Table IV). In particular, the non-gene associated reactions OPHHX3 and R1PK are most likely inactive under aerobic conditions (Alexander and Young, 1978; Hove-Jensen *et al*, 2003), whereas PDX5PO2 is likely to be inactive under glucose minimal conditions (Zhao and Winkler, 1995). These changes would lead to correct predictions of essentiality for the four genes.

We found only two cases of mismatches with experimental results concerning auxotrophy: (*aroL*, *aroK*) and (*tktA*, *tktB*). The SL pair (*tktA*, *tktB*) is auxotrophic for aromatic amino acids and requires the addition of pydxn (pyridoxine) to the medium (Zhao and Winkler, 1994). In contrast, the *in silico* predictions found that it remained a SL even in a rich medium, as it is unable to produce the biomass precursor pydx5p (pyridoxal 5′-phosphate). Pyridoxine is a direct precursor to pydx5p, but inspection of the transport reactions contained in *i*AF1260 reveals that no pyridoxine transport reaction is present. Thus, we resolved the *in vivo/in silico* conflict for this SL through the addition of a pyridoxine uptake pathway to the model. Interestingly, adding this uptake pathway also leads to the corrected prediction that *pdxH* (b1638) is non-essential in rich medium after implementing the regulatory adjustments. Table IV summarizes all suggested *i*AF1260 model modifications.

SL triples

The concept of synthetic (pair) lethality can be extended to SL triples, in which the simultaneous deletion of three genes is lethal. When searching for SL triples, all essential genes and SL pairs are excluded from consideration to eliminate trivial results. We identified 193 SL gene triples involving 114 genes and fifteen non-gene associated reactions (see Supplementary information). Of these predicted SL triples, 111 ($\sim$57%) found to yield auxotrophic strains *in silico* that can restore growth through the supplementation of the growth medium and the rest result in strains that cannot be rescued even in a supplemented medium (see Supplementary information for complete listing).

Similarly to SL gene pairs, a variety of different topological motifs emerge when all SL gene triples are depicted. Note that we pictorially represented them using a triangle with the three members forming the SL triple depicted as edge connected nodes (see Figure 3). Figure 3 shows a number of disjoint triples and *k*-connected clusters of different size. An example of a disjoint triple is cluster A where two genes *mgtA* (b4242) and *corA* (b3816) form a SL triple with a non-gene associated reaction (i.e. Mg2t3_2pp (magnesium (Mg^{+2}) transport in/out through proton antiport (periplasm))). All the components of this cluster are responsible for magnesium transport under different mechanisms. Cluster H is an example of a 1-connected cluster, where the presence of at least one gene (i.e. *pitA* or *pitB*) can prevent lethality. As seen in Figure 3, unlike SL gene pairs, only a small number of SL gene triples participate in disjoint triples. Instead the majority of them form complex *k*-connected clusters (e.g. clusters K, L and M). We used the mixed-integer optimization formulation proposed by Burgard *et al* (2001) to identify the minimum required set of genes (and non-gene associated reactions) in each of these clusters to prevent lethality. Surprisingly, for clusters K and L we found that the minimal sets contained only a single member (i.e. *k*=1). For example, by maintaining only the activity of *purT* (b1849) in cluster K or the activity of either *purU* (b1232) or *purN* (b2500) in cluster L, we can prevent lethality. Unlike clusters K and L, cluster M has fourteen alternative minimal sets each containing nine members (i.e. *k*=9) that need to be active to prevent lethality (see Supplementary information for complete listing).

SL reaction triples

The application of the exhaustive enumeration procedure described in the Materials and methods section for single-reaction deletions led to the identification of 277 essential reactions ($\sim$13.5% of the total number of reactions). Note that we did not allow any of the 304 exchange reactions and 29 spontaneous reactions in the *i*AF1260 model to participate in any SL. After excluding all essential, exchange, spontaneous and 981 blocked reactions, we first considered applying the exhaustive enumeration procedure (see Materials and methods section) on the 792 remaining reactions to identify what pair or triple combinations of reaction eliminations negate biomass formation. For the case of pairs, we found 96 SL reaction pairs (see Supplementary information for a complete list). However, applying this approach to identify all SL reaction triples would have required exhaustively exploring 83 million triple combinations. To avoid this computational burden, we developed a targeted enumeration procedure (see Materials and methods section) relying on a bilevel optimization procedure to identify all synthetic reaction triples without having to explicitly test all 83 million triple combinations. It identified a total of 243 SL triples involving 163 reactions.
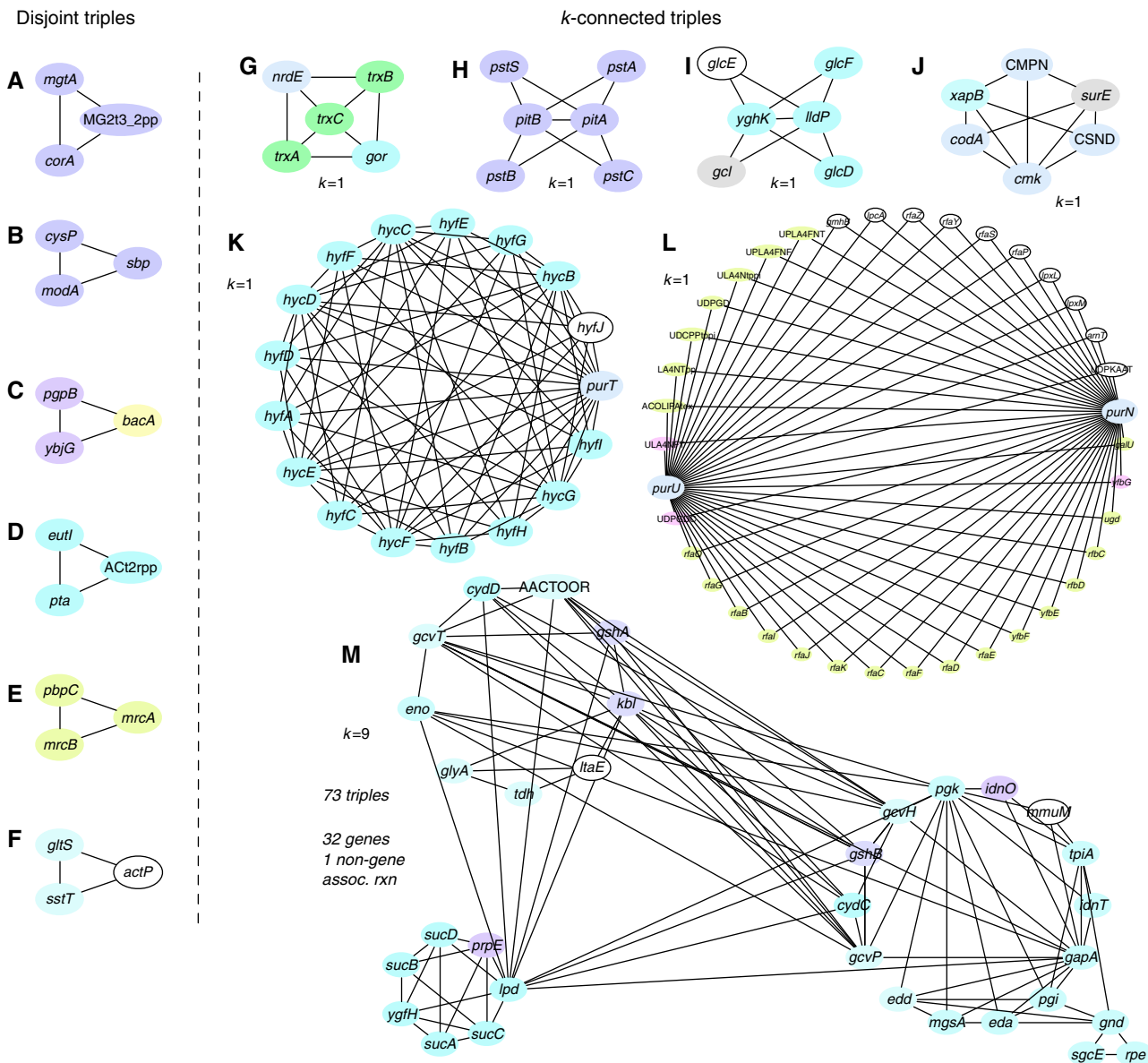
Figure 3 Topological classification of motifs in SL gene triples. Both disjoint triples (left) and *k*-connected triples (right) are seen. Names of genes are set in italics and the names of non-gene associated reactions are set in roman type. Note that all the reaction abbreviations follow those in *i*AF1260 (Feist *et al*, 2007).

Table V Comparison of the approximate CPU time (single 3 GHz) for finding each essential reaction, SL reaction pair, triples and quadruples using the exhaustive and targeted enumeration approaches, respectively

Order of SLs	Exhaustive enumeration			Targeted enumeration CPU time/SL
	Possible combinations	SL (%)	CPU time/SL	
Single	~2050	13.5	~1 s	~5 s
Double	~313,000	0.03	~28 min	~12 min
Triple	~8.3 × 10⁷	2.9 × 10⁻⁴	~2 days	~40 min
Quadruple	~1.6 × 10¹⁰	ND	ND	~5 h/SL

ND, not determined.

Table V depicts the CPU times for the targeted enumeration procedure versus the exhaustive enumeration procedure revealing orders of magnitude improvement.

Similarly to gene SLs, elimination of these reaction triples can yield auxotroph strains capable of restoring growth through the supply of missing nutrients or strains that lack essential functionalities that cannot be rescued by adding extra components to the growth medium. Of the 243 predicted essential reaction triples, as many as 202 (83%) were found to yield *in silico* auxotroph strains that can be rescued through supplementation (see Supplementary information for complete listings). For example, elimination of PGK (phosphoglycerate kinase), TALA (transaldolase) and TPI (triosephosphate isomerase) results in a strain that can be rescued (according to the model) through the supplementation of the growth medium by murein5px4p (two disaccharide linked murein units, pentapeptide crosslinked tetrapeptide). In contrast, eliminating AGMHE (ADP-D-glycero-D-manno-heptose epimerase), RPE (ribulose 5-phosphate 3-epimerase) and
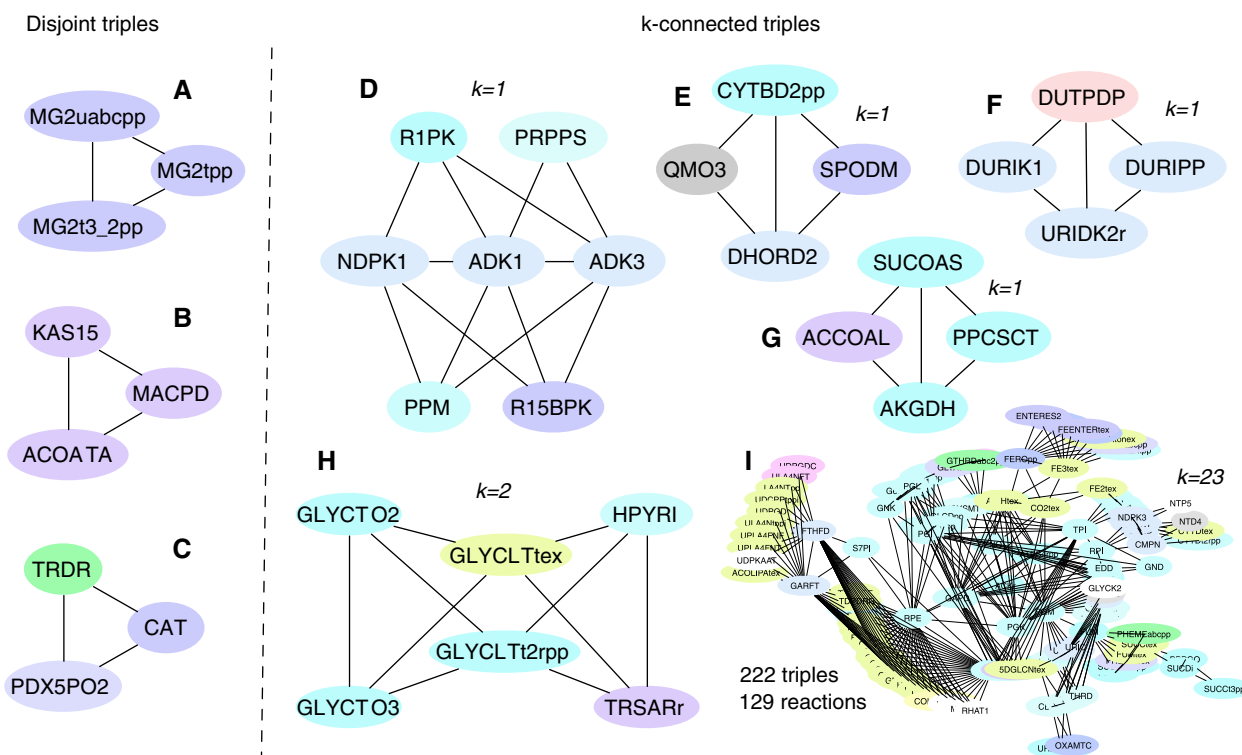
Figure 4 Topological classification of motifs in SL reaction triples. Similarly with SL gene triples in Figure 3, SL reaction triples occur as both disjoint triples (left) and *k*-connected triples (right). Note that all the reaction abbreviations follow those in *i*AF1260 (Feist *et al*, 2007).

TALA (transaldolase) yields a strain, which cannot produce *in silico* all biomass precursors even for a supplemented medium as it is unable to make or uptake the precursor metabolite pydx5p (pyridoxal 5′-phosphate).

Reaction SL triples are also pictorially shown as triangles with the three reactions depicted as edge connected nodes (see Figure 4). This graphical representation reveals that only a small number of reactions (i.e. 34) form small clusters (see Figure 4) while most of them (i.e. 129) are joined together into the highly connected large cluster I (i.e. giant component). For example, all three reactions in cluster A shown in Figure 4 are responsible for magnesium transport under different mechanisms. Interestingly, by looking at GPR associations we find that this cluster maps exactly to cluster A of Figure 3 as the two reactions MG2uabcpp and MG2tpp are coded for by the genes *mgtA* (b4242) and *corA* (b3816), respectively. The giant component (cluster I) consists of 129 reactions forming 222 SL triples. We used the mixed-integer optimization formulation proposed by Burgard *et al* (2001) to identify nine alternative reaction sets, each with 23 reactions (i.e. *k*=23) that allow for all biomass components formation (see Supplementary information). These nine minimal reaction sets spanned 29 different reactions, with seventeen of them present in all nine alternative minimal sets (see Supplementary information).

We analyzed further the reaction SL triples by determining the number of SL triples in which each reaction participates (see Table VI). We find a wide range of participation for different reactions. Most of the reactions (i.e. 85%) appear in seven or fewer triples, whereas only twelve reactions participate in more than ten triples. Notably, TPI is the most

Table VI Frequency of participation of reactions in multiple SL triples

# of SL triples	# of reactions (Rxn abb.)
1	42
2	67
3	8
4	12
5	3
6	5
7	1
8	4
9	4
10	5
11	1 (RPI)
12	1 (PPS)
17	1 (PGI)
20	1 (PGM)
27	3 (GAPD, PGK, TALA)
33	1 (RPE)
34	3 (ATPS4rpp, FTHFD, GARFT)
35	1 (TPI)

All abbreviations follow those in Feist *et al* (2007). The complete list is given in the supplementary information.

highly triple-participating reaction, with membership in 35 different SL triples. Not surprisingly, these twelve reactions appear in the complex cluster of SL triples (cluster I). Interestingly, almost half of these reactions belong to glycolysis, whereas the rest of them (except ATPS4rpp) are involved in pentose phosphate pathway (PPP), folate metabolism and purine and pyrimidine biosynthesis. Among the other interesting patterns, we find that some reactions that participate in a large number of SLs are catalyzed by proteins encoded by genes that also participate in a large number of

SLs. Specifically, the two reactions FTHFD and GARFT are associated with the genes *purU* (b1232) and *purN* (b2500) that are highly connected nodes in cluster L of Figure 3. Similar observations for cluster J of Figure 1 and clusters A of Figures 3 and 4 indicate that gene synthetic lethality can be explained by analyzing the corresponding reaction synthetic lethality.

Higher-order SLs

We identified a number of SL quadruples for the *i*AF1260 model under aerobic glucose minimal medium conditions.
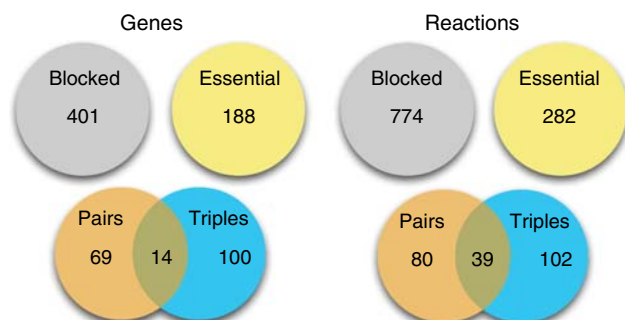

Figure 5 Venn diagram of the number of genes and reactions participating in SLs of order one, two and three.

Upon excluding spontaneous, exchange, blocked and essential reactions, 229 SL reaction quadruples with 137 reactions involved were identified (see Supplementary information for complete list). Using the targeted enumeration approach we were able to elucidate even some higher-order SL interactions, such as SL reaction quintuples. For example, the set of reactions F6PA (fructose 6-phosphate aldolase), FBA (fructose-bisphosphate aldolase), GLCptspp (D-glucose transport through PEP:Pyr PTS [periplasm]), GLCt2pp (D-glucose transport in through proton symport [periplasm]) and RPI (ribose-5-phosphate isomerase) is a SL quintuple.

Unlike essential reactions that cannot participate in any SLs, it is possible for a gene/reaction involved in SL pairs to also participate in one or more SL triples or even higher-order SLs. Figure 5 shows in the form of a Venn diagram the number of genes/reactions that participate in various orders of SLs. We note that 183 non-essential, 'non-blocked' genes participate in at least one SL pair or triple. Interestingly, we see that fourteen genes participating in at least one SL pair also participate in some of the SL triples. In the case of SL reactions, we identified 39 that participate in both SL pairs and triples. For example, the glycine cleavage system (GLYCL), which is involved in the degradation of glycine to ammonia and CO_2, participates in four SL pairs, ten SL triples and nine SL quadruples (see Figure 6). We also identified up to nineteen SL quintuples for GLYCL. This implies that the deletion of any of the reactions
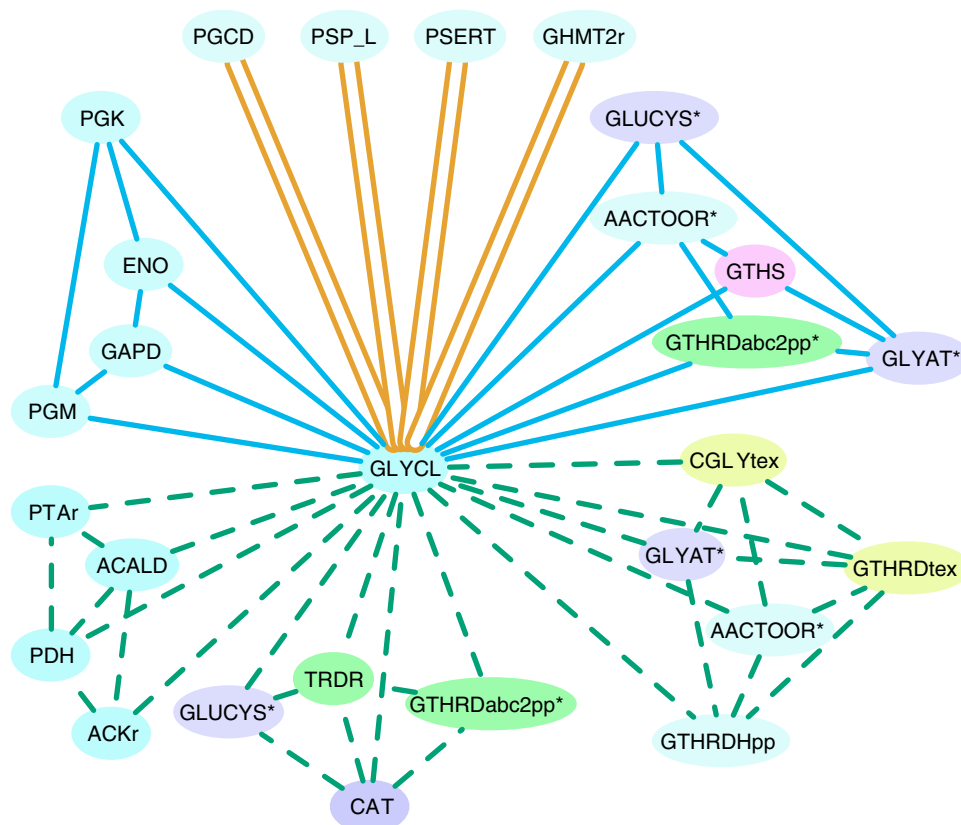

Figure 6 Pictorial view of all SLs for reaction GLYCL (glycine cleavage system). Double, solid and dashed lines depict SL pairs, triples and quadruples, respectively. The reaction abbreviations follow those in *i*AF1260 (Feist *et al*, 2007). Note that for ease of presentation reactions that occur in both SL triples and SL quadruples are repeated (marked with a star) and we show neither the SL quadruples containing the coupled fluxes nor any of the SL quintuples. The complete list of all SL quadruples and quintuples for GLYCL are given in Supplementary information.

depicted in Figure 6 can be compensated by GLYCL alone. Alternatively, the removal of GLYCL would render PGCD, PSP_L, PSERT and GHMT2r essential. Notably, all the reactions present in Figure 6 forming an SL pair with GLYCL (except PSERT) are associated with the genes of cluster J of Figure 1. Among these reactions, GHMT2r converts glycine to serine, whereas, PGCD, PSERT and PSP_L serve as the first, second and third committed steps in the serine biosynthesis pathway, respectively. Therefore, the primary reason for the synthetic lethality of GLYCL with any of these four reactions is the requirement for serine production directly or through conversion from glycine (see Supplementary information for a complete list of all precursor metabolites that cannot be produced due to the reaction eliminations). Furthermore, the elimination of GAPD or PGK with either ENO or PGM prevents the production of metabolite 3pg (3-phospho-D-glycerate), thereby blocking the serine biosynthesis pathway. Finally, the removal of GLYCL with any combination of two from GLUCYS, GLYAT, AACTOOR and GTHRDabc2pp prevents the formation of the biomass precursor metabolites coa (co-enzyme A), amet (s-adenosyl-L-methionine) and sheme (sirohem). This comprehensive synthetic lethality analysis of GLYCL demonstrates that by looking at higher-order SLs one can unravel non-intuitive biomass component deficiencies that may not be apparent from a visual inspection of the metabolic map.

A similar pattern emerged in the identified set of SL reaction quadruples. Most of the highly participating reactions in SL triples also appear with high frequency in the list of identified SL quadruples. For example, TPI participates in 55 SL quadruples. Notably, we found many instances of more than one highly participating member of SL triples occurring together in the SL quadruples. For instance, out of 55 SL quadruples found for TPI, 21 contained PGM (phosphoglycerate mutase). The reason these reactions appear in many SLs of different orders is that they serve as key branch points of central metabolism. The simultaneous removal of multiple branch points will require flux rerouting through other bypass reactions or latent pathways (Fong *et al*, 2006) for the production of essential biomass precursors such as amino acids.

Degree of essentiality

To quantify the dispensability of a gene or reaction in a metabolic network with respect to biomass formation, we introduce the concept of degree of essentiality (DOE). This metric is defined as the size of the smallest SL that the gene or reaction is a member of. Therefore, essential genes or reactions have a DOE of one, whereas genes or reactions that participate in SL pairs (and perhaps in higher-order SLs) have a DOE of two. It should be noted that the DOE metric for genes is akin to the '*k*-robustness' term introduced by Deutscher *et al* (2006). We determined the DOE of up to three for all genes and reactions and the DOE of up to four for all reactions of central metabolism active under aerobic glucose conditions. The distribution of DOE for genes and reactions present in different COG classifications (Tatusov *et al*, 2003) is shown in Table VII and Figure 7. Data in Table VII show that genes and reactions in different COGs have quite different DOE statistics. Figure 7b pictorially delineates the percentage reaction participation in

Table VII Distribution of the degree of essentiality (DOE) for all genes/reactions in the network. Table entries represent how many gene/reactions with a specific degree of essentiality belong to each COG functional class (see Table I for COG abbreviations). Note that some gene/reactions may belong to more than one COG functional class

Cog Class	Genes						Reactions					
	DOE				Blocked	Total	DOE				Blocked	Total
	1	2	3	4+			1	2	3	4+		
J	1	0	1	0	24	26	1	0	1	3	25	30
K	0	0	0	2	2	4	0	1	0	6	1	8
L	0	2	0	3	0	5	1	0	1	12	0	14
V	1	0	1	4	0	6	4	1	1	38	0	44
T	1	0	0	2	1	4	1	0	0	9	8	18
M	19	6	20	37	13	95	42	7	25	173	210	457
U	0	0	0	2	0	2	0	1	0	4	7	12
O	0	1	3	11	8	23	1	10	2	8	16	37
C	7	8	28	62	67	172	7	19	11	55	99	191
G	2	6	9	62	103	182	3	12	24	74	123	236
E	52	28	4	116	30	230	72	20	12	140	70	314
F	22	13	2	25	5	67	28	18	13	49	43	151
H	49	6	3	11	28	97	60	12	4	37	47	160
I	17	1	4	19	13	54	48	25	3	118	70	264
P	7	9	10	34	33	93	11	19	6	112	58	206
Q	0	0	0	6	4	10	22	4	2	39	22	89
R	0	0	2	16	14	32	15	6	9	97	78	205
S	0	0	0	3	1	4	1	0	2	9	3	15
None	10	3	13	73	55	154	30	8	17	123	125	303

each DOE across all COGs and reveals the differing buffering capacity of each functional category for biomass formation (Deutscher *et al*, 2006).

Next, we focus our attention to the DOE results for the reactions participating in central metabolism (spanning glycolysis, PPP, TCA cycle and anaplerotic reactions). Figure 8 illustrates the color-coded degree of essentiality of all reactions in central metabolism up to DOE of four. We can see that the majority of reactions in central metabolism have a DOE of two, three or more. This is most likely due to the presence of multiple diverging and converging branches in pathways of central metabolism. The relatively small fraction of essential reactions was expected, as earlier reports noted that conserved metabolic pathways such as the TCA cycle or glycolysis generally contain few essential reactions (Gerdes *et al*, 2003; Ghim *et al*, 2005). Most of the reactions in the PPP have a DOE of two, whereas glycolyics reactions involve DOEs of three or more. No reaction in glycolysis or PPP has a DOE of one, whereas four of the five reactions with an essentiality degree of one belong to the TCA cycle. Eliminating any of these four essential reactions will prevent the formation of the same list of precursor metabolites including sheme (sitroheme), pheme (protoheme) and murein5px4p (two disaccharide linked murein units, pentapeptide crosslinked tetrapeptide). All four of these reactions are subsequent steps in TCA cycle converting oxaloacetate to 2-ketoglutarate. Application of flux coupling analysis (Burgard *et al*, 2004) showed that three of them are fully coupled (i.e. ACONTa (aconitase [half-reaction B, Isocitrate hydro-lyase]), ACONTb (aconitase [half-reaction B, Isocitrate hydro-lyase]) and ICDHyr (isocitrate dehydrogenase)). It is important to note that reactions operating in opposite directions can have different DOEs. Such examples include reaction pairs FBP and PFK as well as PPC and PPCK.
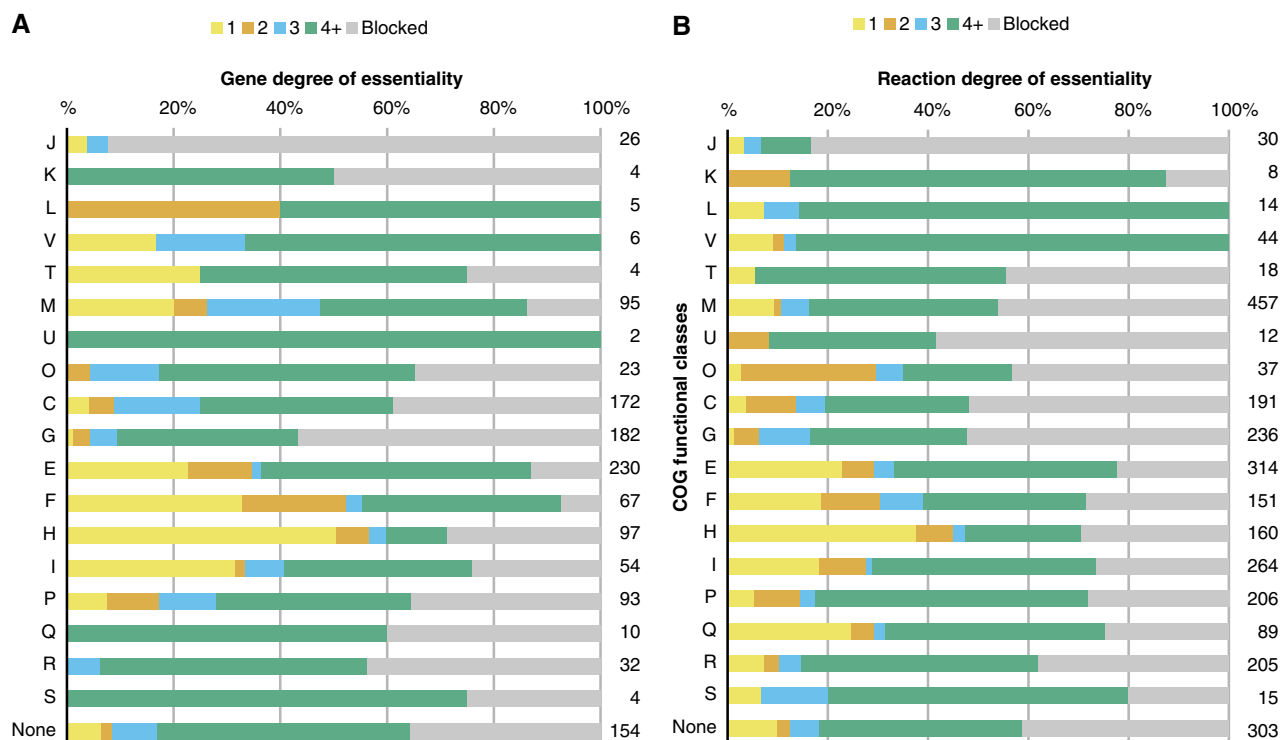
Figure 7 Percentage of degree of essentiality in different parts of the network arranged by the COG (Tatusov *et al*, 2003) functional classifications. (**A**) Genes. (**B**) Reactions. The numbers on the right indicate the total number of items in each COG classification. Note that some genes/reactions belong to more than one COG classification.

Discussion

We presented a comprehensive *in silico* study of gene and reaction synthetic lethality in *E. coli* K12 based on the recent genome-scale metabolic model *i*AF1260. This computationally intensive goal was made possible by developing an efficient procedure for the targeted enumeration of all SL interactions relying on bilevel optimization. Unlike earlier efforts that relied on incomplete sampling to elucidate partial lists of higher-order SLs (Behre *et al*, 2008; Deutscher *et al*, 2008), here we provided complete enumerations of high-order SL interactions for both gene and reaction centric representations. The network organization of the elucidated SLs recapitulated modular lethality relationships consistent with previous observations in yeast (Segre *et al*, 2005). By coloring network nodes using the COG classification, surprising compensatory interactions between seemingly unrelated gene reaction classes were revealed. Earlier efforts for the *in silico* elucidation of SLs (Wunderlich and Mirny, 2006; Palumbo *et al*, 2005, 2007) did not anticipate lethality caused by the inability to meet the non-growth associated ATP maintenance requirement. Notably, in this study, we found over 120 SLs that were triggered by this deficiency.

By contrasting literature data on gene essentiality and synthetic lethality against predicted SLs, 27 instances of inconsistencies (false-positive SLs) were identified. Similar examples of false-positive predictions can be also found for reaction SLs. For example, reaction RPI (ribose-5-phosphate isomerase) was found to be essential *in vivo* (Neidhardt and Curtiss, 1996), however, according to the model *i*AL1260, it

forms SL pairs with RPE (ribulose 5-phosphate 3-epimerase), TALA (transaldolase), TKT1 and TKT2 (transketolase) and also participates in a number of higher-order SLs (see Table IV). These erroneous model predictions are due to the fact that genome-scale metabolic model reconstructions tend to over, rather than under, predict the metabolic capabilities of the organism. This arises from the inclusion in the model of functionalities that are not active at a sufficient level (e.g. due to regulation) to ensure biomass formation. The list of SLs reported in this study should, therefore, be interpreted as a conservative depiction of synthetic lethality, as we recognize that many additional SLs are likely present that stoichiometric models alone cannot reveal (i.e. false-negatives).

We exploited the identified *in vivo/in silico* mismatches to suggest a number of *i*AF1260 model modifications (see Table IV). This is motivated by the observation that false-negative/positive model predictions provide opportunities for not only model improvement but also re-evaluation of experimental data (Thiele *et al*, 2005). Many of these model corrections do not involve the addition or removal of reactions to *i*AF1260 but rather the incorporation of regulatory constraints (i.e. condition-dependent presence of different reactions). Unfortunately, existing experimental data on SLs account for only a small portion of the predicted SL thus limiting the potential for model improvement. This calls for the development of combinatorial methods for the rapid generation and screening of SLs such as the recently developed GIANT-coli technology that allows for the high-throughput generation of double mutants in *E. coli* (Typas *et al*, 2008).
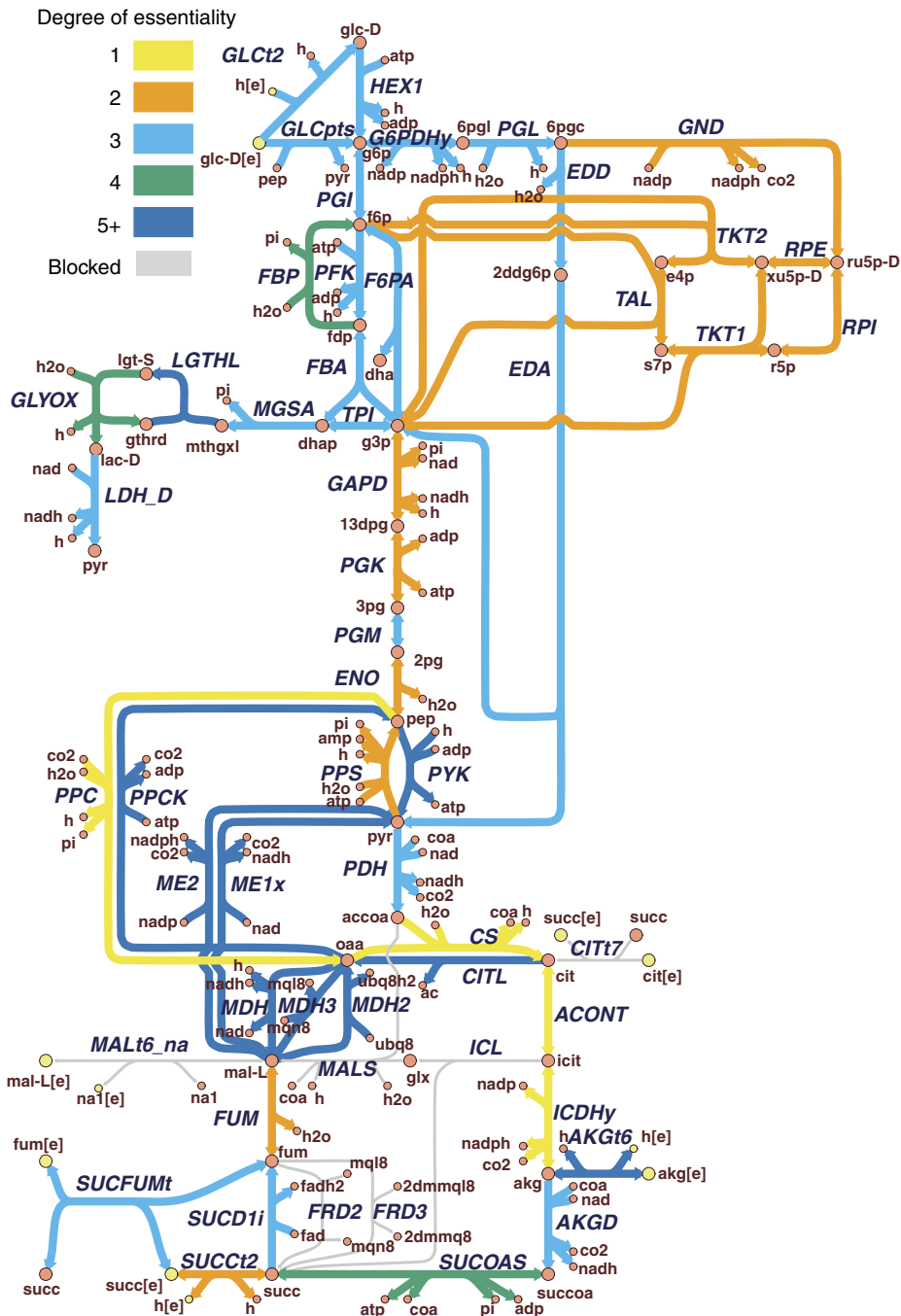
Figure 8 Color-coded representation of the reactions in central metabolism according to their degree of essentiality.

One of the key challenges in correctly interpreting the obtained predictions is the delineation of the true function of isozymes and complementation under the studied conditions. For example, the experimental evidence for the predicted SL (*ubiX, ubiD*) is conflicting. Some analyses propose that the *ubiD* knockout is lethal (Baba *et al*, 2006; Joyce *et al*, 2006; Feist *et al*, 2007), whereas the data in Covert *et al* (2004) suggest that *ubiX* is not expressed under the examined conditions. In contrast, Gulmezian *et al* (2007) recently observed

that both UbiX and UbiD are required for decarboxylation, especially during logarithmic growth, implying that both genes are essential. These inconsistencies among *in silico* predictions and *in vivo* observations call for more nuanced model modifications that are dependent on not only conditions but also on growth phase. In another example, *metL* (b3940) and *thrA* (b0002) form a disjoint SL pair and are isozymes for the aspartate kinase activity. Surprisingly, both genes are reported to be essential (Table IV), suggesting that the OR operator must

be changed into an AND operator in the gene–protein–reaction association. Therefore, simply knowing that a gene is essential is not always sufficient. In some cases, elucidating the functional reason(s) (e.g. loss of sufficient aspartate kinase activity, etc.) for this essentiality is needed before properly correcting the model. As indicated in Table III, the classification of *entD* in cluster F of Figure 1 as essential is likely erroneous. This observation is supported by the fact that eliminating FE3abcpp to make *entD* essential would also render another twelve genes essential that are known to be non-essential. Finally, predictions on whether the disruption of a particular SL results in an auxotroph strain can likewise be exploited for model correction as we demonstrated for SL (*tktA*, *tktB*).

The introduction of the concept of degree of essentiality enabled the quantitative assessment of the dispensability of any gene or reaction in the metabolic network. Moreover, by querying the derived list of SLs, one can examine how the removal of a gene/reaction affects the dispensability characteristics of other genes/reactions. Results for GLYCL (see Figure 6) revealed surprising compensatory interactions with reactions in seemingly unrelated pathways. The elucidation of SL and DOE in human metabolism has implications in the identification of drug targets. For example, it has been suggested that the SL partners of missing enzymatic functionalities of tumors cells would be promising drug targets (Hartman *et al*, 2001; Kamb, 2003). The idea is that healthy cells would remain unaffected due to the ability to compensate for the drug-suppressed functionality, whereas tumor cells, with the missing enzymatic functionality, would not. On another front, SL predictions could be used to pinpoint multi-gene disease mappings (Hoh and Ott, 2003) and identify combinations of genes most likely to interact in disease phenotypes (Wong *et al*, 2004).

The procedure proposed herein can be used to rapidly predict the growth phenotypes of multiple knockout mutants for a variety of other organisms such as *S. cerevisiae*, *B. subtilis* and *H. pylori* in various media. The effect of different conditions (i.e. alternate carbon substrates, aerobic vs anaerobic, etc.) and/or certain regulatory constraints on the membership to the SL sets can be assessed in a straightforward manner by adjusting appropriate model constraints based on exchange reaction usage and gene/reaction availability.

Materials and methods

Two separate optimization-based procedures for the enumeration of all SLs in the gene and reaction levels are described. The first one relies on the exhaustive biomass formation capability evaluation of all single, double, triple, etc. combinations of gene and/or reaction deletions. This method becomes computationally prohibitive when searching for higher-order SLs ($n>2$). Therefore, an alternative much more efficient and targeted method, relying on bilevel optimization, is described that identifies all SL combinations without relying on the exhaustive enumeration of all possible gene/reaction eliminations. To reduce the search space in both of these approaches, a flux coupling analysis (Burgard *et al*, 2004) was performed as a pre-processing step to allow the removal of only one representative of each fully coupled reaction set. It should be noted, however, that the complete lists of reactions present in SL of degree two, three and four are provided in the Supplementary information.

The analysis of synthetic lethality for the metabolic networks requires the introduction of the following sets:

$$I = \{i|i = 1, 2, ..., N\} = \text{set of metabolites}$$
$$J = \{j|j = 1, 2, ..., M\} = \text{set of reactions}$$
$$K = \{k|k = 1, 2, ..., G\} = \text{set of genes}$$

where, N, M and G denote the total number of metabolites, reactions and genes in the network, respectively. On imposing metabolite balances across the entire metabolic model under steady-state conditions we obtain:

$$\sum_j s_{ij} v_j = 0, \quad \forall i \in I \tag{1}$$

where, s_{ij}, represents the stoichiometric coefficient of the metabolite i, in reaction j, and v_j, denotes the flux of reaction j. Next, we describe the exhaustive and targeted SL enumeration approaches in detail.

Exhaustive enumeration of SL interactions

The computational prediction of synthetic lethality hinges on the calculation of the maximum biomass formation in the presence of the gene or reaction deletions implied by the examined SL. We chose 1% of the maximum theoretical biomass yield as the cutoff for computationally predicted growth. We found that the prediction of *in silico* lethality was not particularly sensitive on the selected biomass formation cutoff value. For example, when considering single-gene mutants only eight mutants (which is only about 0.6% of all genes in the model) involve biomass formation values between 1–50%. We are aware of the definition of *in silico* synthetic lethality in the Deutscher *et al* (2006) study. However, we believe that using a conservative universal cutoff of 1% for all mutants safeguards against the possibility of misclassifying as SL gene mutants that could be viable. In other words, we want to minimize the occurrence of false-positives perhaps at the expense of missing some SLs. If D is the set of reactions that is set to zero either directly or as a consequence of the GPR associations implied by the gene deletions then the problem of determining the maximum biomass formation can be formulated as the following linear program:

$$\text{Maximize} \quad v_{\text{biomass}} \quad [\text{MaxBiomass}] \tag{2}$$
$$s.t.$$

$$\sum_j s_{ij} v_j = 0 \quad \forall i \in I \tag{1}$$

$$LB_j \leqslant v_j \leqslant UB_j \quad \forall j \in J \tag{3}$$

$$v_d = 0 \quad d \in D \subset J \tag{4}$$

$$v_{\text{glucose}} \geqslant v_{\text{glucose}}^{\text{uptake limit}} \tag{5}$$

$$v_{\text{oxygen}} \geqslant v_{\text{oxygen}}^{\text{uptake limit}} \tag{6}$$

$$v_{\text{ATPM}} = v_{\text{ATPM}}^{\text{maintenance}} \tag{7}$$

$$v_j \in \mathfrak{R} \quad \forall j \in j$$

Here, v_{biomass} denotes the biomass flux while $v_{\text{glucose}}^{\text{uptake limit}}$, $v_{\text{oxygen}}^{\text{uptake limit}}$ and $v_{\text{ATPM}}^{\text{maintenance}}$ denote the minimum required glucose and oxygen uptake rates and the non-growth associated ATP for maintenance, respectively. The values of the upper and lower bounds, UB_j and LB_j, in Equation (3) were chosen as not to exclude any physiologically relevant metabolic flux values. The upper bound for all reactions was set to 1000. The lower bound was set to zero for irreversible reactions and to −1000 for reversible reactions. For any external carbon containing metabolite, the maximum transport rate into the cell was

set to $20\,\mathrm{mmol\,gDW^{-1}\,h^{-1}}$. For the remaining source exchange fluxes, the lower bound was set to $-1000\,\mathrm{mmol\,gDW^{-1}\,h^{-1}}$ (Feist *et al*, 2007). Glucose minimal conditions were modeled by restricting the glucose uptake rate at $10\,\mathrm{mmol\,gdW^{-1}\,h^{-1}}$ and the oxygen uptake rate at $20\,\mathrm{mmol\,gdW^{-1}\,h^{-1}}$. The non-growth associated ATP maintenance was fixed at $8.39\,\mathrm{gDW^{-1}\,h^{-1}}$ (Feist *et al*, 2007).

The elucidation of SL reactions using formulation MaxBiomass is straightforward as the membership in set D (the set of reaction deletions) is known *a priori*. In the case of gene deletions, information gleaned from the GPR association relations needs to be encoded to elucidate the effect of gene deletions onto reaction deletions accounting for isozymes, multi-meric enzymes and combination thereof. Formulation MaxBiomass is iteratively solved to identify SL for genes or reactions involving a pre-specified number n of deletions (where n is the order of SL sought-after). This exhaustive evaluation becomes computationally intractable for higher-order (higher than two) SLs (see Table V) motivating the development of the following targeted enumeration procedure.

Targeted enumeration of SLs

The proposed targeted enumeration procedure relies on the solution of a bilevel optimization formulation that identifies n simultaneous gene/reaction deletions suppressing biomass formation. This bilevel formulation identifies the set of n gene/reaction deletions that minimizes the maximum biomass formation potential of the network. If the minimal value of the maximum biomass is found to be below the imposed cutoff (i.e. 1% of maximum biomass) then the corresponding combination of n gene/reaction deletions forms a SL. It is important to note that this obviates the need to explore exhaustively all deletion combinations as the bilevel formulation 'homes in' in only the biomass negating combinations.

The mathematical description of the bilevel formulation to determine SL reactions requires the definition of binary variable y_j that encodes which reactions are deleted:

$$y_j = \begin{cases} 0, & \text{if reaction } j \text{ is eliminated} \\ 1, & \text{if reaction } j \text{ is active} \end{cases} \quad \forall j \in J \quad (8)$$

This allows us to put forth the following min–max bilevel optimization formulation (SL Finder):

$$\underset{y_j}{\text{Minimize}} \quad v_{\text{biomass}} \quad \text{[Outer]} \quad \text{[SL Finder]}$$

$$s.t.$$

$$\left[\begin{array}{l} \underset{v_j}{\text{Maximize}} \quad v_{\text{biomass}} \quad \text{[Inner]} \\ s.t. \\ \quad \sum_j s_{ij}v_j = 0 \qquad \forall i \in I \\ \quad v_j \leqslant UB_j y_j \qquad \forall j \in J \\ \quad v_{\text{glucose}} \leqslant v_{\text{glucose}}^{\text{uptake limit}} \\ \quad v_{\text{oxygen}} \leqslant v_{\text{oxygen}}^{\text{uptake limit}} \\ \quad v_{\text{ATPM}} = v_{\text{ATPM}}^{\text{maintenance}} \\ \quad v_j = 0 \qquad \forall j \in J \end{array} \right.$$

$$\sum_j (1 - y_j) \leqslant n$$

$$y_j \in \{0, 1\} \qquad \forall j \in J$$

Formulation (SL Finder) is a min–max mixed-integer linear program. The inner problem adjusts the fluxes to achieve maximum biomass production, subject to network stoichiometry, reaction deletions imposed by the outer problem and other possible growth and environmental constraints. The outer problem on the other hand, aims at finding synthetic reaction eliminations that lower the maximum biomass production below the imposed cutoff. Here, we split the reversible fluxes into forward and backward reaction steps. To solve the bilevel formulation shown above the inner maximization is recast as a set of constraints by appending to the formulation the list of constraints corresponding to the dual of the inner problem and setting

the primal objective function equal to the dual as introduced before in (Burgard *et al*, 2003):

$$\text{Minimize} \quad v_{\text{biomass}} \quad (9)$$

$$s.t.$$

$$v_{\text{biomass}} = \sum_j \mu_j UB_j y_j + v_{\text{ATPM}}^{\text{maintenance}} \cdot \mu_{\text{ATPM}} \quad (10)$$

$$\sum_j s_{ij}v_j = 0, \quad \forall i \in I \quad (1)$$

$$v_j \leqslant UB_j y_j \quad \forall j \in J \quad (11)$$

$$v_{\text{glucose}} \leqslant v_{\text{glucose}}^{\text{uptake limit}} \quad (12)$$

$$v_{\text{oxygen}} \leqslant v_{\text{oxygen}}^{\text{uptake limit}} \quad (13)$$

$$v_{\text{ATPM}} = v_{\text{ATPM}}^{\text{maintenance}} \quad (7)$$

$$\sum_i \lambda_i s_{ij} + \mu_j \geqslant 0 \quad \forall j \in J - \{\text{biomass, ATPM}\} \quad (14)$$

$$\sum_i \lambda_i s_{i\,\text{biomass}} + \mu_{\text{biomass}} \geqslant 1 \quad (15)$$

$$\sum_i \lambda_i s_{i\text{ARPM}} + \mu_{\text{ATPM}} \geqslant 0 \quad (16)$$

$$\sum_j (1 - y_j) \leqslant n \quad (17)$$

$$v_j, \mu_j \geqslant 0 \qquad \forall j \in J - \{\text{ATPM}\}$$

$$\lambda_i, \mu_{\text{ATPM}} \in R \quad \forall i \in I$$

$$y_j \in \{0, 1\} \qquad \forall j \in J$$

Here, λ_i, μ_j and μ_{ATPM} are the dual variables associated with the stoichiometric constraints (equation (1)), inequalities in equation (11) and the constraint for non-growth associated ATP maintenance (equation (7)), respectively. Equations (12) and (13) replace (5) and (6), respectively, because of the split of the reversible fluxes into forward and backward reaction steps. Note that the non-linear term $\mu_j y_j$ in equation (10) are exactly linearized as follows:

$$\alpha_j = \mu_j y_j \quad (18)$$

$$\mu_j^{\text{min}} y_j \leqslant \alpha_j \leqslant \mu_j^{\text{max}} y_j \quad (19)$$

$$\mu_j - \mu_j^{\text{max}}(1 - y_j) \leqslant \alpha_j \leqslant \mu_j - \mu_j^{\text{min}}(1 - y_j) \quad (20)$$

where μ_j^{min} and μ_j^{max} are the lower and upper bounds on the dual variable μ_j. If the optimal objective function value of the above optimization problem is less than the imposed cutoff, then the reactions for which $y_j = 0$, are reported as a SL of degree n. All alternative SL reaction sets of size n are successively obtained by excluding the previously identified SLs using integer cuts and resolving the bilevel formulation. For example, if reactions $j_1, j_2, \dots, j_n$ are found to form a SL set of size n, we can exclude this solution and obtain the next one by appending the following constraint to the formulation that ensures that at least one of the reactions forming the previously identified SL is active.

$$y_{j_1} + y_{j_2} + \dots + y_{j_n} \geqslant 1 \quad j_1, j_2, \dots, j_n \in J \quad (21)$$

Note that while searching for the set of all SL reactions of a particular order n, we need to preclude the removal of all reactions forming lower order SLs. This is accomplished by appending constraints of the following form to the outer problem of the bilevel program for all reactions $j_1, j_2, \dots, j_p$ forming a SL set of size p ($p < n$):

$$y_{j_1} + y_{j_2} + \dots + y_{j_p} \geqslant 1 \quad j_1, j_2, \dots, j_p \in J, \quad \forall p < n \quad (22)$$

It is important to note that the elimination of certain reactions can prevent equation (7) from being satisfied, which precludes the

identification of some SLs. Merging the required non-growth ATP for maintenance into the biomass equation resolves this problem (see Supplementary information for details).

The formulation (SL Finder) introduced above can be modified to find the set of all SL genes by adding a set of constraints to the outer problem describing GPR associations. To this end, a binary variable w_k, representing if a gene k should be deleted is defined as following:

$$w_k = \begin{cases} 0, & \text{if gene } k \text{ is deleted} \\ 1, & \text{if gene } k \text{ is active} \end{cases}, \quad \forall k \in K \qquad (23)$$

The impact of gene deletions on reaction eliminations through GPR relationships can be mathematically described and incorporated into the model by using appropriate equations relating the binary variables w_k and y_j (see Supplementary information for details).

Supplementary information

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

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ahmed A (1973) Mechanism of repression of methionine biosynthesis in Escherichia coli. I. The role of methionine, s-adenosylmethionine, and methionyl-transfer ribonucleic acid in repression. *Mol Gen Genet* **123**: 299–324

Alexander K, Young IG (1978) Alternative hydroxylases for the aerobic and anaerobic biosynthesis of ubiquinone in Escherichia coli. *Biochemistry* **17**: 4750–4755

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**: 2006.0008

Becker SA, Palsson BO (2008) Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst Biol* **2**: 14

Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S (2008) Structural robustness of metabolic networks with respect to multiple knockouts. *J Theor Biol* **252**: 433–441

Bender A, Pringle JR (1991) Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in Saccharomyces cerevisiae. *Mol Cell Biol* **11**: 1295–1305

Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* **14**: 301–312

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

Burgard AP, Vaidyaraman S, Maranas CD (2001) Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnol Prog* **17**: 791–797

Chen CY, Graham TR (1998) An arf1Delta synthetic lethal screen identifies a new clathrin heavy chain conditional allele that perturbs vacuolar protein transport in Saccharomyces cerevisiae. *Genetics* **150**: 577–589

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**: 92–96

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

Deutscher D, Meilijson I, Schuster S, Ruppin E (2008) Can single knockouts accurately single out gene functions? *BMC Syst Biol* **2**: 50

Diestel R (2005) *Graph Theory*, 3rd edn. Berlin, New York: Springer

Dolma S, Lessnick SL, Hahn WC, Stockwell BR (2003) Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer Cell* **3**: 285–296

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**: 121

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

Forsburg SL (2001) The art and design of genetic screens: yeast. *Nat Rev Genet* **2**: 659–668

Forster J, Famili I, Fu P, Palsson BO, Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res* **13**: 244–253

Gerdes SY, Scholle MD, Campbell JW, Balazsi 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, Barabasi AL, Oltvai ZN *et al* (2003) Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185**: 5673–5684

Ghim CM, Goh KI, Kahng B (2005) Lethality and synthetic lethality in the genome-wide metabolic network of Escherichia coli. *J Theor Biol* **237**: 401–411

Grate L, Ares Jr M (2002) Searching yeast intron data at Ares lab Web site. *Methods Enzymol* **350**: 380–392

Gruer MJ, Bradbury AJ, Guest JR (1997) Construction and properties of aconitase mutants of Escherichia coli. *Microbiology* **143** (Part 6): 1837–1846

Guarente L (1993) Synthetic enhancement in gene interaction: a genetic tool come of age. *Trends Genet* **9**: 362–366

Gulmezian M, Hyman KR, Marbois BN, Clarke CF, Javor GT (2007) The role of UbiX in Escherichia coli coenzyme Q biosynthesis. *Arch Biochem Biophys* **467**: 144–153

Hartman JLt, Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. *Science* **291**: 1001–1004

Hashimoto M, Kato J (2003) Indispensability of the Escherichia coli carbonic anhydrases YadF and CynT in cell proliferation at a low CO2 partial pressure. *Biosci Biotechnol Biochem* **67**: 919–922

Hoh J, Ott J (2003) Mathematical multi-locus approaches to localizing complex human trait genes. *Nat Rev Genet* **4**: 701–709

Hove-Jensen B, Rosenkrantz TJ, Haldimann A, Wanner BL (2003) Escherichia coli phnN, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. *J Bacteriol* **185**: 2793–2801

Jordan A, Gibert I, Barbe J (1994a) Cloning and sequencing of the genes from Salmonella typhimurium encoding a new bacterial ribonucleotide reductase. *J Bacteriol* **176**: 3420–3427

Jordan A, Pontis E, Atta M, Krook M, Gibert I, Barbe J, Reichard P (1994b) A second class I ribonucleotide reductase in Enterobacteriaceae: characterization of the Salmonella typhimurium enzyme. *Proc Natl Acad Sci USA* **91**: 12892–12896

Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S (2006) Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **188**: 8259–8271

Kaelin WG (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer* **5**: 689–698

Kamb A (2003) Mutation load, functional overlap, and synthetic lethality in the evolution and treatment of cancer. *J Theor Biol* **223**: 205–213

Kim PJ, Lee DY, Kim TY, Lee KH, Jeong H, Lee SY, Park S (2007) Metabolite essentiality elucidates robustness of Escherichia coli metabolism. *Proc Natl Acad Sci USA* **104**: 13638–13642

Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* **15**: 1421–1430

Kumar VS, Maranas CD (2009) GrowMatch: an automated method for reconciling in silico/*in vivo* growth predictions. *PLoS Comput Biol* **5:** e1000308

Le Meur N, Gentleman R (2008) Modeling synthetic lethality. *Genome Biol* **9:** R135

Lee YJ, Cho JY (2006) Genetic manipulation of a primary metabolic pathway for L-ornithine production in Escherichia coli. *Biotechnol Lett* **28:** 1849–1856

Lindner HA, Nadeau G, Matte A, Michel G, Menard R, Cygler M (2005) Site-directed mutagenesis of the active site region in the quinate/shikimate 5-dehydrogenase YdiB of Escherichia coli. *J Biol Chem* **280:** 7162–7169

Lobner-Olesen A, Marinus MG (1992) Identification of the gene (aroK) encoding shikimic acid kinase I of Escherichia coli. *J Bacteriol* **174:** 525–529

McCoy AJ, Maurelli AT (2005) Characterization of Chlamydia MurC-Ddl, a fusion protein exhibiting D-alanyl-D-alanine ligase activity involved in peptidoglycan synthesis and D-cycloserine sensitivity. *Mol Microbiol* **57:** 41–52

Meganathan R (2001) Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol Lett* **203:** 131–139

Meredith TC, Woodard RW (2005) Identification of GutQ from Escherichia coli as a D-arabinose 5-phosphate isomerase. *J Bacteriol* **187:** 6936–6942

Michel G, Roszak AW, Sauve V, Maclean J, Matte A, Coggins JR, Cygler M, Lapthorn AJ (2003) Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. *J Biol Chem* **278:** 19463–19472

Mullen JR, Kaliraman V, Ibrahim SS, Brill SJ (2001) Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in Saccharomyces cerevisiae. *Genetics* **157:** 103–118

Neidhardt FC, Curtiss R (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. Washington, DC: ASM Press

Novick P, Osmond BC, Botstein D (1989) Suppressors of yeast actin mutations. *Genetics* **121:** 659–674

Nygaard P, Smith JM (1993) Evidence for a novel glycinamide ribonucleotide transformylase in Escherichia coli. *J Bacteriol* **175:** 3591–3597

Ooi SL, Pan X, Peyser BD, Ye P, Meluh PB, Yuan DS, Irizarry RA, Bader JS, Spencer FA, Boeke JD (2006) Global synthetic-lethality analysis and yeast functional profiling. *Trends Genet* **22:** 56–63

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

Palumbo MC, Colosimo A, Giuliani A, Farina L (2005) Functional essentiality from topology features in metabolic networks: a case study in yeast. *FEBS Lett* **579:** 4642–4646

Palumbo MC, Colosimo A, Giuliani A, Farina L (2007) Essentiality is an emergent property of metabolic network wiring. *FEBS Lett* **581:** 2485–2489

Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD (2004) A robust toolkit for functional profiling of the yeast genome. *Mol Cell* **16:** 487–496

Plaimas K, Mallm JP, Oswald M, Svara F, Sourjik V, Eils R, Konig R (2008) Machine learning based analyses on metabolic networks supports high-throughput knockout screens. *BMC Syst Biol* **2:** 67

Saito K, Kurosawa M, Murakoshi I (1993) Determination of a functional lysine residue of a plant cysteine synthase by site-directed mutagenesis, and the molecular evolutionary implications. *FEBS Lett* **328:** 111–114

Schoner D, Kalisch M, Leisner C, Meier L, Sohrmann M, Faty M, Barral Y, Peter M, Gruissem W, Buhlmann P (2008) Annotating novel genes by integrating synthetic lethals and genomic information. *BMC Syst Biol* **2:** 3

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

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41

Thiele I, Vo TD, Price ND, Palsson BO (2005) Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol* **187:** 5818–5830

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

Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L *et al* (2004) Global mapping of the yeast genetic interaction network. *Science* **303:** 808–813

Tucker CL, Fields S (2003) Lethal combinations. *Nat Genet* **35:** 204–205

Typas A, Nichols RJ, Siegele DA, Shales M, Collins SR, Lim B, Braberg H, Yamamoto N, Takeuchi R, Wanner BL, Mori H, Weissman JS, Krogan NJ, Gross CA (2008) High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat Methods* **5:** 781–787

Urbanowski ML, Stauffer LT, Plamann LS, Stauffer GV (1987) A new methionine locus, metR, that encodes a trans-acting protein required for activation of metE and metH in Escherichia coli and Salmonella typhimurium. *J Bacteriol* **169:** 1391–1397

Wang T, Bretscher A (1997) Mutations synthetically lethal with tpm1delta lie in genes involved in morphogenesis. *Genetics* **147:** 1595–1607

Wild J, Hennig J, Lobocka M, Walczak W, Klopotowski T (1985) Identification of the dadX gene coding for the predominant isozyme of alanine racemase in Escherichia coli K12. *Mol Gen Genet* **198:** 315–322

Wilhelm T, Behre J, Schuster S (2004) Analysis of structural robustness of metabolic networks. *Syst Biol (Stevenage)* **1:** 114–120

Willemoes M, Kilstrup M (2005) Nucleoside triphosphate synthesis catalysed by adenylate kinase is ADP dependent. *Arch Biochem Biophys* **444:** 195–199

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

Wunderlich Z, Mirny LA (2006) Using the topology of metabolic networks to predict viability of mutant strains. *Biophys J* **91:** 2304–2311

Ye P, Peyser BD, Pan X, Boeke JD, Spencer FA, Bader JS (2005) Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol Syst Biol* **1:** 2005.0026

Zdych E, Peist R, Reidl J, Boos W (1995) MalY of Escherichia coli is an enzyme with the activity of a beta C-S lyase (cystathionase). *J Bacteriol* **177:** 5035–5039

Zhao G, Winkler ME (1994) An Escherichia coli K-12 tktA tktB mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth. *J Bacteriol* **176:** 6134–6138

Zhao G, Winkler ME (1995) Kinetic limitation and cellular amount of pyridoxine (pyridoxamine) 5′-phosphate oxidase of Escherichia coli K-12. *J Bacteriol* **177:** 883–891