available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/biochempharm

Use of constraint-based modeling for the prediction and validation of antimicrobial targets

John D. Trawick *, *Christophe H. Schilling*

Genomatica, Inc., 5405 Morehouse Dr., Suite 210, San Diego, CA 92121, United States

ARTICLE INFO

Article history:
Received 31 August 2005
Accepted 25 October 2005

Keywords:
Constraint-based modeling
Metabolism
Antimicrobial research
In silico modeling
Essential gene
Flux balance analysis

ABSTRACT

The overall process of antimicrobial drug discovery and development seems simple, to cure infectious disease by identifying suitable antibiotic drugs. However, this goal has been difficult to fulfill in recent years. Despite the promise of the high-throughput innovations sparked by the genomics revolution, discovery, and development of new antibiotics has lagged in recent years exacerbating the already serious problem of evolution of antibiotic resistance. Therefore, both new antimicrobials are desperately needed as are improvements to speed up or improve nearly all steps in the process of discovering novel antibiotics and bringing these to clinical use. Another product of the genomic revolution is the modeling of metabolism using computational methodologies. Genomic-scale networks of metabolic reactions based on stoichiometry, thermodynamics and other physico-chemical constraints that emulate microbial metabolism have been developed into valuable research tools in metabolic engineering and other fields. This constraint-based modeling is predictive in identifying critical reactions, metabolites, and genes in metabolism. This is extremely useful in determining and rationalizing cellular metabolic requirements. In turn, these methods can be used to predict potential metabolic targets for antimicrobial research especially if used to increase the confidence in prioritization of metabolic targets. The many different capacities of constraint-based modeling also enable prediction of cellular response to specific inhibitors such as antibiotics and this may, ultimately find a role in drug discovery and development. Herein, we describe the principles of metabolic modeling and how they might initially be applied to antimicrobial research.

1. Introduction

The age of antibiotics blossomed in the 1940s with the introduction of penicillin, followed by several other classes of antibiotics. Unfortunately, not only does resistance frequently evolve to antibiotics but the chemical diversity of antibiotics and the cellular processes targeted by commercially viable antibiotics is limited [1]. Unfortunately, the FDA has approved very few new classes of antibiotics such as the oxazolidinones and daptomycin in the last decade [2]. Though both are useful clinically, neither is a broad-spectrum drug that can be used in the community. It was hoped that the promise of genomics would break this impasse [3,4] by generating lists of novel targets that could be screened for novel inhibitors that could in turn, be developed into new drugs. Instead, thousands of novel genes have been discovered, nearly all still remain uncharacterized as to function, a few have had qualities suitable as antibiotic targets [5,6], and practically no novel antibiotics have survived preclinical development. This situation and its causes have been reviewed extensively [2,5–10].

* *Corresponding author*. Tel.: +1 858 362 8569; fax: +1 858 824 1772.
E-mail address: jtrawick@genomatica.com (J.D. Trawick).
0006-2952/$ – see front matter
doi:10.1016/j.bcp.2005.10.049

High-throughput methods might help with these issues. For instance, essential gene targets can be identified via insertional mutagenesis [11–15], using antisense expression [16–18], and by bioinformatic conservation [13,19,20]. However, broad conservation of a gene does not mean that it is a good target nor is it without artifacts due to the somewhat arbitrary choice of conservation filters [20,21]. High-throughput experimentation has generated large numbers of potential essential genes ('essential' will be defined as necessary for viability on standard rich medium in the laboratory even though this may vary for different species/media while 'in vivo essential' or 'virulence' are necessary for infection of the host); sometimes up to 25% of the cellular complement has been found essential [12,14,18, 22,23,11] but these results appear to be in conflict with the traditional, gene by gene knockout methods that have found essential genes in number between 6 and 17% of the cell's total [24,25]. This means that each 'essential' gene found by a high-throughput method must be confirmed through directed knockout and its orthologs in each target pathogen must be similarly validated. High-throughput methods to discover the mode of action of potential antibiotics have been described [26–28] that promise to identify functions of potential drugs or to deprioritize non-specific inhibitors. However, a target profile generated by one of these methods does not prove the actual target; lacking are the means to predict what a profile should look like. In place of this, controls using known antibiotics and antifungals are utilized but these cannot tell researchers how unrelated compounds targeting a different spectrum of targets will act. Similar profiling, or lack of it, could handicap drug screening, including methods based on cell-based assays. In these systems, hit screening compounds are usually chosen with statistical assumptions based on control well performance [29,30]. It is possible to imagine that improved modeling of the performance of cell-based assays could improve identification of specific inhibitors regardless of potency and allow for elimination of highly potent poisons. Doing so, might speed up drug development by both advancing useful compounds and rapidly eliminating problematic entities. Metabolism has been a potent source of antibiotic targets and metabolism is particularly amenable to theoretical modeling [1].

2. In silico modeling

In less than 20 years, the fields of theoretical microbiology and industrial microbiology have experienced a revolution in methodology partly due to the introduction of mathematical modeling of cellular, specifically microbial, metabolism that can be applied to create simulations of the flux through the metabolic pathways of a hypothetical, 'in silico', cell [31,32]. As shown in Fig. 1, genome-scale metabolic models can today be constructed from the knowledge of the organism's genes,

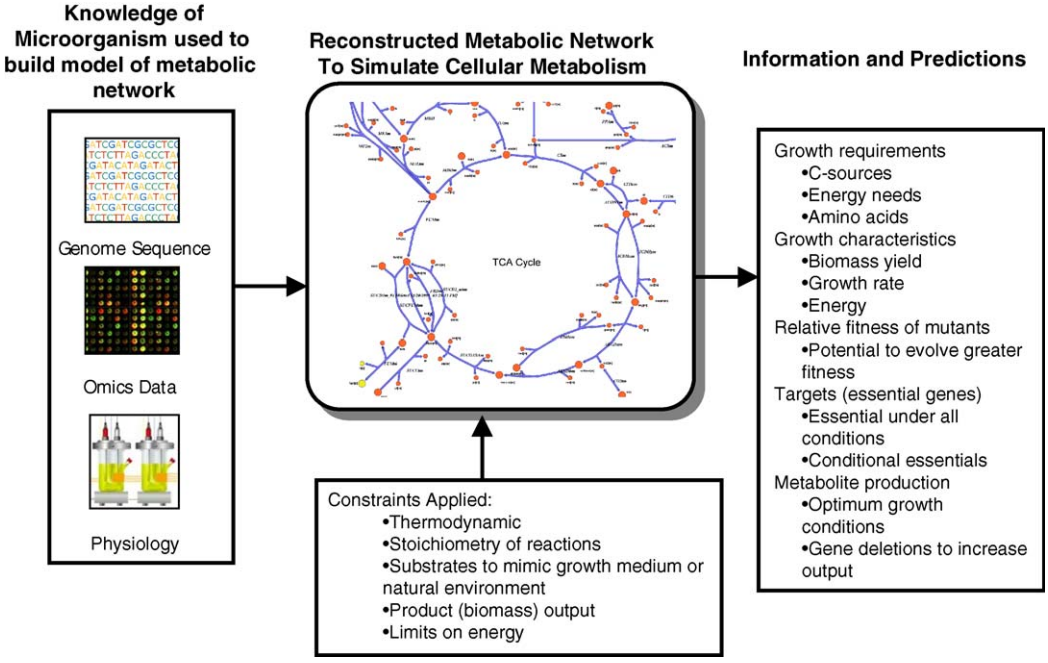

Fig. 1 – Constraint-based modeling with genome-scale metabolic models. This schematic illustrates the nature of an in silico metabolic model including input and output information. These models are built with information from genome sequences, annotation, bioinformatic processing, relevant literature, and experimental data to create a network of the metabolic reactions of a microorganism. In SimPheny™ [33,42], the network is mass and charge balanced and operates using reaction stoichiometry and thermodynamics. Considerable effort is needed to ensure that the model is representative of the actual cellular metabolic network and is as complete as possible. Input information consists of in silico "nutrients" which, with modeling of the reaction network, predict the output in the form concentrations of cellular components, total biomass, and growth rate. Constraints include simulation of growth conditions, reaction stoichiometry, and thermodynamics. A variety of computational methods are used to simulate metabolism. Results are in the form of predictions about gene essentiality, conditional gene essentiality, growth rate, biomass, energy needs, and production of metabolites.

physiology, proteomics, and anything else that might aid in an accurate depiction of the metabolic network unique to each species. These models provide accurate representation of metabolism in a manner that is consistent with the known genetics and biochemistry of the organism, and in a format amenable for computational interrogation. Constraints can be applied to the model in the form of substrate availability, energy requirements, and objectives such as growth, reaction stoichiometry, and thermodynamics. Finally, the in silico model, really a set of programs, produces results in the form of ability to generate biomass, metabolic byproduct formation, energy usage, growth rate, and growth yield as shown on the right of Fig. 1. Simulations can be modified by varying the inputs, altering any component of the network, or by specifying the objective (i.e., biomass, growth rate, and a specific metabolite). This general in silico modeling approach of building a genome-scale model, applying constraints to impose restrictions on the system, and then exploring feasible phenotypes is referred to as the constraint-based modeling (CBM) methodology [32]. This technology is useful because it generates predictions about the state, path, or outcome of metabolic processes. In the field of metabolic engineering, this can be applied as a quality control, as means of predicting improved solutions to fermentation problems, and for identifying feasibility of a particular process [33,34].These modeling methods have been used to predict gene essentiality under a variety of conditions [35], response or alterations in metabolism by changing or limiting certain reactions or pathways [36], and for predicting optimal growth [37]. Some of these predictions have been tested against experimental evidence both to test the quality of the model [37] and also to predict, successfully, the capacity of a cell to evolve a more optimal solution to growth [38].

Each of these characteristics of metabolic modeling demonstrates the applicability of applying predictive modeling solutions to real life issues in microbiology. However, there are many methods for modeling. For instance, one might model each reaction of cellular metabolism using the kinetic constants and constraints unique to each reaction [39]. In practice, kinetic modeling is extremely complex and is further limited because of the radically different milieu of the cell versus the biochemist's assay [39]. Descriptive pathway databases such as KEGG [40] compile huge amounts of information but these do not constitute models capable of running simulations that can in turn, generate testable hypotheses [41].

If the goal of modeling is to construct models of systems that exactly parallel and exactly predict the real metabolic network, the level of complexity will be too high for any model to currently achieve. If instead, a genome-scale network based on reaction stoichiometry were created, the number of possible solutions would appear to be too high to feasibly determine [32,42]. However, by imposing constraints, including realistic growth and biomass yields and finite inputs to the network, the models generate a more limited numbers of solutions. This constraint-based modeling [32,42] enables researchers to generate realistic solution sets that are testable experimentally. The conceptual construction of this process is shown in Fig. 1. In stoichiometric CBM, the computer program describes a matrix of the reactions of the reconstructed metabolic network. A variety of mathematical solutions, not discussed here, are applied to operate this matrix and to return results. If the model were unconstrained, an infinite number of possible outcomes could be generated. Such unbounded results might reflect infinite carbon and energy availability or other unlikely scenarios. To avoid this nonsense, and to create simulations that generate testable hypotheses, constraints to the model performance are introduced. These constraints may include limits on available substrates, known thermodynamic reversibility of reactions, and the required outputs of the cell in terms of predicted growth or energy yields. Clearly, any type of constraint that can be represented quantitatively can be introduced. For instance, individual reactions can be turned off to simulate gene deletion mutations or enzymatic inhibition. A map of a simplified cellular model and corresponding reaction network is presented in Fig. 2. This example depicts how a reaction (and corresponding) gene is essential, non-essential or how synthetic lethals may result. A reaction may be conditionally essential as in the example in Fig. 2, if component C was supplied, the lethality caused by deletion of reaction 1 (*Rxn*1) would be blocked and biomass production allowed. Alternatively, reactions can be limited or turned down to minimal flux values that still permit growth but otherwise represent perturbations of the network. Applying appropriate limitations or constraints generates testable hypothesis by running simulations and then comparing the results with the results of controls with 'wild-type' constraints [32,42]. In these myriad ways, CBM can mimic, and model, growth rate, mutation of genes to measure essentiality, prediction of how an impaired cell responds, production of metabolites, and measurement of optimality. In turn, this can test hypotheses; experimental data can be used to refine and correct models [37,43] and expression and metabolomic data can be used as constraints [32,44,45]. CBM is a mathematical way of constructing and running genome-scale networks that creates a matrix based on mass balance, known transporter (uptake) capacity, thermodynamic limits, and other physico-chemical constraints that can be incorporated into the system and are then used to generate data on the range of fluxes permitted through the metabolic network [32,42]. CBM allows building and testing of these models in an iterative way and, though not the only modeling method, has a long and successful record in the field of in silico modeling and computational systems biology [32,42]. The ability to apply almost any sort of constraint, including gene expression [43–45] or other regulatory features can be used to increase the concordance with reality of the model [44,46]. Critical to the success of a CBM model is accurate representation of the metabolic network and of the organism's biomass constituents [32].

Practical application of CBM has been made in the field of metabolic engineering. When the goal is to optimize synthesis of a metabolite through alterations of a microbial strain, the large number and complexity of the potential answers demand systematic means to generate the required strains. An example of this is in the metabolic engineering of *Escherichia coli* for improved lycopene production [47,48]. Stoichiometric flux balance analysis was used to predict the necessary gene knockouts that ultimately resulted in a 40% increase over an already high-producing parent [47]. These and other [49] studies indicate the facility of a form of CBM to predict a useful path for strain engineering but also

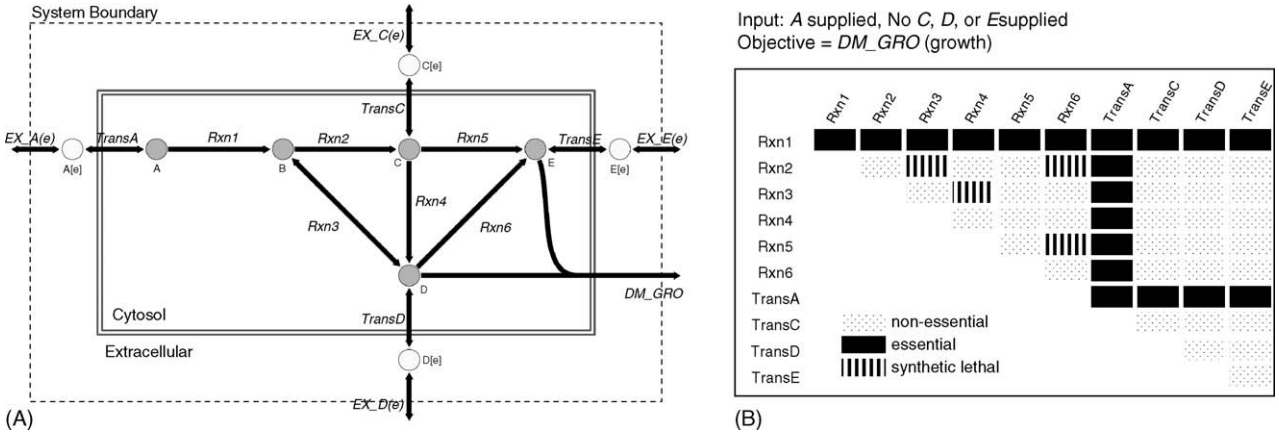

Fig. 2 – Metabolic modeling of reaction dispensability. A. Highly simplified model of a cellular metabolic network. Growth of the hypothetical cell is measured by the objective function "*DM_GRO*", consisting of biomass products D and E. In this network, transporters (*TransA*, etc.) for A, C, and D allow input of these extracellular medium components unless constrained. The intracellular network of reactions (*Rxn*1, 2, etc.) takes each component to produce the biomass (D and E). Constraints applied include extracellular medium (A supplied, no C, D, or E in this example), the direction of the reactions, and the output in the form of biomass. Due to the simplicity of this network, there are three possible results if one or more reactions or transporters are removed: non-essential, essential, or synthetic lethal. In the particular example shown, A is supplied but no C, D, or E are in the extracellular medium. The objective function is production of biomass (D and E) though in real networks, many more objectives are possible. B. Results of deleting reactions or transporters in the network given in A. Therefore, deletion of reaction 1 (*Rxn*1) or transporter A (*TransA*) would block biomass output in the form of *DM_GRO* and be deemed essential (or conditionally essential if supply of C were possible). Certain combinations of deletions are synthetic lethals (e.g., *Rxn*2 and *Rxn*3) because synthesis of a biomass constituent is blocked by removal of two reactions and not by a single reaction. Non-essential deletions result when loss does not block the biomass (objective function) synthesis either directly or indirectly. Imagining that A is a carbon and energy source such as glucose while C, D, and E are amino acids or other nutrients, and D and E are components of the cellular biomass. The constraints applied might represent minimal medium (A present but no C, D, or E), the thermodynamic reversibility of the reactions within the network, and output composition (biomass) whereas supply of all extracellular components, as with rich medium would substantially alter the results.

demonstrated that gene regulation and kinetic effects also may influence the results [47,48].

Already, such models have been employed in pathogens of interest to antibiotic developers including *Helicobacter pylori* [50,51], *Haemophilus influenzae* [46,52], *Staphylococcus aureus* [53], and *E. coli* [54–57]. These, and other models, have been utilized to predict gene essentiality [52,54,55,58], nutrient requirements [35,37,46,59], and evolution towards optimal growth [38,60–62]. The latter capacity may seem non-obvious but actually offers potential examples of how bypass or other compensatory changes in metabolism can evolve under the right conditions [61].

3. In silico essential gene prediction: antimicrobial target identification and prioritization

Over the last decade, computer models of the metabolism of several species of bacteria and yeast have been built [32] and some subsequently updated or rebuilt [56,57,63]. Retrospective validation of models is often by deletion of genes and comparison of predicted growth to 'real' growth. This is an independent validation of model performance because in silico models are not constructed from essentiality data, per se, but from knowledge of metabolism; if the metabolic knowledge is correct and complete then the predicted gene deletions will be concordant with the experimental work. If alternative or bypass pathways do exist and are not recorded in the model, then the model will fail to correctly predict the essentiality of a given gene or reaction. Alternatively, a model might overestimate the number of redundant reactions or pathways; this can be due to errors in annotation and unaccounted regulation. Equally true, each component of the reaction network may be present in the target organism but not expressed under the conditions to be examined.

Assuming that a model is complete enough to replicate the metabolism of the target organism, then it can be tested via simulation. For instance, if the objective function is the cellular constituents or biomass produced by the organism, then each gene in the model may be tested by deletion. Blocking production of biomass (i.e., growth) allows one to identify the reactions and associated genes that are essential for viability or growth under the substrate environment/nutrient conditions used. Therefore, within a metabolic model, one can predict the complement of essential genes, the complement of growth impairing genes, and the complement of non-essential genes computationally. Between 75%

and 89% of the predicted deletions were found correct via experimental confirmation [35,50,55,58,63]. In one of the very first of these efforts, Edwards and Palsson [55] deleted 79 genes of an *E. coli* metabolic model, 68 of these proved to be concordant with the experimental observations (86% performance). Subsequently, other models have performed similarly. For example, a large-scale deletion of 599 genes in a model of *Saccharomyces cerevisiae* metabolism had 87.8% of the deletions agreeing with experimental observations [58]. An expanded model of the gastric ulcer pathogen, *H. pylori*, correctly predicted 75% of gene deletions [63] while earlier versions of the same model were correct 69% of the time [50], showcasing the effects of updating and re-evaluating model content. A variety of phenotypes resulting from null mutations were correctly predicted in a *S. cerevisiae* model approximately 70–80% of the time [35]. While gene deletion studies can be performed with almost any model, a lack of published, high quality data on viability of gene knockouts can inhibit verifying model performance; this was true with a *S. aureus* model [53] because a complete table of gene deletions in this organism have yet to be published [18]. A very recent study constructed a flux balance analysis (FBA) model of the mycolic acid pathway in *Mycobacterium tuberculosis*, prediction of gene essentiality correlated about 79% with previous transposon-based gene mutations [64]. The authors concluded that FBA, coupled with conventional homology determinations against human sequences is a useful means of prioritizing anti-tuberculosis targets [64]. In all of these examples, gene deletions were typically performed to probe and validate the performance of the model rather than attempt to identify potential antibiotic targets. However, almost all of these deletion studies identified issues with gene annotation, expression, missing pathways, etc. that help improve both genomic annotation and metabolic models [63].

A complete model of metabolism for an organism under the chosen growth conditions might be used to predict the set of metabolic reactions (gene products) that are potential targets for antimicrobial intervention. While large sets of potential targets have been identified bioinformatically [13,19,20], or through various high-throughput essential gene screens [11–13,16–18,21], as we shall see, below, this can be complicated by various artifacts. If a model is incomplete, reliance on the prediction alone is risky because bypass pathways may exist. That is why this sort of target prediction will, if employed in the pharmaceutical industry, likely be used along with traditional methods of validation. However, prediction of metabolic gene essentiality still is an extremely powerful tool, this is because in silico modeling targets genes with defined reactions and can be used to assess many aspects of the cellular role of a reaction.

While it might be tempting to run simulations for a great many growth conditions, it is more likely that useful data regarding the set of essential genes can be gathered from conditions that simulate growth in vitro (as in antibiotic development) and growth in vivo (as in the host) assuming that the two differ significantly. In this manner, a model might identify the set of metabolic genes whose products are required for establishment or maintenance of virulence. While some of these gene products are likely to be in biosynthetic pathways for amino acids or other products limiting in the human host, others could be genes that play roles in synthesis of virulence determinants such as lipooligosaccharide in *H. influenzae* [65].

The performance of an in silico metabolic model can be compared to the set of known essential genes of an organism assuming that the latter is known with high confidence. At present, only *S. cerevisiae* and *Bacillus subtilis* have been subjected to thorough, gene-by-gene knockout performed to avoid artifacts that could generate false positives or false negatives [24,25]. However, *E. coli* has been subjected to partial gene-by-gene knockout over the course of ~60 years of work on this organism, collected at the "PEC" [66] website (http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp). Additionally, high-throughput transposon-based gene deletion of *E. coli* was performed recently [12]. Interestingly, the two data sets only partially overlap; 351 genes annotated by PEC as non-essential, were found to be high-throughput essential genes [12]. This conflict cannot be resolved solely by differences in strains or medium conditions. In fact, many of the 351 'PEC' non-essential genes found essential via mutagenesis have published references describing viable null alleles often with no discernable phenotypes (Trawick, data not shown). Therefore, more directed experimentation is necessary.

In *H. influenzae*, 27 genes involved in cell wall biosynthesis were individually deleted [67]. These were both essential and non-essential genes; 18 of which were in the *H. influenzae* model because their gene products partake in metabolic roles. The in silico gene deletion results using Genomatica's SimPheny™ program [33,42] (http://www.genomatica.com/index.shtml) found agreement with directed deletion 94% of the time (Table 1). Significantly, genes involved in cell wall synthesis are prime antibiotic targets [1]. While it is unlikely that many novel cell wall synthesis targets remain, these data, along with the historical performance (70–90%, see above) of in silico models in predicting essential genes, are very encouraging. While analysis of the updated SimPheny™ *H. influenzae* model is part of an ongoing research effort, it was found that, of genes predicted to be essential in the model and targeted by high-throughput transposon-mediated mutagenesis [14], 93% were deemed essential or required for optimal growth by

Table 1 – Very high agreement between SimPheny™ predictions and directed gene deletions in *H. influenzae*

	Experimental (Trepod and Mott [67])	In silico by SimPheny™	Agree experimental vs. in silico	Disagreement
Essential	14	13	12	1
Non-essential	15	5	5	0

Note: essential is defined as a gene required for growth using rich medium or simulation thereof.

Akerley [14]. When required to be conserved and essential in another organism, *E. coli*, the fraction of correct calls rose slightly to 95% (J.D. Trawick, data not shown). These results will require further validation, however. Correctly identifying an essential gene in this range may be far from flawless but, if coupled with low to medium confidence high-throughput data, could produce very high confidence prioritization of metabolic targets, lessening the need for expensive directed gene knockouts.

4. Optimum solutions, adaptive evolution, antibiotic resistance, and the prediction of response to drugs

When stoichiometric modeling was developed, practical uses of this technology have mostly been targeted towards metabolic engineering, strain development, and other uses in the realm of industrial microbiology. Considerable emphasis was placed on the performance of models relative to bacterial growth [37]. Could non-kinetic simulation methods such as flux balance analysis correctly predict growth and biomass production? FBA, a particular constraint-based modeling algorithm, and similar solutions are configured to generate optimal flux distributions, i.e., growth or biomass production, within their networks [32,68]. An in silico model can often accurately predict experimental growth of the target organism [37]. However, not all experimental conditions confirmed the in silico predictions [37,38]. Interestingly, when *E. coli* was grown on glycerol, the experimentally determined growth rate was sub-optimal compared to the in silico prediction [38] but when selected for maximal growth rate, evolved to reach the calculated optimum over a course of 700 generations. These observations have been repeated and confirmed in numerous ways [61,62]. To date, it is not clear that the optimum growth rate is always reached by mutation of metabolic genes or selection for expression changes within the set of metabolic genes. Indeed, non-metabolic processes such as changes in DNA supercoiling can increase fitness under selection for optimal growth probably by decreasing the time for one division cycle [69]. It will take some time to resolve some of these complexities.

This sort of adaptive evolution is highly dependent upon the ability to select for the desired trait, such as rapid growth rate. To predict antibiotic resistance, the situation would be more complex, requiring the researcher to identify a selective condition. This would also require that the antibiotic in question target a metabolic reaction found in the model. If that were true, then a model might predict if a bypass reaction or set of reactions were present. These would make the antibiotic target appear to be non-essential in the model but essential in vivo. Such a difference could arise from gene regulation or other constraints, and these could be subject to evolution studies. In a practical sense, it would appear that this is unlikely simply because the bulk of easily selected antibiotic resistance mutations tend to be in the target gene product, altering interaction with the drug [1]. Nonetheless, some more complicated conditions have been found for compounds under consideration for clinical development. For example, mutations giving rise to resistance in *E. coli* to compounds targeting the lpxC gene product were found in both lpxC and another metabolic gene, fabZ, with the possibility of other unidentified genes also mutating to resistance [70]. This system might be an ideal one in which to model choices that the cell might make, in principle, to evolve resistance.

FBA determines optimal solutions in metabolic simulations that appear to be accurate for predicting both the wild type growth capabilities of an organism as well as the growth capabilities of mutant strains following sufficient time for the mutant to evolve under growth selection [68]. There are other constraint-based modeling methods/algorithms that can be used to assess the immediate impact of gene/reaction knockout by attempting to calculate less optimal or sub-optimal solutions [71,72]. For instance, if a single reaction is blocked, perturbed or partially turned off, can the effects on the entire network be predicted? There are computational methods such as minimization of metabolic adjustment (MOMA) [71] that attempt to identify the minimum redistribution of flux in a knockout or knockdown from the wild type as a means to represent a homeostatic response. Another method called regulatory on/off minimization (ROOM) [72,73] utilizes different algorithms to approach a similar answer, the minimum number of reactions that need to change their flux in order to redistribute flux following a knockout or knockdown. Tests of these have demonstrated that MOMA was more predictive of growth immediately following deletion of gene function whereas FBA and ROOM were more accurate in predicting the endpoints of 'adaptively' evolved mutants [72]. These concepts of predicting growth rates of knockouts immediate following loss of function and after adaptive evolution are further illustrated in Fig. 3. By definition, FBA identifies the maximal possible growth rate. Hence, in silico knockouts predicted by FBA might have a higher confidence of pinpointing true lethals because FBA will find any and all reasonable bypasses within the existing network.

Collectively, these constraint-based methods enable the modeling of immediate responses to specific perturbations such as antibiotics. In antimicrobial research, the advantages of this sort of modeling should be apparent because it enables prediction of the results of cell based inhibitor assays and of fitness tests. These types of tests are attractive in drug screening or mode of action studies [26,27]. If MOMA correctly predicts the relative fitness of the mutant strain relative to the wild-type strain in each pair, then the frequency of the disabled strains at the end of the fitness test will correspond to the predicted fitness of each strain. Yet another method, multi-perturbation Shapley value analysis (MSA) [74] has been devised to facilitate this approach using multiple knockouts or knockdowns. A research goal of the authors is to test use of each of these modeling solutions for this kind of purpose. Until tested, this must remain speculative but there are several ways in which CBM modeling might be utilized. For instance, if an antibiotic has been proposed to act solely through inhibition of a specific metabolic enzyme, modeling inhibition of the presumed target gene would give both its theoretical fitness compared to uninhibited genes and also the effects of partially blocking the target gene on the network. The altered spectrum of fluxes might predict potential areas that are affected by the primary inhibition or indicate possible bypass pathways. In the case of the former, a list of potential

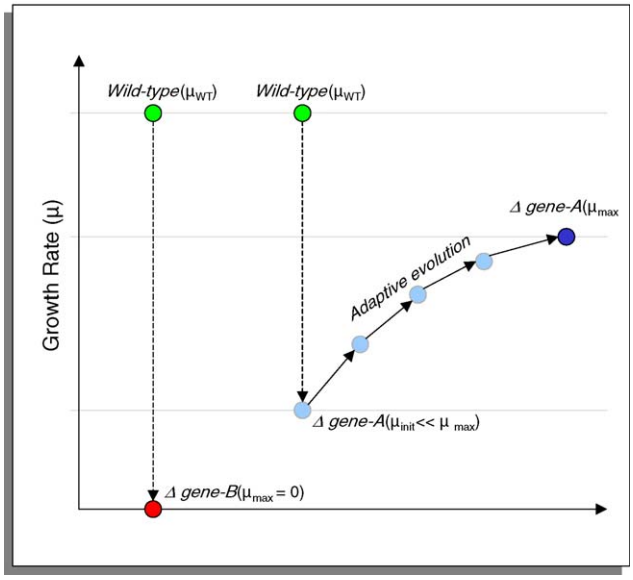

Fig. 3 – Adaptive evolution to reach optimal growth. This illustrates a hypothetical or typical example of adaptive evolution as described [38,60–62]. In this example, deletion of either one of two genes is made, if the deletion is not lethal (gene-A) but necessary for full growth, then growth relative to wild-type will be decreased, if FBA, which predicts optimal solutions, suggests that the network has more growth capacity than shown experimentally, this capacity might be realized through selection of gene-A deletion strains for optimal growth. Though, the figure illustrates a hypothetical example of gene-A deletions never reaching wild-type growth, both this and full return to wild-type have been selected for and observed [61]. Null mutations in other genes, such as gene-B may produce lethal phenotypes or lack capacity for evolution to more optimal growth.

secondary targets, weakened in a sense by having to carry more flux because of the inhibition of the primary target could be obtained. Discovering inhibitors to these, or pulling existing inhibitors off the shelf, might help identify means to increase potency via combination drugs to the primary target and also to the 'Achilles heel' secondary targets. Comparison of the model with the result of a large-scale fitness test such as the *S. cerevisiae* test described several years ago [26] could also pinpoint the reactions and gene products potential off-target effects of an antimicrobial drug under development. If these effects were linked to potential toxicity, then medicinal chemistry could be better focused on correction of the problem. Ultimately, a chemical class that might otherwise be dropped from development could be rescued for further drug development. At present, CBM has not been applied to operation of either cell based assays or fitness tests. How this would be done experimentally will depend partly on the nature of the experimental method. Some of these methods such as the *S. cerevisiae* fitness test provide a list of strains that are sensitive to the drug or antibiotic. Genes on such a list would be compared to the genes with altered flux in an idealized profile generated by the model. In this way, unexpected secondary effects might be pinpointed via comparison with the model. If the models failed to confirm the experimental results and validation of the model carried out, a list of candidate 'toxic' targets could be found. Though still hypothetical, this sort of profiling with either assays or fitness tests could provide idealized pictures of response to specific inhibitors to aid in overall assay performance. Currently, cell based, or biochemical assays are often empirically profiled against sets of known inhibitors, drugs, and antibiotics to evaluate assay performance (J.D. Trawick, data not shown—Elitra Pharmaceutical target-based antifungal assays were profiled in this manner in 2003–2004). Efficacy of any of these in silico methods in making this prediction would be strong evidence for use of CBM as a predictive and corrective tool for this type of assay.

CBM can model relative fitness. This was tested by comparing predictions from an *E. coli* SimPheny^{TM} model [56] with recently published experimental data comparing relative growth rate for *E. coli* strains differing in having either a NAD-dependent isocitrate dehydrogenase or a NADP-dependent isocitrate dehydrogenase [75]. The *E. coli* model was adjusted to mimic the genotypes of each of the tested strains. The FBA algorithm was used for the in silico experiment. The experimental data demonstrated [75] a 3–4% difference in growth rate with the in silico result almost identical to the experimental data (Table 2). If carried out sufficiently long, these very minor differences in growth rate would result in increasingly greater differences in cell numbers. Applied to situations in which the relative fitness differences were slightly greater, the differences in final cell numbers would be extreme. It will be interesting to determine if response to an antibiotic that targets metabolic gene products (trimethoprim, sulfonamides, mupirocin, and other drugs including tunicamycin) can be modeled in the same manner. There should be no impediment to this; inhibition of a gene product by limiting amounts of the drug could affect

Table 2 – SimPheny^{TM} accurately predicts relative fitness as expressed by small differences in growth rate

	Predicted growth rate ratio (icd^{NAD}/icd^{NADP})[a]	Experimentally observed relative growth rate ratio (icd^{NAD}/icd^{NADP})[b]
Growth on acetate with NAD dependent isocitrate dehydrogenase	0.965	0.96 ± 0.02
Growth on acetate with NADP dependent isocitrate dehydrogenase	0.997	0.998 ± 0.008

[a] Data from Radhakrishnan Mahadevan using SimPheny^{TM} FBA.
[b] Ref. [75].

fitness and thus monitored by following growth. As discussed above, when a large set of strains, each with a partial loss of a putative antimicrobial target gene are exposed to inhibitors such as antibiotics, then fitness of each strain will vary according to whether or not the target is affected by the drug [26]. The example in Table 2 shows, in principle, that specific effects on fitness can be measured for loss or partial loss of a single target. Actual experimental design might vary radically depending on the nature of the fitness assay employed. The ability to accurately predict subtle alterations in fitness as an outcome of growth in an assay could be applied in a number of areas including cell based assays for drug screening or the use of fitness tests in determining mode of action [26–28].

5. Summary

Within the last two decades, microbiological research has exploded from several developments including large-scale genome sequencing, revolutionary approaches in understanding bacterial pathogenesis, and the ability to accurately model the metabolism of microorganisms as computer programs generating testable hypotheses. The latter has been primarily confined to the bioengineering and metabolic engineering fields. Yet, all of the capabilities in modeling metabolism and growth that are useful in those fields are similar to methods that could be exploited in antimicrobial research. Identifying and prioritizing the essential metabolic genes constitutes proof-of-concept for broader application of modeling whether or not new targets are needed. Additionally, the metabolic reaction network is composed of gene products with well-defined substrates and products, making most of these reactions, in theory, druggable; new targets might be prioritized in this manner. However, metabolic modeling also has further potential in drug development. One possible use of CBM may be to identify candidate secondary targets, weakened in a sense by having to carry more flux because of the inhibition of the primary target. This could constitute a unique class of antimicrobial targets that cannot be readily discovered by other means. Further use of CBM may be in prediction of alterations in growth or metabolism and application as a model or control might be employed in development of cell-based assay and in mode-of-action determination.

Acknowledgements

We thank Dr. Radhakrishnan Mahadevan for generously supplying the data in Table 2 and for reading the manuscript. We also thank Dr. Iman Famili for her critical reading of the manuscript. The work in this manuscript was supported by SBIR grant R43 AI060216-01 from the NIAID to C.H.S.

REFERENCES

[1] Walsh C. Antibiotics: actions, origins, resistance. ASM Press; 2003.
[2] Bush K, Macielag M, Weidner-Wells M. Taking inventory: antibacterial agents currently at or beyond phase 1. Curr Opin Microbiol 2004;7:466–76.
[3] Payne DJ, Gwynn MN, Holmes DJ, Rosenberg M. Genomic approaches to antibacterial discovery. Methods Mol Biol 2004;266:231–59.
[4] Moir DT, Shaw KJ, Hare RS, Vovis GF. Genomics and antimicrobial drug discovery. Antimicrob Agents Chemother 1999;43:439–46.
[5] Projan SJ. Why is big pharma getting out of antibacterial drug discovery? Curr Opin Microbiol 2003;6:427–30.
[6] Projan SJ, Youngman PJ. Antimicrobials: new solutions badly needed. Curr Opin Microbiol 2002;5:463–5.
[7] Bush K. Antibacterial drug discovery in the 21st century. Clin Microbiol Infect 2004;10(Suppl. 4):10–7.
[8] Projan SJ, Shlaes DM. Antibacterial drug discovery: is it all downhill from here? Clin Microbiol Infect 2004;10(Suppl. 4):18–22.
[9] Overbye KM, Barrett JF. Antibiotics: where did we go wrong? Drug Discov Today 2005;10:45–52.
[10] Walsh C. Where will new antibiotics come from? Nat Rev Microbiol 2003;1:65–70.
[11] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, et al. Global transposon mutagenesis and a minimal mycoplasma genome. Science 1999;286:2165–9.
[12] Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, et al. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. J Bacteriol 2003;185:5673–84.
[13] Gerdes SY, Scholle MD, D'Souza M, Bernal A, Baev MV, Farrell M, et al. From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. J Bacteriol 2002;184:4555–72.
[14] Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, Mekalanos JJ. A genome-scale analysis for identification of genes for growth or survival of *Haemophilus influenzae*. Proc Natl Acad Sci USA 2003;99:966–71.
[15] Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ. Systematic identification of essential genes by in vitro mariner mutagenesis. Proc Natl Acad Sci USA 1998;95:8927–32.
[16] Ji Y, Woodnutt G, Rosenberg M, Burnham MK. Identification of essential genes in *Staphylococcus aureus* using inducible antisense RNA. Methods Enzymol 2002;358:123–8.
[17] Ji Y, Zhang B, Van Horn SF, Warren P, Woodnutt G, et al. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 2001;293:2266–9.
[18] Forsyth RA, Haselbeck RJ, Ohlsen KL, Yamamoto RT, Xu H, Trawick JD, et al. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. Mol Microbiol 2002;43:1387–400.
[19] Freiberg C, Wieland B, Spaltmann F, Ehlert K, Brotz H, Labischinski H. Identification of novel essential *Escherichia coli* genes conserved among pathogenic bacteria. J Mol Microbiol Biotechnol 2001;3:483–9.
[20] Zalacain M, Biswas S, Ingraham KA, Ambrad J, Bryant A, Chalker AF, et al. A global approach to identify novel broad-spectrum antibacterial targets among proteins of unknown function. J Mol Microbiol Biotechnol 2003;6:109–26.
[21] Haselbeck R, Wall D, Jiang B, Ketela T, Zyskind J, Bussey H, et al. Comprehensive essential gene identification as a platform for novel anti-infective drug discovery. Curr Pharm Des 2002;8:1155–72.
[22] Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 2003;48:77–84.

[23] Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 2001;98:12712–7.
[24] Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 2002;418:387–91.
[25] Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci USA 2003;100:4678–83.
[26] Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet 1999;21:278–83.
[27] Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, et al. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc Natl Acad Sci USA 2004;101:793–8.
[28] Ji Y, Yin D, Fox B, Holmes DJ, Payne D, Rosenberg M. Validation of antibacterial mechanism of action using regulated antisense RNA expression in *Staphylococcus aureus*. FEMS Microbiol Lett 2004;231:177–84.
[29] Gribbon P, Lyons R, Laflin P, Bradley J, Chambers C, Williams BS, et al. Evaluating real-life high-throughput screening data. J Biomol Screen 2005;10:99–107.
[30] Brideau C, Gunter B, Pikounis B, Liaw A. Improved statistical methods for hit selection in high-throughput screening. J Biomol Screen 2003;8:634–47.
[31] Palsson B. The challenges of in silico biology. Nat Biotechnol 2000;18:1147–50.
[32] Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–97.
[33] Smid EJ, van Enckevort FJ, Wegkamp A, Boekhorst J, Molenaar D, Hugenholtz J, et al. Metabolic models for rational improvement of lactic acid bacteria as cell factories. J Appl Microbiol 2005;98:1326–31.
[34] Smid EJ, Molenaar D, Hugenholtz J, de Vos WM, Teusink B. Functional ingredient production: application of global metabolic models. Curr Opin Biotechnol 2005;16:190–7.
[35] Famili I, Forster J, Nielsen J, Palsson BO. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA 2003;100:13134–9.
[36] Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 2003;84:647–57.
[37] Edwards JS, Ibarra RU, Palsson BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 2001;19:125–30.
[38] Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 2002;420:186–9.
[39] Visser D, van der Heijden R, Mauch K, Reuss M, Heijnen S. Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. Metab Eng 2000;2:252–75.
[40] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 1999;27:29–34.
[41] Famili I, Mahadevan R, Palsson BO. k-Cone analysis: determining all candidate values for kinetic parameters on a network scale. Biophys J 2005;88:1616–25.
[42] Price ND, Papin JA, Schilling CH, Palsson B. Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol 2003;21:162–9.
[43] Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. J Theor Biol 2001;213:73–88.
[44] Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature 2004;429:92–6.
[45] Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. J Biol Chem 2002;277:28058–64.
[46] Raghunathan A, Price ND, Galperin MY, Makarova KS, Purvine S, Picone AF, et al. In silico metabolic model and protein expression of *Haemophilus influenzae* strain Rd KW20 in rich medium. OMICS 2004;8:25–41.
[47] Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab Eng 2005;7:155–64.
[48] Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 2005;23:612–6.
[49] Pharkya P, Burgard AP, Maranas CD. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol Bioeng 2003;84:887–99.
[50] Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. Genome-scale metabolic model of *Helicobacter pylori* 26695. J Bacteriol 2002;184:4582–93.
[51] Price ND, Papin JA, Palsson BO. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res 2002;12:760–9.
[52] Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem 1999;274:17410–6.
[53] Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. BMC Microbiol 2005;5:8.
[54] Edwards JS, Palsson BO. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. BMC Bioinform 2000;1:1.
[55] Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 2000;97:5528–33.
[56] Reed JL, Vo TD, Schilling CH, Palsson B. *Escherichia coli* iJR904: An expanded genome-scale model of *E. coli* K-12. Genome Biol 2003;4:R54.1–2.
[57] Reed JL, Palsson BO. Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 2003;185:2692–9.
[58] Forster J, Famili I, Palsson BO, Nielsen J. Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*. OMICS 2003;7:193–202.
[59] Burgard AP, Vaidyaraman S, Maranas CD. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Prog 2001;17:791–7.
[60] Fong SS, Marciniak JY, Palsson B. Description and interpretation of adaptive evolution of *Escherichia coli* K12 MG1655 using a genome-scale in silico metabolic model. J Bacteriol 2003;185:6400–8.
[61] Fong SS, Palsson BO. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat Genet 2004;36:1056–8.
[62] Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, et al. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 2005;91:643–8.
[63] Thiele I, Vo TD, Price ND, Palsson BO. Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. J Bacteriol 2005;187:5818–30.

[64] Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. PLoS Comput Biol 2005;1:e46.
[65] Wong SM, Akerley BJ. Environmental and genetic regulation of the phosphorylcholine epitope of *Haemophilus influenzae* lipooligosaccharide. Mol Microbiol 2005;55: 724–38.
[66] Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol Microbiol 2005;55: 137–49.
[67] Trepod CM, Mott JE. Elucidation of essential and nonessential genes in the *Haemophilus influenzae* Rd cell wall biosynthetic pathway by targeted gene disruption. Antimicrob Agents Chemother 2005;49:824–6.
[68] Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 2002;4:133–40.
[69] Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. Genetics 2005;169:523–32.
[70] Clements JM, Coignard F, Johnson I, Chandler S, Palan S, Waller A, et al. Antibacterial activities and characterization of novel inhibitors of LpxC. Antimicrob Agents Chemother 2002;46:1793–9.
[71] Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 2002;99:15112–7.
[72] Shlomi T, Berkman O, Ruppin E. Constraint-based modeling of perturbed organisms: a ROOM for improvement. ISMB 2004; http://www.cs.tau.ac.il/~ruppin/ISMB_ROOM.pdf.
[73] Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci USA 2005;102:7695–700.
[74] Kaufman A, Kupiec M, Ruppin E. Multi-knockout genetic network analysis: the Rad6 example. CSB 2004; http://www.cs.tau.ac.il/~ruppin/csb2004.pdf.
[75] Zhu G, Golding GB, Dean AM. The selective cause of an ancient adaptation. Science 2005;307:1279–82.