GENOME-SCALE MODELS OF MICROBIAL CELLS: EVALUATING THE CONSEQUENCES OF CONSTRAINTS

Nathan D. Price, Jennifer L. Reed and Bernhard Ø. Palsson

Abstract | Microbial cells operate under governing constraints that limit their range of possible functions. With the availability of annotated genome sequences, it has become possible to reconstruct genome-scale biochemical reaction networks for microorganisms. The imposition of governing constraints on a reconstructed biochemical network leads to the definition of achievable cellular functions. In recent years, a substantial and growing toolbox of computational analysis methods has been developed to study the characteristics and capabilities of microorganisms using a constraint-based reconstruction and analysis (COBRA) approach. This approach provides a biochemically and genetically consistent framework for the generation of hypotheses and the testing of functions of microbial cells.

COBRA
(Constraint-based reconstruction and analysis). The overall philosophy and approach of applying constraints to limit the range of achievable functional (phenotypic) states of GENREs.

The theory of evolution is fundamental to the biological sciences. In essence, it states that organisms exist in particular environments that typically have scarce resources, and over time, the probability that the fit will survive is higher compared with the probability of survival of the less fit. To be fit for survival, a myriad of constraints must be satisfied, which limits the range of available phenotypes. Survival then depends on the best use of resources that will enhance the likelihood of survival subject to governing constraints. The closer an organism is to achieving a relatively optimal or fit function, the more likely it is to survive. Therefore, over a long period of time, survival is enhanced by an organism becoming more finely tuned to the environments that it experiences and by achieving better fitness within governing constraints.

All expressed phenotypes must satisfy the constraints imposed on the molecular functions of a cell. The living process must abide by physical laws, such as the conservation of mass and energy. Therefore, the identification and statement of constraints to define ranges of allowable phenotypic states provides a fundamental approach to enhance our understanding of biological systems that is consistent with our understanding of the operation and evolution of organisms. With the advent of whole-genome sequencing in the mid to late 1990s, the reconstruction of genome-scale networks for microorganisms became possible[1–3]. Some of the constraints under which these networks operate can be identified. A framework for constraint-based reconstruction and analysis (COBRA) has consequently arisen and has been successfully applied to study the possible phenotypes that arise from a genome. Because of its initial success, COBRA has attracted the attention of many investigators and has developed rapidly in recent years. Various *in silico* procedures aimed at determining the capabilities and characteristics of microorganisms have emerged over the past few years, and these procedures are forming the basis for *in silico* analysis of microorganisms. The plethora of methods that have developed within the COBRA framework is reviewed here.

Constraints on cellular functions

Different types of constraints limit cellular functions. Here, we briefly describe constraints in four categories: fundamental physico-chemical constraints, spatial or topological constraints, condition-dependent environmental constraints and regulatory or self-imposed constraints.

*Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA
Correspondence to B.Ø.P. e-mail: palsson@ucsd.edu*
doi:10.1038/nrmicro1023

Table 1 | **Genome-scale networks reconstructed to date**

Organism/organelle	Number of genes	Number of metabolites	Number of reactions	Year	Reference
Haemophilus influenzae	296	343	488	1999	45
Escherichia coli	660 904	436 625	720 931	2000 2003	91 19
Helicobacter pylori	291	340	388	2002	43
Saccharomyces cerevisiae	708 750	584 646	842 1,149	2003 2004	48 92
Geobacter sulfurreducens	588	541	523	2004	*
Mitochondria	N/A	230	189	2004	113

*R. Mahadevan, personal communication.

Physico-chemical constraints. Many physico-chemical constraints are found in a cell, and these constraints are inviolable and provide 'hard' constraints on cell functions. Mass, energy and momentum must be conserved. The contents of a cell are densely packed and form an environment where the viscosity can be about 100–1,000 times greater than that of water (see REF. 4 for compelling images). The diffusion rates of macromolecules inside a cell are generally slow and limiting[5,6], depending on molecule size. The confinement of many molecules in a semi-permeable membrane causes high osmolarity, and therefore cells require mechanisms for dealing with osmotic pressure (such as sodium–potassium pumps or a cell wall)[7–9]. Reaction rates are determined by local concentrations inside cells and might be limited by mass-transport. Enzyme-turnover numbers are generally less than 10^4 sec^{-1} and maximal reaction rates are equal to the turnover-number multiplied by the enzyme concentration[10]. Furthermore, biochemical reactions must result in a negative free-energy change to proceed in the forward direction.

Topobiological constraints. The crowding of molecules inside cells leads to topobiological, or three-dimensional, constraints that affect both the form and the function of biological systems[11–15]. For example, bacterial DNA is about 1,000 times longer than the length of a cell. DNA must be tightly packed in a cell without becoming entangled. For DNA to be functional, it must also be accessible for transcription; in *Escherichia coli*, DNA is organized and regulated so that there are spatio-temporal patterns[16]. Therefore, two competing needs — to be tightly packed but easily accessible — constrain the physical arrangement of DNA in the cell. As a further example, the ratio between the number of tRNAs and the number of ribosomes in an *E. coli* cell is typically only ten[17]. As there are 43 different types of tRNA, there is less than one full set of tRNAs per ribosome, indicating that it might be necessary to configure the genome so that rare codons are located close together[18]. Another example of interesting insights gained by analysing a cell's composition and size is that, at a pH of 7.6, *E. coli* would typically contain about 16 hydrogen ions[15]. This extremely low number indicates that it might be meaningless to discuss a cell's intracellular pH in terms of bulk averages, and that tracking the availability of hydrogen ions is crucial in the context of genome-scale models[19].

Environmental constraints. Environmental constraints on cells are time and condition dependent. Examples of environmental constraints are: nutrient availability, pH, temperature, osmolarity and the availability of electron acceptors. For example, *Helicobacter pylori* is constrained by its environment — the human stomach — to produce ammonia at a rate that will maintain its immediate surrounding at a pH that is sufficiently high to allow survival. Elemental nitrogen is needed to make ammonia, and therefore *H. pylori* has adapted by using amino acids instead of carbohydrates as its primary carbon source.

Environmental constraints are of fundamental importance for the quantitative analysis of microorganisms. Defined media and well-documented environmental conditions are needed to integrate data from various laboratories into quantitative models that are both accurately descriptive and predictive. Laboratory experiments with undefined media composition are often of limited use for quantitative *in silico* modelling.

Regulatory constraints. Regulatory constraints differ from the three categories discussed above as they are self-imposed and are subject to evolutionary change. For this reason, these constraints may be referred to as regulatory restraints, in contrast to 'hard' physico-chemical constraints and time-dependent environmental constraints. On the basis of environmental conditions, regulatory constraints allow the cell to eliminate suboptimal phenotypic states and to confine itself to behaviours of increased fitness. Regulatory constraints are implemented by the cell in various ways, including the amount of gene products made (transcriptional and translational regulation) and their activity (enzyme regulation).

Mathematical representations of constraints

After the recognition and definition of constraints, they need to be described mathematically. Once in a mathematical form, they can be used to perform *in silico* analysis.

Two fundamental types of constraints: balances and bounds. Constraints can generally be classified as either balances or bounds. Balances are constraints that are associated with conserved quantities, such as energy, mass, redox potential and momentum, as well as with phenomena such as solvent capacity, electroneutrality and osmotic pressure. Bounds are constraints that limit numerical ranges of individual variables and parameters such as concentrations, fluxes or kinetic constants.

The conservation of mass is an example of a balance constraint. At steady-state, there is no accumulation or depletion of metabolites in a metabolic network, so the rate of production of each metabolite in the network must equal its rate of consumption. This balance of fluxes can be represented mathematically as $S \bullet \mathrm{v} = 0$, where v is a vector of fluxes through the metabolic network and S is the stoichiometric matrix containing the stoichiometry of all reactions in the network[1,20]. Similar balance equations can be written for osmotic pressure[7,21], electroneutrality[22] and free energy around biochemical loops[23,24]. Balances result in equality constraints.

Bounds that further constrain the values of individual variables can be identified, such as fluxes, concentrations and kinetic constants. Upper and lower limits can be applied to individual fluxes ($\mathrm{v}_{\mathrm{min}} \leq \mathrm{v} \leq \mathrm{v}_{\mathrm{max}}$). For elementary (and irreversible) reactions, $\mathrm{v}_{\mathrm{min}} = 0$. Specific upper limits ($\mathrm{v}_{\mathrm{max}}$) that are based on enzyme capacity measurements are generally imposed on reactions. Concentrations must always be non-negative; so this constraint places a lower bound on concentration values. Upper bounds for concentrations can arise from solvent constraints and crowding. Similarly, kinetic constants also have constraints; they are constrained to be positive and have an upper bound that is based on collision frequency ($0 \leq \mathrm{k} \leq \mathrm{k}_{\mathrm{max}}$). Transmembrane potentials are limited to about 240–270 mV, as lipid bi-layers destabilize above this potential[25].

Constraining reconstructed networks defines achievable cellular functions. Taken together, both bound and balance constraints limit the allowable functional states of reconstructed networks. In mathematical terms, the range of allowable network states is described by a solution space that represents the phenotypic potential of an organism[26,27]. All allowable network states are contained in this solution space. If the balances and bounds are described by linear equations, then the solution space is a polytope in a high-dimensional space, allowing the use of CONVEX analysis[28]. If the constraints are bi-linear, such as those arising from mass-action kinetics of elementary association reactions, the solution space can be CONCAVE. There are now genome-scale network reconstructions (GENRE) for many microorganisms (TABLE 1). These GENREs correspond to biochemically and genetically structured databases that can represent multiple 'omics' data types. GENREs also form the basis for COBRA of a particular organism. Imposition of constraints on GENREs leads to a genome-scale model *in silico* (GEMS) of an organism that can be studied to define its capabilities and characteristics.

CONVEX SPACE
A convex space is one that satisfies the following condition: given any two points in the space, the line segment in between the points is completely contained in the space. Examples of convex objects include a square, triangle or circle.

CONCAVE SPACE
A space that is not convex. Examples of concave objects include a doughnut shape or a crescent.

GENRE
(Genome-scale network reconstruction). Applies to a particular organism, for example, GENRE of *Escherichia coli*. A GENRE contains a list of all the chemical transformations that take place in the particular network. These transformations can be represented stoichiometrically. These stoichiometric representations form a matrix, the rows of which represent the compounds, the columns of which represent the chemical transformations and the entries of which are the stoichiometric coefficients.

GEMS
Genome-scale models *in silico* of a particular organism, for example, GEMS of *E. coli*. The COBRA approach is used to analyse the properties of GENREs by assessing allowable functional states.

Tools for analysing network states

The analysis of microorganism phenotypic functions on a genome-scale using COBRA has developed rapidly in recent years[3,26]. Until recently, it has focused on the steady-state flux distributions through a reconstructed network, but is now being used to study all allowable concentration[29] and kinetic states (I. Famili and colleagues, unpublished observations). COBRA consists of two fundamental steps (FIG. 1): first, a GENRE is formed[1], and second, the appropriate constraints are applied to form the corresponding GEMS. In recent years, many new *in silico* methods have been developed using the COBRA framework. This plethora of methods can be broadly classified into the following categories: finding best or optimal states in the allowable range; investigating flux dependencies; studying all allowable states; altering possible phenotypes as a consequence of genetic variations; and defining and imposing further constraints. The following sections describe the development in each category.

Optimal or best states

Mathematical programming can be used to identify biochemical-reaction network states that maximize a particular network function (FIG. 2). The desired network function is defined and is described mathematically, which takes the form of an objective function (Z). Objective functions are generally formed for three primary purposes: first, the exploration of the phenotypic potential of a GENRE[30,31]; second, the determination of likely physiological states by choosing the objective function that represents probable physiological functions[32,33] such as biomass or ATP production; and third, the design of strains[34–36] to satisfy an engineering goal such as the improved production of a desired secreted product. Z can be either a linear or nonlinear function. Linear objective functions are used to maximize biomass ($\mathrm{Z} = \mathrm{v}_{\mathrm{biomass}}$) or ATP production ($\mathrm{Z} = \mathrm{v}_{\mathrm{ATP}}$), where the flux (v) through a reaction that drains biomass constituents or ATP is maximized. All the methods that are discussed in this category compute optimal phenotypes on the basis of assumed physiological objectives.

Single optima. Once the desired network function is defined, finding the best function of a GENRE is, in mathematical terms, a constrained optimization problem. Frequently, the constraints and the objective function are linear functions, so linear optimization or linear programming (LP) can be used[20,37,38]. The result from the linear optimization is a single network state (in the form of a flux distribution) that maximizes the chosen objective. Optimization is the basis for many of the other analysis methods described below (FIG. 1; icons 1–3, 5–7, 10–14).

The enumeration of optimal biochemical-reaction network states has been used to study several organisms including: *Saccharolytic clostridium*[39], *E. coli*[40–42], *H. pylori*[43], *Haemophilus influenzae*[44,45] and *Saccharomyces cerevisiae*[46–48]. The calculated optimal use of the metabolic network can be compared to experimental flux measurements[42,49] or to experimental phenotypic data[50,51]. LP calculates one optimal reaction network state. Interestingly, for genome-scale networks in particular, there can be multiple network states or flux distributions with the same optimal value of the objective function; therefore the need for enumerating alternate optima arises. The implication of this property is just

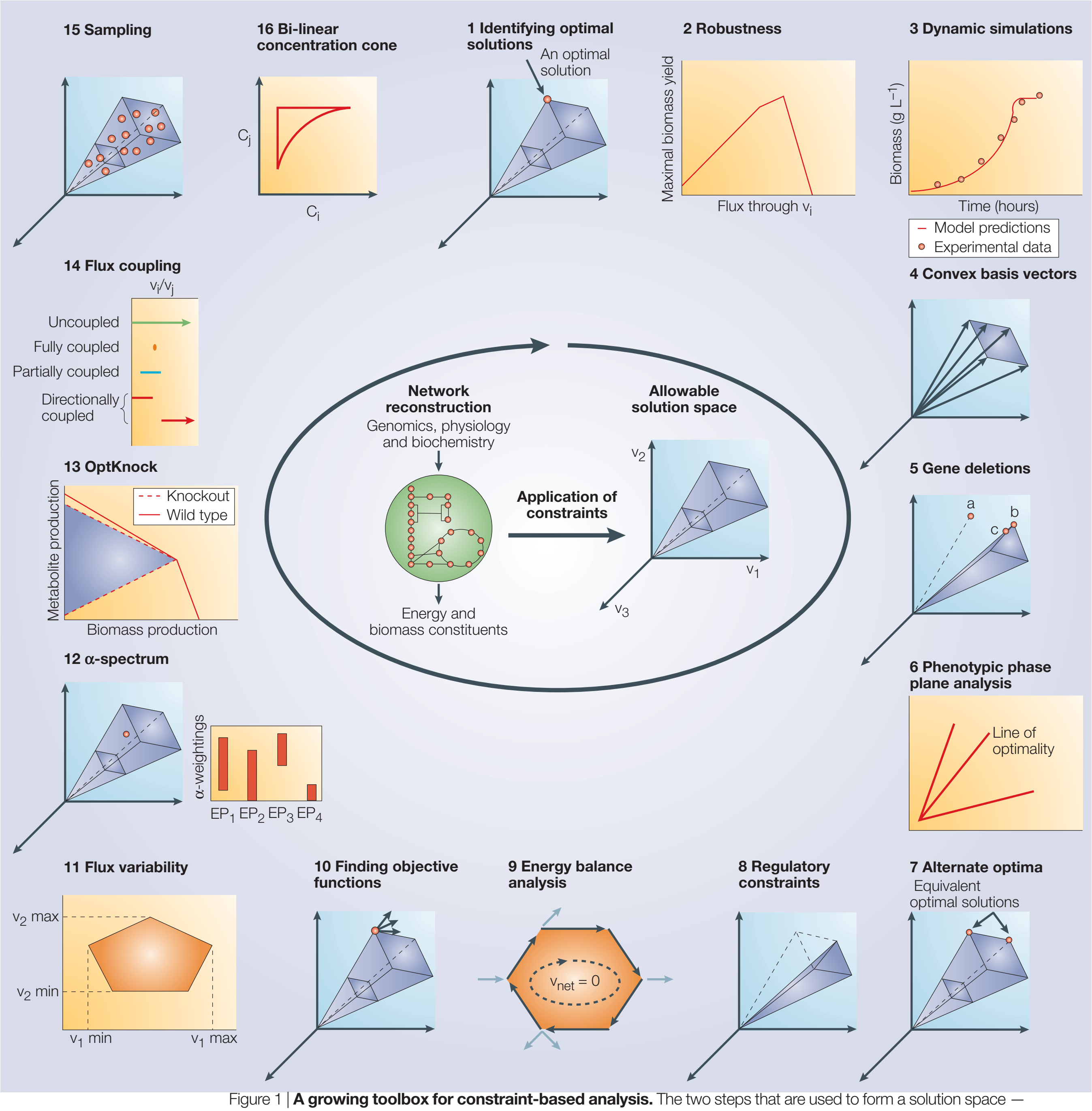

Figure 1 | **A growing toolbox for constraint-based analysis.** The two steps that are used to form a solution space — reconstruction and the imposition of governing constraints — are illustrated in the centre of the figure[1,20,37,111]. As indicated, several methods are being developed at various laboratories to analyse the solution space. The primary references for the methods indicated are: 1, REF. 40; 2, REFS 41, 61; 3, REFS 50, 99; 4, REFS 70, 71; 5, REFS 45, 49; 6, REFS 45, 62, 112; 7, REF. 55; 8, REF. 97; 9, REF. 23; 10, REF. 59; 11, REF. 58; 12, REF. 83; 13, REF. 35; 14, REF. 64; 15, REF. 85; 16, REF. 29. C_i, concentration of compound i; C_j, concentration of compound j; EP, extreme pathway; v_i, flux through reaction i; v_j, flux through reaction j; v_1, flux through reaction 1; v_2, flux through reaction 2; v_3, flux through reaction 3; v_{net}, net flux through loop.

beginning to be realized. A GENRE can reproduce the same function in many different ways. The mathematical notion of equivalent optimal states is coincident with the biological notion of silent phenotypes[52–54]. This property distinguishes *in silico* modelling in biology from that in the physico-chemical sciences where a single and unique solution is sought.

Alternate optima. Alternate flux distributions that lead to equivalent optimal network states are a property of genome-scale networks. The number of such alternate optima varies depending on the size of the metabolic network, the chosen objective function and the environmental conditions[55,56]. In general, the larger and more interconnected the network, the higher the number of

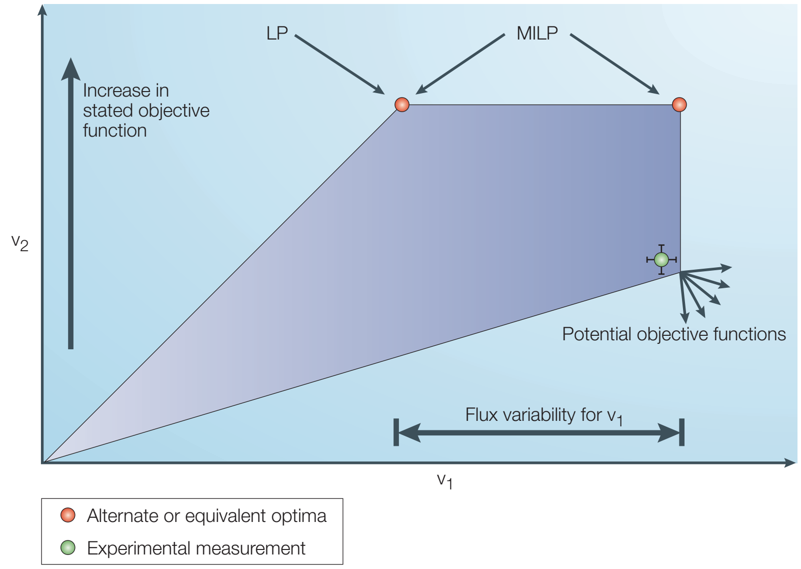
LP
MILP
Increase in stated objective function
v2
Potential objective functions
Flux variability for v1
v1
Alternate or equivalent optima
Experimental measurement

Figure 2 | **Determining optimal states.** If an objective is stated for a biochemical network, optimal solutions for the objective can be calculated. Linear programming (LP) will find one particular optimal solution, whereas mixed integer LP (MILP) can be used to find all of the basic (corner) optimal solutions. Flux variability analysis can be used to find ranges of values for all the fluxes in the set of alternate optima. In the figure, only v_1 is variable across the alternate optima. Conversely, if an objective function is not known for a biochemical network, experimental measurements can be used to identify potential objectives that would lead the cell towards that network state. v_1, flux through reaction 1; v_2, flux through reaction 2.

alternate optimal phenotypes. A recursive mixed-integer LP (MILP) algorithm has been developed to exhaustively enumerate all the alternate optima[55].

The method has been applied to study the central metabolic network in *E. coli*[55,57]. All the alternate optima were calculated and were then used to design NMR experiments for measuring *in vivo* intracellular fluxes[57]. In GENREs, there are several redundant pathways, which make the enumeration of all optima computationally challenging. Therefore, only a subset of alternate optima has been calculated for the *E. coli* GENRE[56] for various different minimal media conditions. To understand the complete set of alternate optima, flux variability analysis was developed to investigate their properties.

Flux variability. Flux variability analysis determines the full range of numerical values for each flux in the network, while still satisfying the given constraints and optimizing a particular objective[58]. The maximum value of the objective function is first computed (as described above), and is used as a further constraint on the network to ensure that only optimal network states are considered (see FIG. 2). Multiple optimizations are then carried out to calculate the maximum and minimum flux values through each reaction. Flux variability analysis can also be used to study the entire range of achievable cellular functions as well as the redundancy in optimal phenotypes[58].

Finding objective functions. The computation of an optimal network state requires a statement of the inferred — but in fact, unknown — cellular objective. The 'inverse' problem is to calculate all putative objective functions on the basis of measured intracellular fluxes to find the objective function(s) that, if used in an optimization (as described above), would result in a flux distribution that most approximates the measured flux distribution[59]. This approach was applied to a central metabolic model of *E. coli* for which there were experimentally measured flux distributions[59]. The calculated objective functions for aerobic and anaerobic growth were similar to biomass generation and to each other, indicating that one metabolic objective function can predict both aerobic and anaerobic flux distributions.

The four methods developed in this category focus on computing optimal network states. However, calculated optimal states may or may not be experimentally observed[32,49]. Therefore, other methods have been developed to study the range of achievable cellular functions of GENREs without optimizing an *a priori* stated objective.

REDUCED COST
A mathematical programming term; it is the smallest change in the objective function coefficient needed for a zero variable to become a non-zero variable.

LP to quantify flux dependencies

Optimization of GENRE functions has been used to obtain particular solutions, to perform sensitivity analysis and also to determine the relationship between reactions in the range of achievable network states (FIG. 3).

Single parameter perturbation: robustness calculations. The consequences of enzyme defects on functional states of GENREs can be determined. The value of the flux is simply constrained through the affected reaction and the optimal state is recomputed with this new constraint, which represents the enzyme defect. If the exact amount of reduction of enzymatic function is unknown, the flux can be sequentially changed through the reaction of interest and the objective function can be optimized at each step (FIG. 3). Plotting the resultant optimal value versus the flux value through the reaction of interest creates a curve that is piecewise linear. The slope in each of the linear regions of this curve represents the REDUCED COST[60]. The reduced cost describes the change in the objective function per unit change in the flux through the reaction of interest. Using this information, the robustness of the overall network function to the change in a flux through a particular reaction can be determined[60,61].

Such robustness analysis has been applied to an *E. coli* GENRE to analyse the impact of enzyme activity on growth rate[41,61]. Seven essential reactions were identified in central *E. coli* metabolism[61] by determining if the growth rate is zero when the activity of a reaction is zero. Additionally, the metabolic flux through transketolase and the tricarboxylic-acid-cycle reactions could be reduced to 15% and 19% of their optimal values, respectively, with no significant effect on predicted growth rate[60,61].

Variation of two parameters: phenotypic phase planes. Phenotypic phase planes (PhPPs) represent a method to perform sensitivity analysis as a function of two variables. They are used to visualize and characterize many optimal network states as a function of two fluxes of interest. Each solution can be characterized in terms

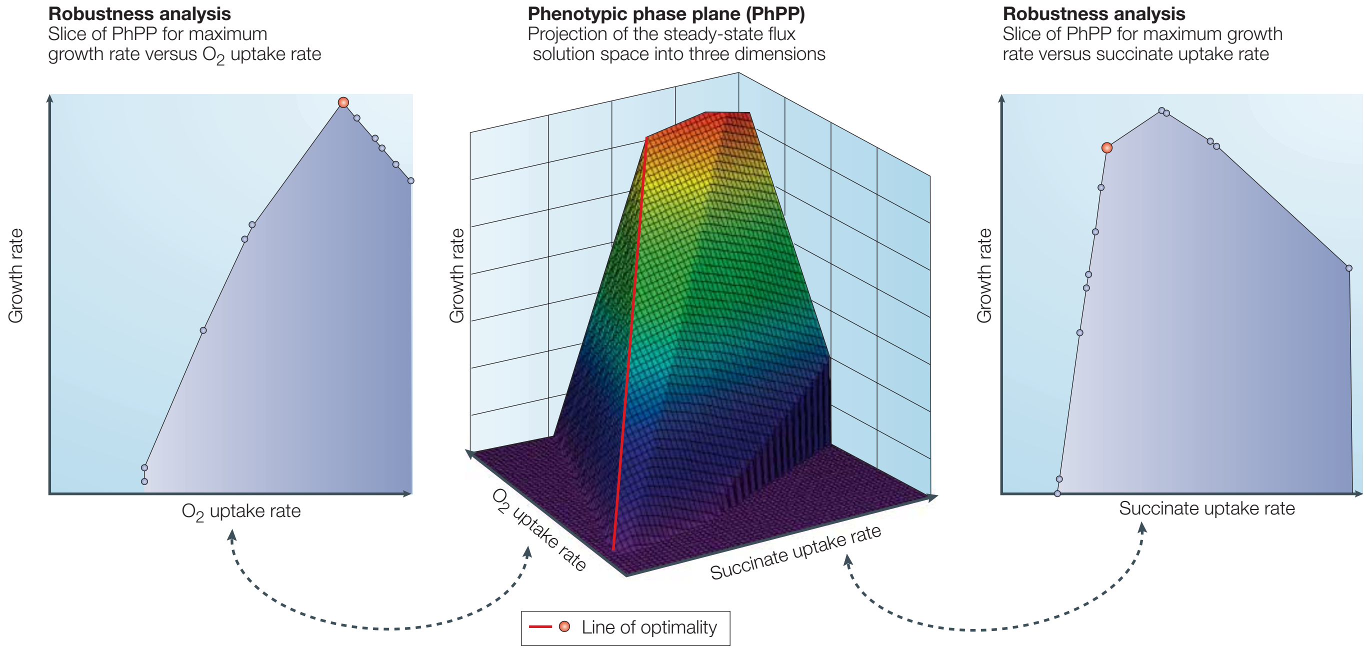

Figure 3 | **Flux dependencies.** The central figure represents a phenotypic phase plane (PhPP). It shows the maximum biomass production that is achievable at every possible combination of O_2 and succinate uptake rates. A phase plane is a projection of the solution space into two or three dimensions. The line of optimality corresponds to the conditions that are necessary for maximal biomass yield (g DW cell mmol^{-1} carbon source, where DW is dry weight). Robustness analysis of the two uptake rates is shown in the two side panels. The graph on the left shows the effect on growth rate of varying O_2 uptake at a fixed succinate uptake rate. Conversely, the graph on the right shows the effect on biomass generation of varying the succinate uptake rate at a fixed O_2 uptake rate. This figure was generated using SimPheny (Genomatica, Inc.).

of the SHADOW PRICES of the two fluxes that are being examined. A shadow price is the rate at which the objective value changes in response to an increase in the supply of a particular resource — in the case of a metabolic network, the resource would be a metabolite. The shadow price structure is finite in number, dividing the plane formed by the flux levels through the two reactions of interest into phases. Lines in the phase plane separate these phases, which correspond to changes in the shadow prices as described above. In other words, the incremental change in the value of the objective function that occurs owing to an incremental change in the availability of metabolites is different in each phase. The lines separating the phases can have particular designations, such as the line of optimality[62] that shows the conditions corresponding to the maximum biomass yield (g DW cell mmol^{-1} carbon source, where DW is dry weight). The phases in the PhPP can be categorized as: first, futile — an increase in either flux lowers the objective; second, single substrate limited — an increase in only one flux will increase the objective; or third, dual-substrate limited — an increase in either flux will increase the objective[33,62].

PhPPs are useful to interpret data and to design experiments[32,33,62,63]. For example, a PhPP analysis of *E. coli* growth on succinate and acetate showed that, under the conditions tested, *E. coli* grows on the line of optimality[33]. Also, the oxygen and glycerol uptake rates during the course of evolution can be traced out in a phase plane, showing that *E. coli* has reproducibly evolved to grow along the line of optimality from a suboptimal initial state[32]. Similar results have been obtained for the adaptive evolution of *E .coli* that is grown on lactate, but on pyruvate, the line of optimality is passed and *E. coli* grows as a partial anaerobe with a higher overall growth rate[87]. So, the GEMS of *E. coli* allowed the *a priori* prediction of the end point of an adaptive evolution[32].

Flux coupling finder. The flux coupling finder (FCF) was developed to analyse the relationship between fluxes at steady state in GEMs[64]. This approach uses an LP framework to minimize and maximize the ratio between all pairwise combinations of fluxes in a reaction network. Reaction pairs are found to be directionally coupled — if a non-zero flux for v_i implies a non-zero flux for v_j but not necessarily the reverse; partially coupled — if a non-zero flux for v_i implies a non-zero, but variable, flux for v_j and vice versa; or fully coupled — if a non-zero flux for v_i implies not only a non-zero but also a fixed flux for v_j and vice versa[64].

The FCF was used to analyse GENREs of *E. coli*, *S. cerevisiae*, and *H. pylori*[64]. The percentage of the reactions in each microorganism that was contained in coupled sets was ~60% for *H. pylori*, ~30% for *E. coli* and ~20% for *S. cerevisiae*. This percentage depends on the growth condition that is being considered, but is indicative of the flexibility of a network and the degrees of freedom available in the GENRE. The percentage of the total number of reactions found to be essential for aerobic growth on glucose-minimal medium was 59% for *H. pylori*, 28% for *E. coli* and 14% for *S. cerevisiae*.

SHADOW PRICE
A mathematical programming term; it is the rate at which the objective value changes by increasing the supply of a particular resource (for example, a metabolite).

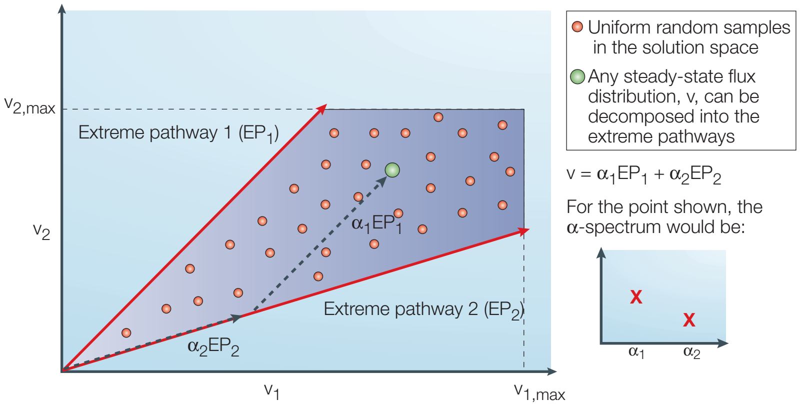
Uniform random samples in the solution space
Any steady-state flux distribution, v, can be decomposed into the extreme pathways
v = α_1EP$_1$ + α_2EP$_2$
For the point shown, the α-spectrum would be:
X
X
α_1
α_2
$v_{2,max}$
v_2
Extreme pathway 1 (EP$_1$)
α_1EP$_1$
Extreme pathway 2 (EP$_2$)
α_2EP$_2$
v_1
$v_{1,max}$

Figure 4 | **Characterizing the whole solution space.** The range of functions possible within the solution space can be characterized in two ways: first, through the definition of network-based pathways (such as the elementary modes and extreme pathways) or second, through the calculation of uniform random points within the space. The extreme pathways (EPs) are the edges of a convex space. Therefore, any point inside the space can be reached with a non-negative linear combination of the extreme pathways. In two dimensions, the decomposition of any point into two extreme pathways is unique, but in higher dimensions, the decomposition is generally non-unique. The range of possible weightings (α_1) on extreme pathways that can lead to a particular network state is called the α-spectrum. Uniform random sampling yields probability distributions for each flux based on the size and shape of the solution space and also provides a means for analysing the independence of different fluxes. v_1, flux through reaction 1; v_2, flux through reaction 2.

Characterization of all allowable phenotypes

Non-optimization based techniques have also been developed to study the full range of achievable biochemical network states that are provided by the solution spaces (FIG. 4). These methods do not look only at the properties of solutions selected by the statement of an objective, but at all the solutions in the space. The results are therefore not biased by a statement of an objective, but indicate properties of the GENRE as a whole.

Convex basis vectors: network-based pathway definitions. Network-based pathways, such as the extreme pathways or elementary modes, describe the full capabilities of a reconstructed biochemical network, and have already been reviewed extensively in the literature[65–69]. In mathematical terms, a solution space can be spanned by a set of basis vectors. So, every point in the space can be decomposed into a combination of the basis vectors. For convex polytopes, these basis vectors are referred to as extreme pathways[65,68]. The extreme pathways are edges of the convex solution space and therefore form a convex set of basis vectors. All possible network states in the solution space can be described by a non-negative linear combination of network-based pathways. Elementary modes[67,70] are a superset of the extreme pathways[71] — that is, combinations of extreme pathways. An elementary mode is a minimal set of enzymes that can operate at steady state — the enzymes are weighted by the relative flux that they need to carry for the mode to function[66]. There are useful and readily available software packages for computing elementary modes[72,73]. Unlike the LP-based methods discussed above, the use of network-based pathways for analysing GENREs is currently computationally difficult[74], and calculations of network-based pathways for genome-scale networks have only been achieved with limited input and output constraints[30,31].

Network-based pathways are useful for studying microorganisms. These pathways have been used to study the inherent redundancy in metabolic networks and have shown that there is more redundancy in the production of amino acids by the *H. influenzae* metabolic network than in the *H. pylori* metabolic network[30,75]. The robustness of an organism to gene deletions and changes in gene expression has also been studied using network-based pathways in central *E. coli* metabolism[76] and *Methylobacterium extorquens* metabolism[77]. Enzyme subsets (or correlated reaction subsets) have also been calculated using network-based pathways[78] and, in central *S. cerevisiae* metabolism, these enzyme subsets correlated to changes in gene expression during a diauxic shift[79]. Network-based pathways have also been used to assign functions to orphan genes based on metabolomic data[80] and to design strains[81]. Computing the singular value decomposition of extreme pathway matrices[75] has been used to quantify the magnitude of the problem of regulating a metabolic network[82]. Therefore, mathematical definitions of basis vectors have been useful to study the biological properties of the solution spaces that are formed in COBRA.

Decomposition of any steady state into extreme pathways. A steady-state flux distribution through a biochemical network can be described by non-negative linear combinations of extreme pathways. Because the extreme pathways form a convex set of basis pathways, the decomposition of a particular flux distribution into the extreme pathways is not unique[83]. This leads to a range of allowable weightings on the extreme pathways. By minimizing and maximizing the weighting on each pathway in the decomposition of a flux vector, the allowable ranges can conservatively be determined and are generally referred to as the α-spectrum[83].

The α-spectrum has been studied for human red blood cell metabolism, a skeleton representation of bacterial central metabolism[83] and central metabolism in *E. coli*[84]. For the skeleton network, it was shown that the incorporation of transcriptional regulatory rules significantly reduced the α-spectrum[83]. The α-spectrum was also used to assess the effects of adding additional flux constraints. Incorporating experimental flux measurements as constraints reduced the solution space, leading to a reduced α-spectrum[84]. Further work is needed to better define the α-spectrum in a less conservative manner, representing research opportunities in this field.

Uniform random sampling. The contents of a solution space can be studied by uniform random sampling of points throughout the space. Uniform random sampling is computationally tractable for genome-scale models[85]. Therefore, it is now possible to quantitatively characterize the full range of capabilities of genome-scale networks.

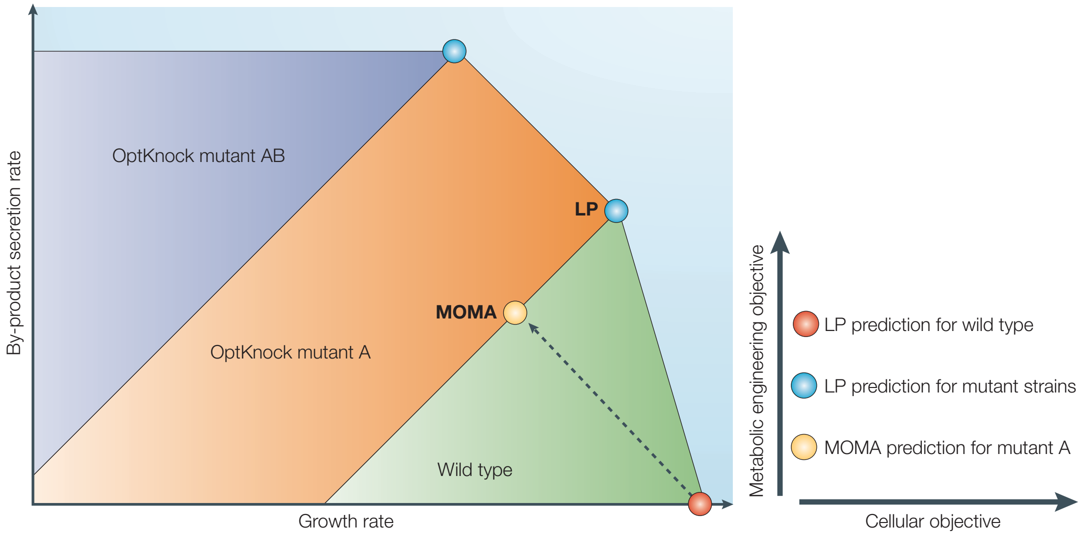
OptKnock mutant AB
OptKnock mutant A
LP
MOMA
Wild type
By-product secretion rate
Growth rate
Metabolic engineering objective
Cellular objective
LP prediction for wild type
LP prediction for mutant strains
MOMA prediction for mutant A

Figure 5 | **Altered solution spaces.** Solution spaces are altered by changes in the underlying biochemical network, such as occur with gene deletions. The projections of a wild-type solution space and two smaller knockout solution spaces are depicted. Optimization of growth rate (*x*-axis) in the wild-type solution space (red point) would not produce any by-product (*y*-axis), whereas optimization of growth rate in the two OptKnock mutant strains A and AB (blue points) finds solutions with by-product secretion. Minimization of metabolic adjustment (MOMA) — another method that is used for knockout predictions — assumes that, instead of being optimal for growth, the mutant will minimize the difference between its metabolic state and the metabolic state that is optimal for the wild-type strain (yellow dot). If the by-product is important, OptKnock can be used to identify knockout strains that couple optimal biomass production with by-product secretion. So, OptKnock identifies gene knockouts that require a cell to produce the desired by-product for optimal growth. In essence, the knockouts align the cell's objective with that of the metabolic engineer. Adapted with permission from REF. 35 © (2003) Wiley Interscience. LP, linear programming.

Uniform random sampling of the steady-state flux space has recently been used to study the properties of *E. coli* and human red blood cell metabolism. For example, this approach was used to show that, whereas low flux levels are common in *E. coli*, a high flux backbone exists that dominates metabolism[85]. This high flux backbone is related to the high fluxes in the principal eigenvector obtained from the singular value decomposition of a matrix of sampled network states, similar to the principal eigenvector of the extreme pathway matrix[75,82].

Pairwise correlation coefficients can be calculated between all reaction fluxes based on uniform random sampling. Perfectly correlated reactions ($R^2 = 1$) operate as functional modules within a biochemical network, whereas uncorrelated reactions ($R^2 \sim 0$) operate independently of each other. The degree of independence between reactions is an important consideration when choosing a set of fluxes to measure that will best determine the operating state of a biochemical network. Uniform random sampling has also been used to study the size and shape of steady-state flux spaces associated with red blood cell metabolism[86].

Uniform sampling provides an unbiased assessment of the impact of physico-chemical constraints on the achievable biochemical reaction network states. If a random set of points has been obtained for a solution space, then its segmentation by further constraints leads to exclusion of a subset of points of the full sample set. Therefore, one can readily determine which functional states are eliminated by the imposition of new constraints.

Altering phenotypic potential

Changes to the metabolic network, either adding or deleting reactions, can result in an enlarged or reduced range of achievable cellular functions, respectively (FIG. 5). The assessment of the functional states that have been removed or added provides information about the phenotypic consequences of genetic perturbations.

Gene additions and deletions. Gene additions and deletions modify the allowable states of a metabolic network. Regarding gene deletions, if there are no isozymes in the genome, the reactions associated with the deleted gene product are removed from the network. The effects of gene deletions can also be simulated by constraining the fluxes through the corresponding reactions to zero. Gene additions might result in new reactions in the metabolic network. Gene additions often result in an expansion of the wild-type solution space, whereas gene deletions often result in a reduction of the wild-type solution space. Optimization can be used to find an optimal phenotype for the mutant strain.

Wild-type strains can evolve towards optimal states within the range of allowable solutions[32,87] under the correct selection pressure. Since knockout strains cannot be expected to behave optimally in their functions, another approach, minimization of metabolic adjustment (MOMA), uses quadratic programming to find a point in the altered solution space that is closest to an optimal point in the wild-type solution space[49]. This closest point, where closeness is defined as the Euclidian distance, is usually not an optimal solution in the

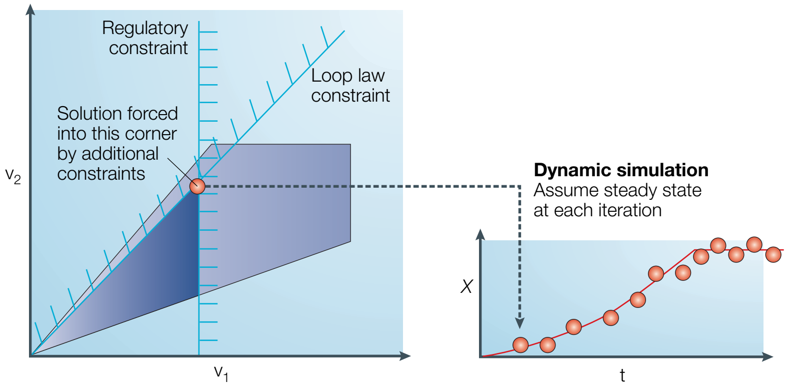
Regulatory constraint
Loop law constraint
Solution forced into this corner by additional constraints
v_2
v_1
Dynamic simulation
Assume steady state at each iteration
X
t

Figure 6 | **Additional constraints.** The successive addition of constraints shrinks the solution space, where the blue hatch lines indicate the infeasible side of the line. The figure illustrates how the addition of the thermodynamic loop law and regulatory constraints changes the solution space and the optimal growth-rate prediction (red dot). This predicted steady-state solution is calculated at each time point in a dynamic simulation. The steady-state solution is then used by way of a quasi-steady state assumption to approximate derivatives which are integrated over a small time set to describe system behaviour. t, time; v_1, flux through reaction 1; v_2, flux through reaction 2; *X*, biomass.

reduced space (FIG. 5). Optimizing in the reduced space will always result in an equal or better network performance than that predicted using MOMA. If a knockout strain is suboptimal, it can be evolved to achieve optimal functions[88].

Qualitative growth comparisons for knockout strains using computational predictions and experimental data — growth versus no growth — show good agreement with standard optimization, and improved agreement by using MOMA[49] or incorporating regulatory constraints[51,89]. Knockout studies have been performed in *E. coli*[49,51,76,89–91], *S. cerevisiae*[47,92,93], *H. pylori*[43], *H. influenzae*[45], and *M. extorquens*[77] with an accuracy rate of around 60–90%. Gene addition studies have also been performed to identify genes that could enhance the maximal production of amino acids in *E. coli*[94].

More recently, a gene deletion study investigated the reason for gene dispensability in yeast, to understand why a large fraction of yeast genes (80%) are not essential under normal laboratory conditions[92]. The authors found that many of these dispensable genes are required in other growth environments; others are compensated for by isozymes, and a smaller fraction is compensated for by alternative metabolic pathways. The authors also showed that a better explanation for the possession of multiple isozymes is the need for high flux rates through specific reactions, rather than the provision of redundancy for essential genes. Therefore, GEMS of GENREs are already proving to be useful to address fundamental biological questions.

Bi-level optimization. COBRA of metabolic and regulatory networks has been surprisingly successful in predicting the effects of gene deletions on growth. A computational procedure called OptKnock was developed to solve the reverse problem[95]; it identifies the gene deletions needed to generate a desired phenotype. In this approach, the desired phenotype will show an increase in biomass yield coupled to an increase in the production rate of a desired by-product. In other words, the cell will be able to grow faster only by producing more of the desired by-product. The resulting knockout strain will have significant by-product production at its maximal growth rate. The problem is formulated as a bi-level optimization problem and can be used on genome-scale models[35].

OptKnock has been used to identify gene knockout strategies for the production of intermediate metabolites (succinate and lactate)[35] as well as downstream metabolites (1,3-propandiol, chorismate, alanine, serine, aspartate and glutamate)[35,95]. These knockout strains would theoretically be stable strains that can produce metabolites in continuous culture, as increases in growth efficiency will only lead to higher by-product secretion rates. Initial experimental verification of these predictions has yielded promising results.

Application of additional constraints

The constraints typically imposed to form the flux solution space are flux-balance ($S \cdot v = 0$), enzyme capacity ($v_i \leq v_{max}$), and thermodynamic ($0 \leq v_{min}$). Frameworks for imposing other kinds of constraints have been developed (FIG. 6). These include transcriptional regulatory constraints[51,96,97], energy balance constraints[23,98] and slow dynamic change in the growth environment[50,51,99].

Regulatory constraints. Gene expression is dependent on an organism's growth environment. The regulation of gene expression might lead to the repression of enzyme synthesis and therefore the effective removal of a reaction from the network. To account for the effects of transcriptional regulation, a Boolean representation of the transcriptional regulatory network can be constructed[89,97]. With this framework, genes can only be found in two states, either expressed or not expressed. If the gene is not expressed, the enzyme will not be present in the cell and so the associated reaction is inactive ($v_i = 0$).

The imposed regulatory constraints further limit the phenotypic capabilities of an organism, and vary in a condition-dependent fashion[96]. Any of the previously described methods can be used to probe the reduced solution space. Constraint-based regulatory models have been built for central and genome-scale metabolic models of *E. coli*[51,89]. These combined regulatory and metabolic models now account for 1,010 genes. They have been used to predict growth phenotypes of knockout strains (with 79% accuracy compared with 65% when regulatory effects are ignored)[89], to predict changes in gene expression[51,89] and to simulate time courses of batch culture experiments[51]. Comparison with high-throughput data enabled the iterative development of the genome-scale metabolic and transcriptional regulatory model and led to hypotheses regarding the metabolic and transcriptional regulatory networks in *E. coli*[89]. The incorporation of regulatory constraints enhances the predictive capabilities of GEMS.

Energy balance analysis. The mass balance constraint, $S \cdot v = 0$, is analogous to Kirchhoff's first law for electrical circuits — which states that the sum of currents that enter a node must be balanced by the sum of currents that leave a node. Similar to Kirchhoff's second law — which states that the sum of voltage drops around a loop is zero — the sum of Gibbs free-energy changes around a loop in a biochemical reaction network must be zero[23,98]. These loops have been classified as type III extreme pathways[71], and network-based pathway analysis can be used to identify biochemical loops in genome-scale networks[98]. In agreement with thermodynamic principles, each reaction must have a negative Gibbs free-energy drop (bound constraint) in order to proceed, and the summation of the free-energy drops around a biochemical loop must be zero (balance constraint). To satisfy these two criteria, biochemical loops are constrained to have zero net flux. Therefore, the application of the loop law places a bound constraint on the fluxes in the network and results in a non-convex solution space, which will certainly be a subject of further investigation as this field moves forward.

Energy balance analysis (EBA) has been used to analyse a genome-scale model of *E. coli*[23]. In this study, the EBA constraint on the fluxes was imposed so that a set of free-energy drops that satisfy the loop law had to be present to allow a flux distribution through the network. The incorporation of the energy constraints resulted in a more tightly constrained range of allowable internal flux distributions, although these additional constraints did not affect the maximum growth prediction or the range of possible values of any of the exchange fluxes. It is important to consider the incorporation of the loop law when randomly sampling a phenotypic space[100] or when identifying multiple alternate optima[56].

Slow changes in the growth environment. The methods presented so far have all dealt with steady-state flux distributions in networks; however, these steady states can be used to model slower dynamic processes. The timescales associated with internal cellular processes may be much shorter than those associated with an organism's environment. Temporal decomposition leads to the assumption that cells are in an internal quasi-steady state relative to the dynamics of their environment. The quasi-steady state assumption (QSSA) can be used to approximate the time derivatives at each point, and a dynamic curve over longer time periods can be generated to simulate the dynamics of batch and fed-batch experiments[50,51]. QSSA assumptions are routinely used in several fields for many applications. For example, it is a standard procedure to apply the QSSA for the intermediates of enzymatic reactions in deriving enzymatic rate laws[101].

This approach has been used to simulate *E. coli* growth experiments in many different environments. Both regulated and unregulated constraint-based models have been used in dynamic simulations. For aerobic growth on glucose, both models predict acetate secretion and re-utilization; only the regulated model, however, is able to predict the time lag after glucose is consumed and before acetate begins to be reused[50,51]. Concentrations of by-products, such as acetate, formate and ethanol were also in quantitative agreement for anaerobic batch growth profiles[50,51].

Another method, dynamic flux balance analysis, includes additional constraints that limit the rate of change of internal fluxes in attempts to describe the dynamic change in intracellular flux levels[99]. This method was applied to study *E. coli* growth in a batch culture and was found to match experimental observations.

Future directions

The initial success of COBRA has spurred the development of the described methods in a few years. It is likely that this process will continue over the coming years. With a large number of constraints existing in nature and being self imposed by cells through regulation, it should be possible to continue to narrow in on physiological functions by identifying constraints and successively imposing them on GENREs, thereby limiting the range of allowable phenotypes.

Expanding the scope of reconstructed networks. Most of the existing GENREs are used for analysis of metabolism. However, any type of biochemical reaction network can be represented with a stoichiometric matrix, and the analysis tools discussed in this review can be applied to analyse its properties. For example, sufficient data to reconstruct signalling networks with a stoichiometric matrix are becoming available[102–107], as shown by the recent reconstruction of a stoichiometrically balanced JAK–STAT (Janus kinase and signal transducers and activators of transcription) signalling network in the human B cell[108]. Another type of network that can be reconstructed is the process of transcription and translation — this has already been completed for a prototypic network[109].

Other solution spaces. Most of the COBRA studies describe the range of allowable steady-state fluxes. However, several different mathematical subspaces are associated with the stoichiometric matrix that describes a network reconstruction. Four fundamental subspaces are associated with a stoichiometric matrix, each with significance for biochemical reaction network functions: first, the null space, which contains the feasible steady-state flux distributions; second, the left null space, which contains the metabolite conservation quantities; third, the row space, which contains the dynamic flux vectors; and fourth, the column space, which contains the metabolite time derivatives. Through mass-action kinetic representations, the space of possible concentration states or the range of possible kinetic values associated with measured experimental data can be studied. Initial studies on the kinetic and concentration spaces have been performed[29].

Alternate methods. The study of alternate solution spaces and additional constraints will lead to the definition of solution spaces formed by nonlinear constraints.

For example, most elementary reactions result from the interaction of two compounds. Therefore, the flux through these elementary reactions is described by mass action kinetics as $v_i = k_iC_AC_B$. The form of this equation leads to the definition of bi-linear constraints. This bilinearity means that the convexity of these phenotypic spaces will often be lost, necessitating the development and application of new analysis methods. For example, in a convex space, the finding of an optimal state always leads to the finding of the global optimum. However, in non-convex spaces, there are also local optima, and it becomes a more challenging problem to identify global optimal states or to uniformly sample high-dimensional solution spaces.

Practical applications. Industrial- and academic-grade software is now available for companies and academic laboratories to easily implement many of the types of analyses presented in this review (GAMS, Matlab, Mathematica, SimPheny). Constraint-based modelling[26] similar to that discussed in this review has the potential for important application to both metabolic engineering and drug discovery in the near future. There are recent articles that describe the use of constraint-based models for designing new microbial strains for metabolite production[35,36,81,95]. The use of these models to identify non-intuitive strain designs will advance the field of metabolic engineering. GENREs and GEMS are now becoming available for organisms that are important in bioremediation[110]. Constraint-based models can also be used to identify proteins that are essential for growth. These essential enzymes could serve as potential targets for the development of new antibiotics.

Constraint-based models have been constructed for several microorganisms, and have been useful for predicting and understanding phenotypic behaviour. Incorporation of new constraints reduces the available solution space and will increase the predictive capabilities of the models. Continued method development will be needed to further understand and better predict cellular behaviour.

1. Covert, M. W. *et al.* Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* **26**, 179–186 (2001).
2. Edwards, J. S., Covert, M. & Palsson, B. Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol.* **4**, 133–140 (2002).
3. Reed, J. L. & Palsson, B. O. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J. Bacteriol*. **185**, 2692–2699 (2003).
4. Goodsell, D. S. *The Machinery of Life* (Springer, New York, 1993).
5. Weisz, P. B. Diffusion and chemical transformation. *Science* **179**, 433–440 (1973).
6. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **181**, 197–203 (1999).
7. Werner, A. & Heinrich, R. A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary "*in vivo*" state and of time dependent variations under blood preservation conditions. *Biomed. Biochim. Acta* **44**, 185–212 (1985).
8. Hallows, K. & Knauf, P. in *Cellular and Molecular Physiology of Cell Volume Regulation* (ed. Strange, K.) 3–29 (CRC, Boca Raton, 1994).
9. Lew, V. L. & Bookchin, R. M. Volume, pH, and ion-content regulation in human red cells: analysis of transient behavior with an integrated model. *J. Membr. Biol*. **92**, 57–74 (1986).
10. Stryer, L. *Biochemistry* (Freeman, New York, 1988).
11. Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604 (2001).
12. Minton, A. P. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10577–10580 (2001).
13. Hall, D. & Minton, A. P. Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim. Biophys. Acta* **1649**, 127–139 (2003).
14. Ellis, R. J. & Minton, A. P. Cell biology: join the crowd. *Nature* **425**, 27–28 (2003).
15. Danchin, A. By way of introduction: some constraints of the cell physics that are usually forgotten, but should be taken into account for *in silico* genome analysis. *Biochimie* **78**, 299–301 (1996).
16. Huang, J., Zhang, Q. & Schlick, T. Effect of DNA superhelicity and bound proteins on mechanistic aspects of the Hin-mediated and Fis-enhanced inversion. *Biophys. J.* **85**, 804–817 (2003).
17. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell* (Sinauer Associates, Sunderland, Massachusetts, 1990).
18. Danchin, A., Guerdoux-Jamet, P., Moszer, I. & Nitschke, P. Mapping the bacterial cell architecture into the chromosome. *Philos. Trans. R. Soc. Lond. B* **355**, 179–190 (2000).
19. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. Ø. An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
20. Varma, A. & Palsson, B. O. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology* **12**, 994–998 (1994).
21. Brumen, M. & Heinrich, R. A metabolic osmotic model of human erythrocytes. *Biosystems* **17**, 155–169 (1984).
22. Marhl, M., Schuster, S., Brumen, M. & Heinrich, R. Modeling the interrelations between the calcium oscillations and ER membrane potential oscillations. *Biophys. Chem.* **63**, 221–239 (1997).
23. Beard, D. A., Liang, S. D. & Qian, H. Energy balance for analysis of complex metabolic networks. *Biophys. J.* **83**, 79–86 (2002).
Introduces the use of thermodynamic constraints to constraint-based analysis methods, resulting in better predictions of ranges of intracellular fluxes.
24. Qian, H., Beard, D. A. & Liang, S. D. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* **270**, 415–421 (2003).
25. Nicholls, D. G. & Ferguson, S. J. *Bioenergetics 3* (Academic, San Diego, California, 2002).
26. Price, N. D., Papin, J. A., Schilling, C. H. & Palsson, B. Ø. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.* **21**, 162–169 (2003).
27. Covert, M. W., Famili, I. & Palsson, B. Ø. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol. Bioeng.* **84**, 763–772 (2003).
28. Rockafellar, R. T. *Convex Analysis* (Princeton Univ. Press, Princeton, 1970).
29. Famili, I. & Palsson, B. Ø. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J.* **85**, 16–26 (2003).
30. Price, N. D., Papin, J. A. & Palsson, B. Ø. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome Res.* **12**, 760–769 (2002).
31. Papin, J. A., Price, N. D., Edwards, J. S. & Palsson, B. Ø. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.* **215**, 67–82 (2002).
32. Ibarra, R. U., Edwards, J. S. & Palsson, B. Ø. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189 (2002).
33. Edwards, J. S., Ibarra, R. U. & Palsson, B. Ø. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19**, 125–130 (2001).
34. Varma, A. & Palsson, B. Ø. Predictions for oxygen supply control to enhance population stability of engineered production strains. *Biotechnol. Bioeng.* **43**, 275–285 (1994).
35. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).
Presents a novel method for metabolic engineering by predicting knockout strains in which the objective of the metabolic engineer and the cell are coupled.
36. Liao, J. C., Hou, S. Y. & Chao, Y. P. Pathway analysis, engineering and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* **52**, 129–140 (1996).
37. Bonarius, H. P. J., Schmid, G. & Tramper, J. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends Biotechnol.* **15**, 308–314 (1997).
38. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14**, 491–496 (2003).
References 37 and 38 are well-written reviews that provide an introduction to flux balance analysis, one of the most common constraint-based modelling methods.
39. Papoutsakis, E. T. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnol. Bioeng.* **26**, 174–187 (1984).
40. Majewski, R. A. & Domach, M. M. Simple constrained optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng.* **35**, 732–738 (1990).
41. Varma, A. & Palsson, B. Ø. Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. *J. Theor. Biol.* **165**, 503–522 (1993).
42. Pramanik, J. & Keasling, J. D. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **56**, 398–421 (1997).
43. Schilling, C. H. *et al.* Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* **184**, 4582–4593 (2002).
44. Raghunathan, A. *et al. In silico* metabolic model and protein expression of *Haemophilus influenzae* strain Rd KW20 in rich medium. *OMICS* **8**, 25–41 (2004).
45. Edwards, J. S. & Palsson, B. Ø. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
46. Famili, I., Forster, J., Nielsen, J. & Palsson, B. Ø. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA* **100**, 13134–13139 (2003).
47. Forster, J., Famili, I., Palsson, B. Ø. & Nielsen, J. Large-scale evaluation of *in silico* gene knockouts in *Saccharomyces cerevisiae*. *OMICS* **7**, 193–202 (2003).
48. Forster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).

49. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15112–15117 (2002).
Presents a new method for predicting metabolic flux distributions of knockout strains, and shows that the predictions matched experimental data better than flux balance analysis.
50. Varma, A. & Palsson, B. Ø. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60**, 3724–3731 (1994).
51. Covert, M. W. & Palsson, B. Ø. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28058–18064 (2002).
52. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol.* **19**, 45–50 (2001).
53. Thorneycroft, D., Sherson, S. M. & Smith, S. M. Using gene knockouts to investigate plant metabolism. *J. Exp. Bot.* **52**, 1593–1601 (2001).
54. Bouche, N. & Bouchez, D. Arabidopsis gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117 (2001).
55. Lee, S., Phalakornkule, C., Domach, M. M. & Grossmann, I. E. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comp. Chem. Eng.* **24**, 711–716 (2000).
First use of MILP to identify alternate equivalent optimal flux distributions in metabolic networks.
56. Reed, J. L. & Palsson, B. Ø. Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res.* **14**, 1797–1805 (2004).
57. Phalakornkule, C. *et al.* A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. *Metab. Eng.* **3**, 124–137 (2001).
58. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
Develops the method of flux variability analysis to study the effects that optimal and suboptimal solutions have on the outcome of MOMA calculations and to identify equivalent pathways in metabolic networks.
59. Burgard, A. P. & Maranas, C. D. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol. Bioeng.* **82**, 670–677 (2003).
60. Varma, A., Boesch, B. W. & Palsson, B. Ø. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **59**, 2465–2473 (1993).
61. Edwards, J. S. & Palsson, B. Ø. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* **16**, 927–939 (2000).
62. Edwards, J. S., Ramakrishna, R. & Palsson, B. Ø. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* **77**, 27–36 (2002).
63. Kauffman, K. J., Pajerowski, J. D., Jamshidi, N., Palsson, B. Ø. & Edwards, J. S. Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophys. J.* **83**, 646–662 (2002).
64. Burgard, A. P., Nikolaev, E. V., Schilling, C. H. & Maranas, C. D. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* **14**, 301–312 (2004).
65. Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. Ø. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28**, 250–258 (2003).
66. Schuster, S., Fell, D. A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol.* **18**, 326–332 (2000).
A nice introduction to the definition and uses of elementary modes for analysing biochemical networks.
67. Schuster, S., Dandekar, T. & Fell, D. A. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17**, 53–60 (1999).
68. Schilling, C. H., Schuster, S., Palsson, B. Ø. & Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15**, 296–303 (1999).
69. Papin, J. A. *et al.* Comparison of network-based pathway analysis methods. *Trends Biotechnol.* **22**, 400–405 (2004).
70. Schuster, S. & Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182 (1994).
71. Schilling, C. H., Letscher, D. & Palsson, B. Ø. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* **203**, 229–248 (2000).
72. Klamt, S., Stelling, J., Ginkel, M. & Gilles, E. D. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* **19**, 261–269 (2003).
73. Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J. C., Montero, F. & Schuster, S. METATOOL: for studying metabolic networks. *Bioinformatics* **15**, 251–257 (1999).
74. Klamt, S. & Stelling, J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236 (2002).
75. Price, N. D., Reed, J. L., Papin, J. A., Famili, I. & Palsson, B. Ø. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophys. J.* **84**, 794–804 (2003).
76. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193 (2002).
77. Van Dien, S. J. & Lidstrom, M. E. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C(3) and C(4) metabolism. *Biotechnol. Bioeng.* **78**, 296–312 (2002).
78. Papin, J. A., Price, N. D. & Palsson, B. Ø. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res.* **12**, 1889–1900 (2002).
79. Schuster, S., Klamt, S., Weckwerth, W., Moldenhauer, F. & Pfeiffer, T. Use of network analysis of metabolic systems in bioengineering. *Bioprocess Biosyst. Eng.* **24**, 363–372 (2002).
80. Forster, J., Gombert, A. K. & Nielsen, J. A functional genomics approach using metabolomics and *in silico* pathway analysis. *Biotechnol. Bioeng.* **79**, 703–712 (2002).
81. Carlson, R., Fell, D. & Srienc, F. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.* **79**, 121–134 (2002).
82. Price, N. D., Reed, J. L., Papin, J. A., Wiback, S. J. & Palsson, B. Ø. Network-based analysis of metabolic regulation in the human red blood cell. *J. Theor. Biol.* **225**, 185–194 (2003).
83. Wiback, S. J., Mahadevan, R. & Palsson, B. Ø. Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum. *J. Theor. Biol.* **224**, 313–324 (2003).
84. Wiback, S. J., Mahadevan, R. & Palsson, B. Ø. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. *Biotechnol. Bioeng.* **86**, 317–331 (2004).
85. Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N. & Barabasi, A. L. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843 (2004).
First paper to perform uniform random sampling of the steady-state flux space to analyse the organization of genome-scale metabolic fluxes.
86. Wiback, S. J., Famili, I., Greenberg, H. J. & Palsson B. Ø. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* **228**, 437–447 (2004).
87. Fong, S. S., Marciniak, J. Y. & Palsson, B. Ø. Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale *in silico* metabolic model. *J. Bacteriol.* **185**, 6400–6408 (2003).
88. Fong, S. S. & Palsson, B. Ø. Metabolic gene deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genet.* **36**, 1056–1058 (2004).
89. Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. Ø. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96 (2004).
90. Edwards, J. S. & Palsson, B. Ø. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* **1,** 1 (2000).
91. Edwards, J. S. & Palsson, B. Ø. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97**, 5528–5533 (2000).
92. Duarte, N. C., Herrgard, M. J. & Palsson, B. Ø. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309 (2004).
93. Papp, B., Pal, C. & Hurst, L. D. Metabolic network analysis of the causes and elution of enzyme dispensability in yeast. *Nature* **429**, 661–664 (2004).
Insightful use of GEMS to demonstrate that the presence of isozymes is better explained by the need for a high flux rate through a reaction, rather than by providing redundancy for an essential function. Explains why a high degree of genes are found to be non-essential under laboratory conditions.
94. Burgard, A. P. & Maranas, C. D. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* **74**, 364–375 (2001).
95. Pharkya, P., Burgard, A. P. & Maranas, C. D. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* **84**, 887–899 (2003).
96. Covert, M. & Palsson, B. Ø. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.* **221**, 309–325 (2003).
97. Covert, M. W., Schilling, C. H. & Palsson, B. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88 (2001).
98. Price, N. D., Famili, I., Beard, D. A. & Palsson, B. Ø. Extreme pathways and Kirchhoff's second law. *Biophys. J.* **83**, 2879–2882 (2002).
99. Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **83**, 1331–1340 (2002).
100. Price, N. D., Schellenberger, J. & Palsson, B. Ø. Uniform sampling of steady state flux spaces: means to design experiments and to interpret enzymopathies. *Biophsy. J.* (In the press).
101. Segel, I. H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady–State Enzyme Systems* (Wiley, New York, 1975).
102. Gilman, A. G. *et al.* Overview of the alliance for cellular signaling. *Nature* **420**, 703–706 (2002).
103. Zhu, H. & Snyder, M. "Omic" approaches for unraveling signaling networks. *Curr. Opin. Cell Biol.* **14**, 173–179 (2002).
104. Graves, P. R. & Haystead, T. A. A functional proteomics approach to signal transduction. *Recent Prog. Horm. Res.* **58**, 1–24 (2003).
105. Li, J. *et al.* The molecule pages database. *Nature* **420**, 716–717 (2002).
106. Sivakumaran, S., Hariharaputran, S., Mishra, J. & Bhalla, U. S. The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **19**, 408–415 (2003).
107. Walhout, A. J. *et al.* Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**, 1952–1958 (2002).
108. Papin, J. A. & Palsson, B. O. The JAK–STAT signaling network in the human B-cell: an extreme signaling pathway analysis. *Biophsy. J.* **87**, 37–46 (2004).
109. Allen, T. E. & Palsson, B. Ø. Sequenced-based analysis of metabolic demands for protein synthesis in prokaryotes. *J. Theor. Biol.* **220**, 1–18 (2003).
110. Lovley, D. R. Cleaning up with genomics: applying molecular biology to bioremediation. *Nature Rev. Microbiol.* **1**, 35–44 (2003).
111. Edwards, J. S. & Kauffman, K. J. Biochemical engineering in the 21st century. *Curr. Opin. Biotechnol.* **14**, 451–453 (2003).
112. Schilling, C. H., Edwards, J. S., Letscher, D. & Palsson, B. Ø. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **71**, 286–306 (2000).
113. Vo, T. D., Greenberg, H. J. & Palsson, B. Ø. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J. Biol. Chem.* **279**, 39532–35940 (2004).

Acknowledgements

The authors would like to acknowledge funding from the United States National Institutes of Health and the National Science Foundation.

Competing interests statement

The authors declare competing financial interests: see Web version for details.

Online links

FURTHER INFORMATION

Bernhard Palsson's laboratory: http://systemsbiology.ucsd.edu/
GAMS: http://www.gams.com
Matlab: http://www.mathworks.com
Mathematica: http://www.wolfram.com
SimPheny: http://www.genomatica.com
Access to this links box is available online.