Journal of Biotechnology 261 (2017) 221–228

Contents lists available at ScienceDirect

Journal of Biotechnology

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

Use of *CellNetAnalyzer* in biotechnology and metabolic engineering

Axel von Kamp, Sven Thiele, Oliver Hädicke, Steffen Klamt*

Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany

ARTICLE INFO

Keywords:
Metabolic networks
Metabolic flux analysis
Flux balance analysis
Elementary modes
Strain optimization
Minimal cut sets

ABSTRACT

Mathematical models of the cellular metabolism have become an essential tool for the optimization of biotechnological processes. They help to obtain a systemic understanding of the metabolic processes in the used microorganisms and to find suitable genetic modifications maximizing the production performance. In particular, methods of stoichiometric and constraint-based modeling are frequently used in the context of metabolic and bioprocess engineering. Since metabolic networks can be complex and comprise hundreds or even thousands of metabolites and reactions, dedicated software tools are required for an efficient analysis. One such software suite is *CellNetAnalyzer*, a MATLAB package providing, among others, various methods for analyzing stoichiometric and constraint-based metabolic models. *CellNetAnalyzer* can be used via command-line based operations or via a graphical user interface with embedded network visualizations. Herein we will present key functionalities of *CellNetAnalyzer* for applications in biotechnology and metabolic engineering and thereby review constraint-based modeling techniques such as metabolic flux analysis, flux balance analysis, flux variability analysis, metabolic pathway analysis (elementary flux modes) and methods for computational strain design.

1. Introduction

Industrial (white) biotechnology uses cells or parts of cells for the production of chemicals, biofuels, pharmaceuticals, nutraceuticals, enzymes or other industrially relevant products. The design and optimization of biotechnological processes usually involves genetic modifications in the metabolism of the used production organisms to maximize their production performance. For the targeted (rational) metabolic engineering of cell factories, mathematical modeling of the cellular metabolism has become an essential tool. Various theoretical methods of metabolic modeling have been developed to analyze the capabilities of metabolic networks, to study the behavior of the metabolism under different growth and production conditions, to discover potential bottlenecks and to eventually identify targets for genetic modifications redirecting metabolic fluxes to a desired compound. In particular, methods of stoichiometric and constraint-based modeling have been successfully applied in metabolic and bioprocess engineering (Gutierrez and Lewis, 2015; King et al., 2015; Maia et al., 2015; Simeonidis and Price, 2015; Kim et al., 2015; Machado and Herrgard, 2015). These methods include, for example, metabolic flux analysis (MFA; characterization of metabolic fluxes under controlled conditions), flux balance analysis (FBA; analysis of optimal flux distributions), flux variability analysis (FVA; analysis of feasible ranges of metabolic fluxes), metabolic pathway analysis (discovery and analysis of metabolic pathways) and methods for computational strain design (computation of metabolic engineering strategies optimizing the production behavior of the organism).

Since stoichiometric and constraint-based metabolic models may involve hundreds or even thousands of metabolites and reactions, dedicated software tools are required to support an efficient analysis of (up to genome-scale) metabolic networks. Accordingly, several software packages for constraint-based modeling have been developed in the past years, including, for example, the COBRA toolbox (Schellenberger et al., 2011; Ebrahim et al., 2013), OPTFLUX (Rocha et al., 2010), OMIX (Droste et al., 2011, 2013), MUFINS (Wu et al., 2016), RAVEN (Agren et al., 2013) and *CellNetAnalyzer* (Klamt et al., 2007). COPASI (Hoops et al., 2006), presented in another article in this special issue, also supports analysis of basic stoichiometric features of metabolic networks but focuses more on kinetic modeling of biochemical systems.

CellNetAnalyzer is a package for MATLAB providing various (partially unique) algorithms for analyzing structure and function of biological networks. Metabolic networks can be studied based on stoichiometric and constraint-based models whereas signaling and

Abbreviations: CNA, *CellNetAnalyzer*; SCBM, Stoichiometric and constraint-based modeling; CR(s), Conservation relation(s); MFA, Metabolic flux analysis; FBA, Flux balance analysis; FVA, Flux variability analysis; EFM(s), Elementary flux modes(s); MCS(s), Minimal cut set(s)

* Corresponding author.
E-mail address: klamt@mpi-magdeburg.mpg.de (S. Klamt).

http://dx.doi.org/10.1016/j.jbiotec.2017.05.001
Received 26 January 2017; Received in revised form 28 April 2017; Accepted 3 May 2017
Available online 10 May 2017

regulatory networks can be explored by qualitative and semi-quantitative modeling approaches. The development of the software started more than 15 years ago and since then its scope and functionality has grown steadily. *CellNetAnalyzer* can be used via command-line based operations or via a graphical user interface with embedded network visualizations. Herein we will present key functionalities of *CellNetAnalyzer* for applications in biotechnology and metabolic engineering and thereby briefly review constraint-based modeling techniques.

2. Fundamentals of stoichiometric and constraint-based modeling

We start with a short introduction to the mathematical foundations of stoichiometric and constraint-based modeling (SCBM); detailed descriptions can be found elsewhere (Maarleveld et al., 2013; O'Brien et al., 2015; Klamt et al., 2014). SCBM methods require as (minimal) input the $m \times q$ stoichiometric matrix $\mathbf{N}$ capturing the structure of the metabolic network (columns: q reactions; rows: m metabolites with their reaction stoichiometries). Central to all SCBM methods is the assumption of steady state (concentrations of intracellular metabolites do not change) which implies the metabolite balancing equation

$$\mathbf{Nr} = \mathbf{0} \tag{1}$$

where $\mathbf{r}$ is the vector of net reaction rates (also called flux vector or flux distribution). Usually, several biochemical reactions are known to be irreversible which are collected in the index set *Irr*. These reactions can only proceed in forward direction posing sign restrictions on their rates:

$$r_i \geq 0 \; \forall \, i \in Irr \; . \tag{2}$$

Mathematically, the set of flux vectors $\mathbf{r}$ satisfying (1) and (2) form a convex polyhedral cone ("flux cone"). This cone is often analyzed by means of elementary flux modes (Section 4.5). For some reactions we might additionally know minimum/maximum flux capacities (e.g., maximal substrate uptake rates)

$$\alpha_i \leq r_i \leq \beta_i \tag{3}$$

and for some fluxes we might even have measurements

$$r_k = m_k. \tag{4}$$

Combining constraints (1) and (2) with (3) and/or (4) changes the solution space from a cone to a bounded or unbounded (flux) polyhedron. SCBM is based on Eqs. (1)–(4) and employs techniques from linear algebra, linear programming, and computational geometry to analyze the flux space and properties of feasible flux vectors (see also Section 4).

3. Overview of *CellNetAnalyzer*

CellNetAnalyzer (CNA) is a MATLAB toolbox for analyzing biological networks on the basis of topological, stoichiometric, qualitative (logical) and semi-quantitative modeling approaches requiring no or only few (kinetic) parameters. In particular, CNA includes various methods that facilitate an in-depth analysis of metabolic networks based on techniques of SCBM as detailed in Section 4. Functions to study signaling and regulatory networks via interaction (influence) graphs, logical (Boolean) networks, or logic-based ODEs are also included, however, this type of analysis will not be described herein (we refer the reader to Klamt et al., 2007).

The internal architecture of CNA is depicted in Fig. 1. User-created network project(s) are the central objects in CNA. A network project can be of type "mass-flow" (metabolic) or "signal–flow" (signaling/regulatory). Every network project consists of a formal network representation (model) and, optionally, of a graphical user interface (GUI) with one or several *interactive network maps* visualizing the network and allowing interactive input and output of calculated results (Fig. 2). The user can endow a network project with a GUI by providing suitable graphics (bitmap images) of the network, e.g. by using appropriate drawing programs (such as OMIX (Droste et al., 2011, 2013); see also Supplementary Info) or by using maps from other sources such as KEGG. In this way, CNA is very flexible regarding the visual representation of the network. The connection between model and network graphics − resulting in the interactive network maps − is then established by placing input fields (small text boxes) on the network graphics (Fig. 2). Each input field is associated with one network element, for instance, a reaction. The position of the input fields in the network map can be intuitively defined by the user by clicking on the respective position in the network map once a new reaction is defined. The abstract model of a metabolic network is constructed by declaring metabolites and reactions and their respective properties (names, ID, external/internal metabolites, reaction equation, minimal/maximal reaction rates; coefficients in linear objective function, notes, etc.). CNA supports the convenient definition of biomass constituents (proteins, RNA, DNA, etc.). Prior to computations, the biomass composition can be specified by the percentages of the biomass constituents which enables quick adaptation of the stoichiometry of the growth reaction in the metabolic network model. Within the GUI, models can be constructed and edited via a network composer. A new project can also be instantiated by providing the network's stoichiometric matrix, indices of irreversible reactions (Eq. (2)), and flux constraints (Eq. (3)). Furthermore, stoichiometric and constraint-based models can be imported and exported in SBML format (Hucka et al., 2003; including the recently established flux balance constraint package (Olivier and Bergmann, 2015)) or be converted from or to COBRA (Schellenberger et al., 2011) and Metatool (Pfeiffer et al., 1999) models.

Created network projects can be analyzed by the comprehensive toolbox provided with *CellNetAnalyzer* (major functions are described in Section 4). In the GUI, the user may enter, for example, known reaction rates into the respective input fields and then start the calculation by choosing a function from the CNA's menu bar installed in the interactive network maps (Fig. 2). In return, results of calculations are displayed in the network maps. CNA also provides an Application Programming Interface (API) which supports model/GUI access, command-line mode, and batch calculations (Klamt and von Kamp, 2011). In particular, this allows model analysis without the necessity to have a GUI. Most functions provided in the GUI are also supported in command-line mode via the API. In fact, some functions, where a GUI-based workflow is not practical, are only accessible via API. Furthermore, the API allows access and modifications of the model and if the project is endowed with a GUI it can also be used to read/write values from/to the GUI. In this way, the user may program own calculation routines that make use of the abstract model and then display results of these calculations within the network maps. Generally, for some of its calculations, CNA utilizes external packages including linear programming solvers (CPLEX, glpk) and elementary modes calculation routines (efmtool (Terzer and Stelling, 2008) and Metatool (von Kamp and Schuster, 2006)), to which it interfaces via Java and MEX code.

4. Metabolic network analysis with *CellNetAnalyzer*

4.1. Basic network properties

CNA calculates a number of basic network properties, which is especially useful when a new metabolic network model has been created. This includes *conservation relations* and *coupled or blocked reactions*. Conservation relations (CRs) are weighted sums of metabolite concentrations that remain constant in a metabolic reaction network, irrespective of the chosen reaction kinetics. A typical example for a conservation relation in certain metabolic network models is [NADH] + [NAD^+] = CONST. CRs correspond to linearly dependent rows in the stoichiometric matrix $\mathbf{N}$ and CRs can be represented by vectors of

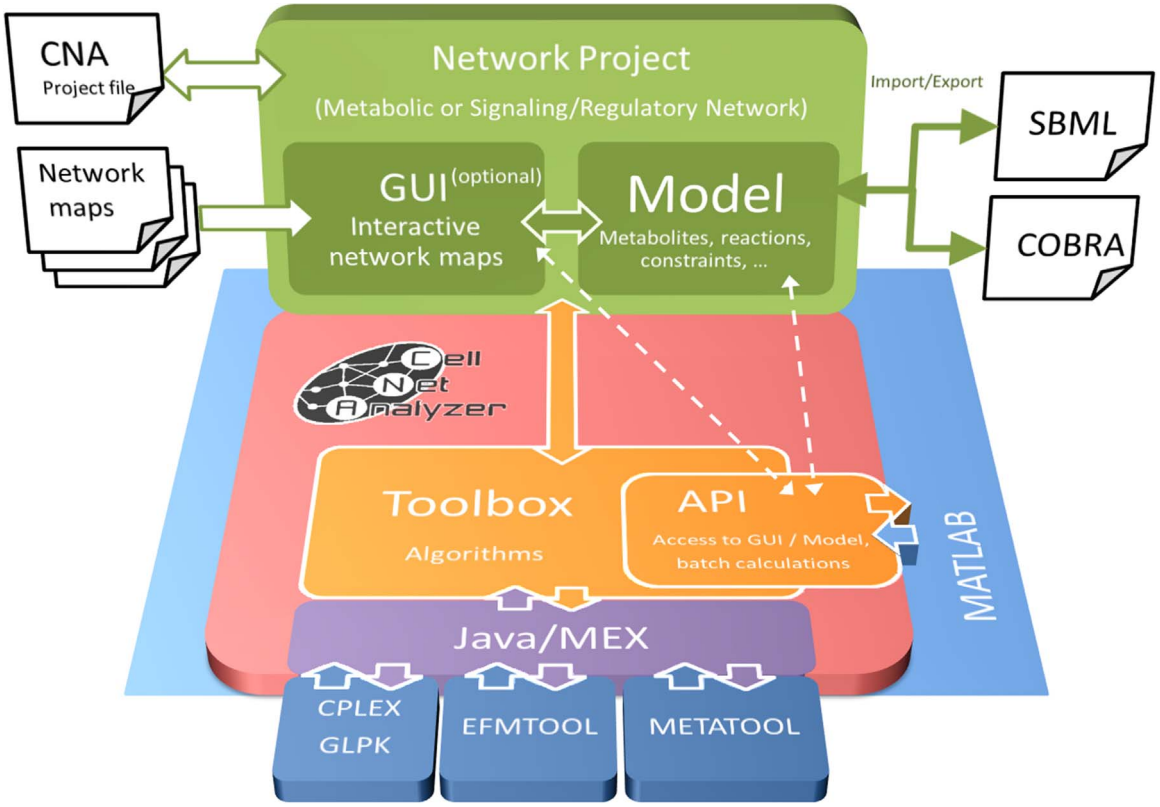

Fig. 1. Architecture of *CellNetAnalyzer*.

the left nullspace of **N** (Klamt et al., 2014). The identification of CRs is important for several reasons. Non-negative CRs (where all coefficients in the sum are non-negative) indicate so-called *conserved moieties*. The case of NADH and NAD^+ is such an example where the NAD^+ molecule is the conserved moiety. CNA provides a procedure to calculate elementary (non-decomposable) CRs which cannot be further reduced. The user can compute either all elementary CRs or only the non-negative elementary CRs (representing the conserved moieties). For stoichiometric (and also kinetic) models, CRs represent some kind of redundancy which can be eliminated − without losing any relevant information − by removing certain metabolites from the stoichiometric matrix. Removing CRs from a reaction network is not only useful to reduce the dimension of the system but also to avoid numerical issues that may arise in certain calculations. CNA provides therefore a function which successively deletes certain metabolites (or marks them as external species) until no further CR remains in the system.

Blocked (strictly detailed balanced) reactions are reactions whose rates are always zero in steady state and are thus of no relevance for stationary flux distributions. A simple example is a reaction that produces a “dead-end” metabolite or lies on a linear pathway producing a “dead-end” metabolite. But there can be also less intuitive cases. Blocked reactions often arise from modeling errors or gaps in the metabolic network. One can search for appropriate corrections or, alternatively, remove blocked reactions when analyzing the network with SCBM methods. Another important network property are coupled reactions (also called enzyme subsets or correlated reaction sets). For any steady-state flux vector, coupled reactions operate with a fixed ratio in their rates (Pfeiffer et al., 1999; Burgard et al., 2004), that is, there is a strong dependency between the fluxes. Typical examples are reactions in a linear pathway but more complicated reaction couplings may also exist.

CNA provides methods that use nullspace and flux variability analysis (see below) to detect blocked and coupled reactions. Alternatively, elementary flux mode analysis can be used for this task which also allows the identification of more complicated hierarchical couplings (one reaction requires another but not the other way around (Burgard et al., 2004)).

4.2. Metabolic flux analysis

The purpose of metabolic flux analysis (MFA) is to determine a steady-state flux distribution in a metabolic network when some fluxes are already known. Typically, MFA is used to calculate the unknown internal fluxes of a cell when measurements of the more easily accessible exchange fluxes between the cell and its environment are available. In this way insight into the physiology of the cell and the operation of the metabolism can be gained which is useful to derive metabolic engineering strategies (see e.g. Schwender 2008; Amaral et al., 2010; Quirós et al., 2013; Lohr et al., 2014). We focus here on standard MFA as supported by CNA where measurements of external fluxes are used as inputs; the much more complicated flux analysis based on isotopic tracer experiments requires a different framework as introduced elsewhere (Weitzel et al., 2013).

For standard MFA, methods from linear algebra are used to calculate unknown fluxes from the values of the known fluxes. After appropriate reordering of the reactions, the steady-state Eq. (1) can always be divided into a measured/known (index k) and unknown (index u) part:

$$\mathbf{Nr} = \mathbf{N}_u\mathbf{r}_u + \mathbf{N}_k\mathbf{r}_k = \mathbf{0} \ .$$

This can be rewritten as

$$\mathrm{N}_u\mathrm{r}_u = -\mathrm{N}_k\mathrm{r}_k \ ,$$

which is the central equation of MFA. Since $\mathbf{N}_k$ and $\mathbf{r}_k$ are both known their product is a vector and the equation above forms an inhomogeneous system of linear equations which has a general least-squares solution (see Klamt et al. (2002)). Importantly, depending on the rank of $\mathbf{N}_u$, a MFA scenario is either determined (all unknown rates are uniquely calculable) or underdetermined (not all rates are deter-

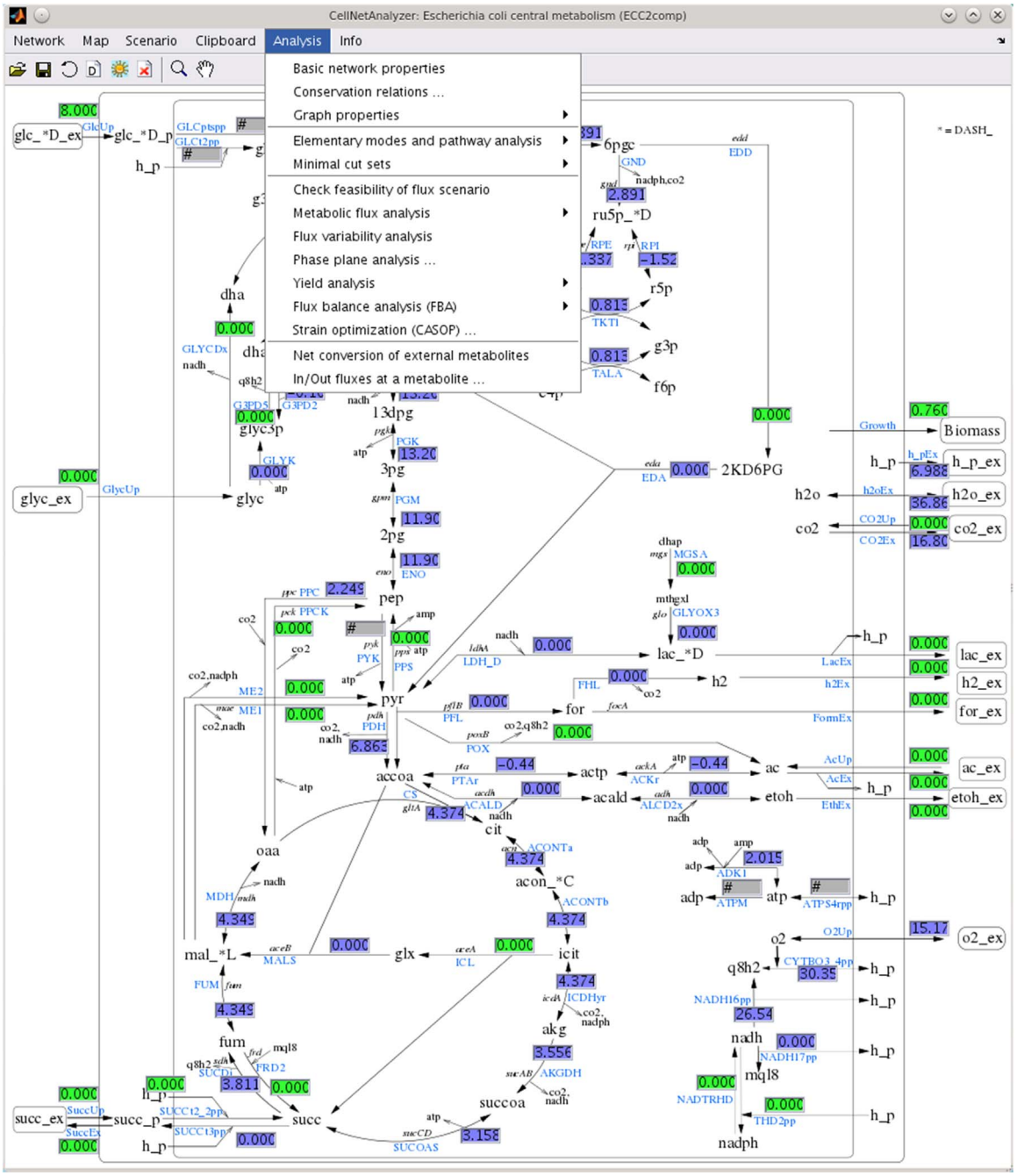

Fig. 2. Screenshot of an interactive network map in *CellNetAnalyzer* showing a calculated flux distribution (green boxes: given values; blue boxes: calculated rates; grey boxes: non-calculable rates). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mined, but some might be calculable). Furthermore, a MFA scenario is either redundant (some of the known rates are not independent) or non-redundant. If the scenario is redundant, the system is either consistent or inconsistent (some rates contradict each other, e.g., due to measurement errors). In the latter case, the known rates can be balanced by certain approaches before computing the uniquely calculable rates (Stephanopoulos et al., 1998; Van der Heijden et al., 1994).

In CNA, after specifying the known rates in the text boxes and starting the MFA calculation, the determinacy and redundancy of the resulting scenario are determined and, if the system is redundant, methods for balancing inconsistent rates are provided. As results, the values of the calculable rates are displayed in the interactive network maps. In standard MFA, flux bounds and reaction directionalities are not explicitly taken into account, however, warnings will be given by CNA if certain constraints are violated. CNA also provides a function to test, as a preliminary step before doing the MFA calculation, the general feasibility of a given MFA scenario thereby detecting different types of infeasibility (inconsistent reaction rates due to redundancies; violation of any irreversibility or capacity constraints).

A concrete example of a MFA scenario in CNA is displayed in Fig. 2. The map shows a network model of the central metabolism of *Escherichia coli* which was derived by reducing and compressing a genome-scale model of *E. coli* (Hädicke and Klamt, 2017). The known values in this MFA scenario are the growth rate and the substrate (glucose) uptake rate. In addition, several internal fluxes are assumed to be inactive (all predefined fluxes are indicated by green boxes). The blue boxes show initially unknown flux values that could be calculated by MFA, and the grey boxes indicate reactions whose fluxes could not

be determined uniquely and remain thus unknown.

4.3. Flux balance analysis and yield space analysis

The main purpose of flux balance analysis (FBA) is the optimization (maximization or minimization) of a linear objective function over the reaction rates:

$$\underset{\mathrm{r}}{maximize}\ z = \mathrm{c}^T\mathrm{r}. \tag{5}$$

The objective function is thus defined by a vector **c** containing a weight (coefficient) for each reaction (in most cases, only one or few coefficients are different from zero). Typical objective functions in FBA applications maximize, for example, the growth rate or the formation rate of a certain product. Importantly, the optimization is subject to the general constraints (1)–(4), that is, optimal flux vectors must fulfill the steady state condition (Eq. (1)) and comply with reversibility (Eq. (2)), capacity (Eq. (3)) and measurement constraints (Eq. (4)) when such are given. FBA problems constitute linear programs (LPs) for which dedicated solvers exist.

A typical FBA task of biotechnological relevance is the maximization of the growth rate or the production of a certain compound under a given substrate uptake limit. If the substrate uptake is set to a fixed value then effectively the biomass or product yield under the specified uptake rate is maximized (but this does not hold true for the general case where the substrate uptake rate is not fixed; see below). Another use of FBA is to check whether a reaction is essential for some specified behavior. For instance, in order to determine if a reaction is essential for growth, the reaction rate is fixed at zero and the growth rate is maximized. If the maximal growth rate is zero as well it follows that the reaction is essential for growth, otherwise it is not. Note that although the optimal value of the objective function is unique, the associated fluxes that are calculated as the optimal FBA solution are, in general, not unique (cf. flux variability analysis below). Further applications of FBA together with extensions of this method have been reviewed by Gianchandani et al. (2010).

In CNA, as for MFA, the user may fix some flux values in the interactive flux maps and then start the optimization (in CNA the objective function is defined by specifying the respective coefficients c_i for each reaction i in the reaction properties). For the actual computation, the user can choose between three supported solvers (Fig. 1: MATLAB's linprog, GLPK, CPLEX). The computed optimal flux distribution is displayed on the interactive network maps and a potential infeasibility would be reported. Such infeasibility occurs when the linear program contains contradictory constraints. In certain situations, the FBA optimum may be unbounded, for example, when the growth rate is maximized without a substrate uptake limit and without capacity constraints on the other reactions. Therefore, care should be taken to use meaningful capacity constraints for FBA.

Related to FBA is the new functionality of CNA to maximize a user-defined *yield function* (or a certain ratio of reaction rates), for example, the yield of a relevant product with respect to a given substrate. As has been pointed out, the maximization of the rate-based linear objective function (5) can, in the general case, not be used to find the flux vector (s) with an optimal product yield within a given flux space and linear-fractional programming must be applied instead (Burgard et al., 2004; Klamt et al., 2017). Furthermore, CNA can now also be used to map the flux space on a two-dimensional *yield space* where two (user-defined) yields are plotted against each other. Yield spaces are of particular relevance for metabolic engineering to analyze the trade-offs between biomass yield and certain product yields (Klamt and Mahadevan, 2015) and they can be calculated also in large (genome-scale) networks.

4.4. Flux variability analysis and phase planes

As mentioned above, an MFA scenario and even optimal flux distributions calculated by FBA can be non-unique. To get a better overview of feasible flux values, flux variability analysis (FVA) can be used. FVA takes the constraints from Eqs. (1)–(4) and iteratively minimizes and maximizes every reaction rate (hence, it also requires LP solvers). In this way the feasible ranges of all reaction rates are determined which provides insight into the variability of the fluxes in the network for a given scenario (see e.g. Hädicke et al., 2011; Flahaut et al., 2013; Hay and Schwender, 2014; Lohr et al., 2014). As an example for a biotechnological application, one can calculate an optimal production rate by FBA and then run FVA with the fixed optimal rate to determine the ranges for the optimal flux distributions that support optimal production. However, FVA does not describe the solution space in as much detail as elementary flux mode analysis (see below). With FVA it is also possible to identify blocked reactions for a given flux scenario as these have a minimal and maximal rate of zero.

In CNA, FVA can be performed in the same way as FBA and some fluxes may be set to a fixed value in the network maps. Again, it is possible to choose one of the three supported LP solvers. After calculation, CNA displays the flux ranges of all reactions and highlights reactions in the network map for which only a single (unique) flux value is possible in the defined scenario.

As a related function to FVA, CNA also allows the analysis of *phase planes* where the flux solution space is projected on two selected reaction rates to investigate their mutual dependencies. In the context of biotechnological applications, this is particularly useful for the analysis of production envelopes where the growth rate is plotted against the production rate of a compound of interest (Machado and Herrgard, 2015).

4.5. Elementary flux modes

An elementary flux mode (EFM) is defined as feasible steady-state flux vector fulfilling the Eqs. (1) and (2) and the further property that it uses a minimal (or irreducible) set of reactions (Schuster and Hilgetag, 1994; Schuster et al., 2000; Trinh et al., 2009). Irreducible means that the steady-state condition can no longer be fulfilled if any reaction that is used by the EFM (with a rate unequal to zero) is removed from the EFM. The concept of an EFM formalizes the notion of a metabolic pathway in that the EFM can be seen as a minimal connected subnetwork which needs to operate in order to keep the system in steady state. In fact, an EFM is uniquely defined by its set of reactions, while its flux distribution may be arbitrarily scaled. The set of all EFMs describes the solution space of a system given by Eqs. (1) and (2) in detail because any feasible flux distribution obeying these constraints can be decomposed into a non-negative combination of EFMs. A further extremely useful property is that, if a reaction is removed from the system, the set of all EFMs of the smaller system can be derived by removing all EFMs in which the deleted reaction participates. This property can also be exploited for the calculation of minimal cut sets (see below).

EFMs can be used to analyze various properties of a given metabolic network and they have also become a standard tool for metabolic modeling in biotechnology and metabolic engineering (Trinh et al., 2009; Horvat et al., 2015; Zanghellini et al., 2013; Klamt et al., 2014). First of all, as already described, EFMs can be used to identify minimal pathways (or subnetworks) leading from a given substrate to a (biotechnologically relevant) product or to biomass. In particular, the set of EFMs allows one to determine the maximal yield of any product/substrate pair because the pathway with maximal yield is always an EFM. In addition, by calculating the yields of all modes one can see how many (and which) optimal pathways exist and how many further solutions can be opened up by allowing a suboptimal yield. EFMs can be used to evaluate the importance of a reaction for a (desired or undesired) phenotype and blocked, essential, or coupled reactions can be immediately identified by the network's EFMs. EFMs may also indicate internal cycles operating without uptake of substrate. These cycles are thermodynamically infeasible and may point to potential

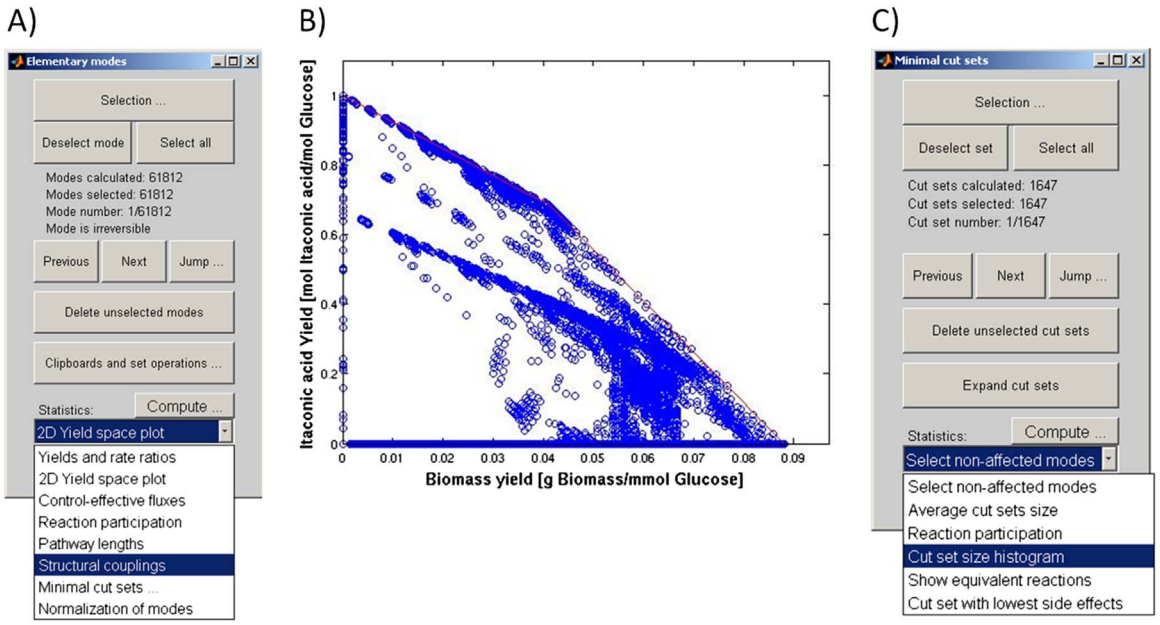

Fig. 3. A) EFM panel, B) example of a two-dimensional yield space plot of EFMs, and C) MCS panel in *CellNetAnalyzer*.

errors in the network. Last but not least, EFMs can be used to calculate metabolic engineering strategies, for example, based on correlation of pathway fluxes to desired properties (Melzer et al., 2009; Neuner and Heinzle, 2011; Poblete-Castro et al., 2013) or via minimal cut sets or the CASOP method (see below).

CNA not only supports EFM computation but also provides methods for a detailed analysis of EFMs. The calculation of EFMs is not trivial in larger networks (and mostly infeasible in genome-scale models), because the number of modes can grow exponentially with the size of the network. Therefore, two dedicated tools for the computation of EFM can be used by CNA, namely Metatool (von Kamp and Schuster, 2006) and efmtool (Terzer and Stelling, 2008). Prior to the computation, the user may specify reactions which are either inactive (zero flux) or must be active in the EFMs to be calculated. When the computation is finished, an "EFM panel" pops up which can be used to analyze and navigate through the EFMs (Fig. 3A). On the network maps the currently selected EFM is displayed with its flux values in the reaction text boxes. Using the EFM panel, subsets of the EFMs can be selected according to various criteria (involved reactions/metabolites, min/max yields, min/max pathway lengths). EFM selections can be stored in clipboards and set operations be performed on different selections. These selections allow the user to identify pathways having certain properties or/and being relevant for a particular condition. The EFM panel also includes a menu from which various methods for the analysis of a selection of EFMs can be launched (Fig. 3A). For example, relative reaction participations, pathway lengths and yields as well as yield space plots can be calculated. In the latter, each EFM is plotted as a point on a plane with its coordinates determined by its yield for two products (typically, biomass and a chemical of interest). Together with the convex hull around the points this graphical representation (Fig. 3B) gives a quick overview over the distributions and possible combinations of the two product yields. As it was already mentioned above, yield space plots are very useful for metabolic engineering applications. Finally, the calculation of minimal cut sets (see below) can also be started from the selected EFMs.

Lastly, the computation of *elementary flux vectors*, a generalization of EFMs where inhomogeneous constraints, e.g. flux bounds, can be taken into account (Urbanczik, 2007; Klamt et al., 2017), is also supported. Calculated elementary flux vectors can be processed and analyzed in the same way as EFMs.

4.6. Minimal cut sets and computational strain design

Minimal cut sets (MCSs) have been introduced as a formal concept to calculate metabolic engineering strategies in constraint-based metabolic models. The calculation of MCSs requires as an initial step the specification of undesired and desired flux vectors (phenotypes or functions) (Klamt and Gilles, 2004; Hädicke and Klamt, 2011). In a biotechnological setting, undesired phenotypes could be flux vectors with low product yield while the desired phenotype could be high product yield with some minimal growth rate. The specification of desired and undesired phenotypes can be done either via the EFMs (resulting in two sets of desired and undesired (target) EFMs, Hädicke and Klamt, 2011) or via appropriate linear inequalities (von Kamp and Klamt, 2014). CNA supports both methods (see below). Given these specifications, all minimal (irreducible) reaction knockout strategies (the MCSs) can be calculated by which all undesired flux vectors will be deleted while keeping at least some desired flux vectors feasible.

The main application of MCSs lies in rational strain design. In particular, MCSs can be used to identify growth-coupled strain designs where growth becomes obligatorily coupled to product synthesis. As a recent real-world application, MCSs have been used to successfully design a high-yield itaconic acid producer strain of *E. coli* (Harder et al., 2016). MCSs have also been used to study the phenotypic roles of genes in the anthocyan production pathways of plants (Clark and Verwoerd, 2011) or to identify knockout strategies to support heterologous terpenoid synthesis in yeast (Gruchattka et al., 2013) as well as to enhance ethanol production in cyanobacteria (Erdrich et al., 2014). As another application, MCSs can be used to evaluate the robustness of a metabolic network with respect to a given functionality (Klamt and Gilles, 2004; Klamt, 2006; Behre et al., 2008; Gerstl et al., 2016). Generally, a large number of small MCSs indicates fragility.

In CNA, the calculation of MCSs via EFMs can be directly started within the EFM panel (Fig. 3A). The specification of desired and undesired EFMs is facilitated by the described EFM selection tool. Then, CNA uses the Berge algorithm (Berge, 1989) to enumerate the corresponding MCSs for the given intervention problem. A large number of distinct MCSs may exist for a given problem. Usually one is interested in the smallest MCSs because they require the least number of interventions (reaction knockouts). Therefore, it is possible to limit the size of MCSs to be calculated. Once calculated, a new MCS panel comes up (similar to the EFM panel; see Fig. 3C) which allows (i) the

display of each MCS within the network maps, (ii) the selection of certain subsets of MCSs (e.g., involving certain reaction knockouts), and (iii) statistical calculations (e.g. cut set size histogram) of those selections.

Using the API of CNA, a direct computation of MCSs is possible without having to calculate the EFMs first so that MCSs even in genome-scale networks can be determined (Klamt and von Kamp, 2014). As mentioned above, in this approach, desired and undesired behaviors are defined by linear inequalities which make it also possible to integrate inhomogeneous constraints like Eq. (3) and to allow also up and down regulation of fluxes as possible interventions (Mahadevan et al., 2015). With this methodology, the smallest MCSs can be enumerated also in large-scale networks, for example, more than 8000 MCSs up to size 7 could be calculated for growth-coupled ethanol production in a genome-scale model of *E. coli* (von Kamp and Klamt, 2014).

Beside minimal cut sets, another method for computational strain design supported by CNA is CASOP (Computational Approach for Strain Optimization aiming at high Productivity; Hädicke and Klamt, 2010). This heuristic approach determines knockout and overexpression candidates for increasing the productivity of a strain. CASOP analyzes EFMs to identify potential targets to increase the flux towards the product (not necessarily the product yield) while keeping a lowered growth rate feasible. The result is a ranked list of knockout and overexpression candidates. For application examples of CASOP see Pande et al. (2014) and Erdrich et al. (2014).

5. Discussion and conclusions

Metabolic modeling based on stoichiometric and constraint-based methods has become a routinely used tool for bioprocess and metabolic engineering. A number of successful studies demonstrated the value of model-based approaches for analyzing and optimizing the metabolism of microbial production organisms (Jang et al., 2012; Kung et al., 2012; Lee and Kim, 2015; Machado and Herrgard, 2015; Maia et al., 2015). Accordingly, several software suits have been developed to support this process, with *CellNetAnalyzer* being one of the largest packages. Other software with a similar scope include the COBRA and COBRApy toolbox (Schellenberger et al., 2011; Ebrahim et al., 2013), OPTFLUX (Rocha et al., 2010), RAVEN (Agren et al., 2013), OMIX (Droste et al., 2011), or the R package sybil (Gelius-Dietrich et al., 2013). Despite of significant overlaps, each package has its own merits and strengths. For example, among these tools, the MATLAB toolbox RAVEN provides the richest functionality for semi-automated reconstruction and gap filling of genome-scale metabolic models. OMIX has its main focus in providing sophisticated tools for drawing and visualizing metabolic networks and related datasets, although basic analysis tools are also included. The COBRA toolbox for MATLAB (and, similarly, the related COBRApy toolbox for Python) provides a comprehensive set of functions for constraint-based modeling. This includes several methods for computational strain design based on bi-level Mixed Integer Linear Programming problems (bi-level MILPs (Maia et al., 2015; Machado et al., 2015; Zomorrodi et al., 2012)); those methods are not (yet) supported by *CellNetAnalyzer*. On the other hand, all functions have to be started from command line (no GUI or visualization) and methods for metabolic pathway analysis are not available. OPTFLUX provides similar analysis tools as the COBRA toolbox although its focus is more on applications in metabolic engineering. Furthermore, OPTFLUX comes with a GUI allowing a more convenient specification and conduction of simulations. Data input (e.g. fixed fluxes) and display of calculated results are mainly done via tables, although some simulation results can also be visualized if suitable network drawings are provided by the user. Generally, compared to the packages mentioned above, unique features of CNA are, in particular, detailed pathway analysis based on elementary flux modes or elementary flux vectors and enumeration of metabolic engineering strategies based on minimal cut sets. Moreover, with its interactive network maps, CNA is, to the best of our knowledge, the only tool allowing both user input AND display of calculated results directly within a network visualization. This eases and speeds up frequent use cases in metabolic network modeling, e.g., the specification of measured/known fluxes and the subsequent (MFA or FBA based) calculation and display of resulting flux distributions. To our experience, displaying numeric results directly within network maps (possibly in combination with colors to highlight, for example, low or high fluxes) greatly enhances the process of interpretation of data and computed results. So far, CNA has mainly been used for analyzing medium-scale metabolic networks, but there is, in fact, no limitation to use it with genome-scale networks as well, either via the provided API (this MATLAB command line mode is similar to the COBRA toolbox) or, if one or several network maps are available, also via the GUI. As an example, Supplementary Figs. 1 and 2 show screenshots of a CNA interactive network map of a genome-scale model of *Corynebacterium glutamicum*. The mathematical model was originally presented in (Zelle et al., 2015) together with a map of the network created with OMIX (Droste et al., 2011, 2013). We implemented this model in CNA and used the OMIX visualization as background network map in CNA. We also note that the API of CNA provides functions to convert CNA models to COBRA models (and vice versa) directly in MATLAB thus allowing parallel application of methods from both tools for analyzing a CNA or a COBRA model.

The *CellNetAnalyzer* package (including a tutorial and manual) can be downloaded for free for academic use from: https://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html.

Acknowledgements

This work was in parts supported by the German Federal Ministry of Education and Research (de.NBI partner project “ModSim” (FKZ: 031L104B) and Biotechnologie 2020+ project CASCOO (FKZ: 031A180B)) and by the European Research Council (ERC Consolidator Grant 721176). We thank Tim Rose for compiling the CNA project for *Corynebacterium glutamicum* shown in the Supplementary Data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2017.05.001.

References

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J., 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput. Biol. 9 (3) e1002980.

Amaral, A.I., Teixeira, A.P., Martens, S., Bernal, V., Sousa, M.F., Alves, P.M., 2010. Metabolic alterations induced by ischemia in primary cultures of astrocytes: merging 13C NMR spectroscopy and metabolic flux analysis. J. Neurochem. 113 (3), 735–748.

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

Berge, C., 1989. Hypergraphs. Combinatorics of Finite Sets. North-Holland, Amsterdam.

Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 14, 301–312.

Clark, S.T., Verwoerd, W.S., 2011. A systems approach to identifying correlated gene targets for the loss of colour pigmentation in plants. BMC Bioinf. 12, 343.

Droste, P., Miebach, S., Niedenführ, S., Wiechert, W., Nöh, K., 2011. Visualizing multi-omics data in metabolic networks with the software Omix: a case study. Biosystems 105, 154–161.

Droste, P., Nöh, K., Wiechert, W., 2013. Omix − a visualization tool for metabolic networks with highest usability and customizability in focus. Chem. Ing. Tech. 85, 849–862.

Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRApy: constraints-based reconstruction and analysis for python. BMC Syst. Biol. 7, 74.

Erdrich, P., Knoop, H., Steuer, R., Klamt, S., 2014. Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling. Microb. Cell Fact. 13, 128.

Flahaut, N.A., Wiersma, A., van de Bunt, B., Martens, D.E., Schaap, P.J., Sijtsma, L., Dos Santos, V.A., de Vos, W.M., 2013. Genome-scale metabolic model for Lactococcus

lactis MG1363 and its application to the analysis of flavor formation. Appl. Microbiol. Biotechnol. 97 (19), 8729–8739.
Gelius-Dietrich, G., Desouki, A.A., Fritzemeier, C.J., Lercher, M.J., 2013. Sybil − efficient constraint-based modelling in R. BMC Syst. Biol. 7 (125).
Gerstl, M.P., Klamt, S., Jungreuthmayer, C., Zanghellini, J., 2016. Exact quantification of cellular robustness in genome-scale metabolic networks. Bioinformatics 32, 730–737.
Gianchandani, E.P., Chavali, A.K., Papin, J.A., 2010. The application of flux balance analysis in systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 2 (3), 372–382.
Gruchattka, E., Hädicke, O., Klamt, S., Schütz, V., Kayser, O., 2013. In silico profiling of Escherichia coli and Saccharomyces cerevisiae as terpenoid factories. Microb. Cell Fact. 12, 84.
Gutierrez, J.M., Lewis, N.E., 2015. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. 10, 939–949.
Hädicke, O., Klamt, S., 2010. CASOP: a computational approach for strain optimization aiming at high productivity. J. Biotechnol. 147, 88–101.
Hädicke, O., Klamt, S., 2011. Computing complex metabolic intervention strategies using constrained minimal cut sets. Metab. Eng. 13, 204–213.
Hädicke, O., Klamt, S., 2017. EColiCore2: a reference model of the central metabolism of *Escherichia coli* and the relationships to its genome-scale parent model. Sci. Rep. 7, 39647.
Hädicke, O., Grammel, H., Klamt, S., 2011. Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. BMC Syst. Biol. 5, 150.
Harder, B.-J., Bettenbrock, K., Klamt, S., 2016. Model-Based metabolic engineering enables high yield itaconic acid production by Escherichia coli. Metab. Eng. 38, 29–37.
Hay, J.O., Schwender, J., 2014. Flux variability analysis: application to developing oilseed rape embryos using toolboxes for constraint-based modeling. Methods Mol. Biol. 1090, 301–316.
Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Kummer, U. Mendes, 2006. COPASI: a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074.
Horvat, P., Koller, M., Braunegg, G., 2015. Recent advances in elementary flux modes and yield space analysis as useful tools in metabolic network studies. World J. Microbiol. Biotechnol. 31 (9), 1315–1328.
Hucka, M., et al., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531.
Jang, Y.S., Park, J.M., Choi, S., Choi, Y.J., Seung do, Y., Cho, J.H., Lee, S.Y., 2012. Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. Biotechnol. Adv. 30, 989–1000.
Kim, B., Kim, W.J., Kim, D.I., Lee, S.Y., 2015. Applications of genome-scale metabolic network model in metabolic engineering. J. Ind. Microbiol. Biotechnol. 42, 339–348.
King, Z.A., Lloyd, C.J., Feist, A.M., Palsson, B.Ø., 2015. Next-generation genome-scale models for metabolic engineering. Curr. Opin. Biotechnol. 35, 23–29.
Klamt, S., Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20, 226–234.
Klamt, S., Mahadevan, R., 2015. On the feasibility of growth-coupled product synthesis in microbial strains. Metab. Eng. 30, 166–178.
Klamt, S., von Kamp, A., 2011. An application programming interface for CellNetAnalyzer. Biosystems 105, 162–168.
Klamt, S., Schuster, S., Gilles, E.D., 2002. Calculability analysis in underdeterminedmetabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. Biotechnol. Bioeng. 77, 734–751.
Klamt, S., Saez-Rodriguez, J., Gilles, E.D., 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2.
Klamt, S., Hädicke, O., von Kamp, A., 2014. Stoichiometric and constraint-based analysis of biochemical reaction networks. In: Benner, P., Findeisen, R., Flockerzi, D., Reichl, U., Sundmacher, K. (Eds.), Large-Scale Networks in Engineering and Life Sciences. Springer, pp. 263–316.
Klamt, S., Regensburger, G., Gerstl, M.P., Jungreuthmayer, C., Schuster, S., Mahadevan, R., Zanghellini, J., Müller, S., 2017. From elementary flux modes to elementary flux vectors: metabolic network and pathway analysis with arbitrary linear flux constraints. PLoS Comput. Biol. 13 e1005409.
Klamt, S., 2006. Generalized concept of minimal cut sets in biochemical networks. Biosystems 83, 233–247.
Kung, Y., Runguphan, W., Keasling, J.D., 2012. From fields to fuels: recent advances in the microbial production of biofuels. ACS Synth Biol 1, 498–513.
Lee, S.Y., Kim, H.U., 2015. Systems strategies for developing industrial microbial strains. Nat. Biotechnol. 33, 1061–1072.
Lohr, V., Hädicke, O., Genzel, Y., Jordan, I., Büntemeyer, H., Klamt, S., Reichl, U., 2014. The avian cell line AGE1. CR. pIX characterized by metabolic flux analysis. BMC Biotechnol. 14, 72.
Maarleveld, T.R., Khandelwal, R.A., Olivier, B.G., Teusink, B., Bruggeman, F.J., 2013. Basic concepts and principles of stoichiometric modeling of metabolic networks. Biotechnol. J. 8, 997–1008.
Machado, D., Herrgard, M.J., 2015. Co-evolution of strain design methods based on flux balance and elementary mode analysis. Metab. Eng. Commun. 2, 85–92.
Mahadevan, R., von Kamp, A., Klamt, S., 2015. Genome-scale strain designs based on regulatory minimal cut sets. Bioinformatics 31, 2844–2851.
Maia, P., Rocha, M., Rocha, I., 2015. In silico constraint-Based strain optimization methods: the quest for optimal cell factories. Microbiol. Mol. Biol. Rev. 80, 45–67.
Melzer, G., Esfandabadi, M.E., Franco-Lara, E., Wittmann, C., 2009. Flux Design: in silico design of cell factories based on correlation of pathway fluxes to desired properties. BMC Syst. Biol. 3, 120.
Neuner, A., Heinzle, E., 2011. Mixed glucose and lactate uptake by Corynebacterium glutamicum through metabolic engineering. Biotechnol. J. 6, 318–329.
O'Brien, E.J., Monk, J.M., Palsson, B.O., 2015. Using genome-scale models to predict biological capabilities. Cell 161, 971–987.
Olivier BG. and Bergmann FT., (2015) Flux Balance Constraints, Version 2 Release 1. Available from COMBINE < http://identifiers.org/combine.specifications/sbml.level-3. version-1. fbc.version-2. release-1 >
Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L.F., Kaleta, C., Kost, C., 2014. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J. 8, 953–962.
Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15, 251–257.
Poblete-Castro, I., Binger, D., Rodrigues, A., Becker, J., Dos Santos, V. A. Martins, Wittmann, C., 2013. In-silico-driven metabolic engineering of Pseudomonas putida for enhanced production of poly-hydroxyalkanoates. Metab. Eng. 15, 113–123.
Quirós, M., Martínez-Moreno, R., Albiol, J., Morales, P., Vázquez-Lima, F., Barreiro-Vázquez, A., Ferrer, P., Gonzalez, R., 2013. Metabolic flux analysis during the exponential growth phase of Saccharomyces cerevisiae in wine fermentations. PLoS One 8 e71909.
Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C., Rocha, M., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst. Biol. 4, 45.
Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., BØ, Palsson, 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–1307.
Schuster, S., Hilgetag, C., 1994. On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2, 165–182.
Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18, 326–332.
Schwender, J., 2008. Metabolic flux analysis as a tool in metabolic engineering of plants. Curr. Opin. Biotechnol. 2, 131–137.
Simeonidis, E., Price, N.D., 2015. Genome-scale modeling for metabolic engineering. J. Ind. Microbiol. Biotechnol. 42, 327–338.
Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering. Academic Press, San Diego.
Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24, 2229–2235.
Trinh, C.T., Wlaschin, A., Srienc, F., 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl. Microbiol. Biotechnol. 81, 813–826.
Urbanczik, R., 2007. Enumerating constrained elementary flux vectors of metabolic networks. IET Syst. Biol. 1, 274–279.
Van der Heijden, R.T., Heijnen, J.J., Hellinga, C., Romein, B., Luyben, K.C., 1994. Linear constraint relations in biochemical reaction systems: i: classification of the calculability and the balanceability of conversion rates. Biotechnol. Bioeng. 43, 3–10.
von Kamp, A., Klamt, S., 2014. Enumeration of smallest intervention strategies in genome scale metabolic networks. PLoS Comput. Biol. 10, e1003378.
von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931.
Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B., Wiechert, W., 2012. 13CFLUX2 − high-performance software suite for 13C-metabolic flux analysis. Bioinformatics 29 (1), 143–145.
Wu, H., et al., 2016. MUFINS: multi-formalism interaction network simulator. npj Syst. Biol. Appl. 2 16032.
Zanghellini, J., Ruckerbauer, D.E., Hanscho, M., Jungreuthmayer, C., 2013. Elementary flux modes in a nutshell: properties, calculation and applications. Biotechnol. J. 8 (9), 1009–1016.
Zelle, E., Nöh, K., Wiechert, W., 2015. Growth and production capabilities of corynebacterium glutamicum: interrogating a genome-scale metabolic network model. In: Burkovski, A. (Ed.), Corynebacterium Glutamicum: From Systems Biology to Biotechnological Applications. Caister Academic Press, pp. 39–56.
Zomorrodi, A.R., Suthers, P.F., Ranganathan, S., Maranas, C.D., 2012. Mathematical optimization applications in metabolic networks. Metab. Eng. 14, 672–686.