Gene 521 (2013) 1–14

Contents lists available at SciVerse ScienceDirect

Gene

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

Review

Comparing methods for metabolic network analysis and an application to metabolic engineering

Namrata Tomar, Rajat K. De *

Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Accepted 7 March 2013
Available online 26 March 2013

Keywords:
Systems Biology
Flux balance analysis
Elementary flux mode analysis
Metabolic pathway modeling
Strain optimization

ABSTRACT

Bioinformatics tools have facilitated the reconstruction and analysis of cellular metabolism of various organisms based on information encoded in their genomes. Characterization of cellular metabolism is useful to understand the phenotypic capabilities of these organisms. It has been done quantitatively through the analysis of pathway operations. There are several in silico approaches for analyzing metabolic networks, including structural and stoichiometric analysis, metabolic flux analysis, metabolic control analysis, and several kinetic modeling based analyses. They can serve as a virtual laboratory to give insights into basic principles of cellular functions. This article summarizes the progress and advances in software and algorithm development for metabolic network analysis, along with their applications relevant to cellular physiology, and metabolic engineering with an emphasis on microbial strain optimization. Moreover, it provides a detailed comparative analysis of existing approaches under different categories.

Contents

1. Introduction . . . 2
2. Sources of data on metabolic pathways . . . 3
   2.1. Kyoto Encyclopedia of Genes and Genomes (KEGG) . . . 3
   2.2. BioCyc . . . 3
   2.3. *ERGO™* . . . 3
   2.4. metaTIGER . . . 3
   2.5. ENZYME . . . 3
   2.6. BRaunschweig ENzyme Database (BRENDA) . . . 3
   2.7. BioCarta . . . 3
   2.8. PUBCHEM . . . 4
   2.9. Universal Protein Resource (UNIPROT) . . . 4
   2.10. REACTOME . . . 4
   2.11. Chemical Entities of Biological Interest (CHEBI) . . . 4
   2.12. ExplorENZ . . . 4
   2.13. Integrated relational Enzyme database (IntEnz) . . . 4
   2.14. Protein ANalysis THrough Evolutionary Relationships (PANTHER) . . . 4
3. Existing metabolic network analysis methods: A Categorization . . . 4

Abbreviations: GRN, gene regulatory network; TRN, Transcriptional Regulatory network; TF, transcriptional factor; GPR, gene–protein–reaction; GSMM, genome scale metabolic model; KEGG, Kyoto Encyclopedia of Genes and Genomes; KGML, KEGG Markup Language; EMP, Enzymes and Metabolic Pathways; IUBMB, International Union of Biochemistry and Molecular Biology; BRENDA, BRaunschweig ENzyme Database; EC number, Enzyme Commission (EC) number; UNIPROT, Universal Protein Resource; CHEBI, Chemical Entities of Biological Interest; IUPAC, International Union of Pure and Applied Chemistry; IntEnz, integrated relational Enzyme database; SIB, Swiss Institute of Bioinformatics; PANTHER, Protein ANalysis THrough Evolutionary Relationships; ODE, ordinary differential equations; PDE, partial differential equations; MILP, Mixed Integer Linear Programming; PN, Petri Net; HFPN, Hybrid Functional Petri Net; CPN, Colored Petri Net; SNA, stoichiometric network analysis; EMA, elementary flux mode analysis; EPA, extreme pathway analysis; COBRA, constraint-based reconstruction and analysis; FASIMU, Flux-balance Analysis based SIMUlations; FBA, Flux Balance Analysis; LP, linear programming; MOMA, minimization of metabolic adjustment; QP, quadratic programming; SLP, successive linear programming; ROOM, regulatory on–off minimization; FVA, Flux Variability Analysis; FCA, Flux Coupling Analysis; MMB-FCA, Minimal Metabolic Behaviors Flux Coupling Analysis; FFCA, Feasibility-based Flux Coupling Analysis; PLS, projection to latent structures; SKM, structural kinetic modeling; OMNI method, optimal metabolic network identification method; MEP method, Metabolic expression placement method; PDO, 1,3-propanediol; GDLS, Genetic Design through Local Search; CASOP, Computational Approach for Strain Optimization aiming at high Productivity; EA, evolutionary algorithm; SA, simulated annealing; EMILiO, Enhancing Metabolism with Iterative Linear Optimization.

* Corresponding author. Tel.: +91 33 25753105; fax: +91 33 25783357.
E-mail addresses: namrata_t@isical.ac.in (N. Tomar), rajat@isical.ac.in (R.K. De).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.03.017

3.1. Kinetic modeling . . . 4
3.1.1. Discrete approach . . . 4
3.1.2. Continuous approach . . . 5
3.2. Structural and stoichiometric modeling . . . 5
3.2.1. Network based approaches . . . 5
3.2.2. Constraints based approaches . . . 5
3.3. Hybrid modeling approaches . . . 6
3.3.1. Limitations of GSMM . . . 7
3.3.2. Automated metabolic network model reconstruction . . . 7
4. Comparison of metabolic network analysis approaches under different categories . . . 7
4.1. Comparison among paradigm . . . 8
4.2. Comparison among approaches . . . 8
4.3. Comparison among methods . . . 8
4.3.1. Comparison of EPA and EMA . . . 8
5. Metabolic engineering . . . 9
5.1. Rational design of efficient microbial host strains: An application to metabolic engineering . . . 10
5.1.1. Pathway based design . . . 10
5.1.2. Optimization based design . . . 11
5.2. Comparison among different methods for metabolic engineering . . . 12
6. Conclusions . . . 12
Acknowledgments . . . 13
References . . . 13

1. Introduction

Biological systems differ from each other because of differences in the architecture and functions of the components. A living system cannot be well understood by analyzing individual components only, as the interactions among these components play an important role behind biological processes. Thus, the branch, called Systems Biology, has evolved. The objective of Systems Biology is to study the complex biological processes as integrated systems of many interacting components. The structure and dynamics of cellular and organismal function is examined to understand molecular interactions at the system level, rather than the characteristics of isolated parts of a cell or an organism. These interactions involve a series of biochemical conversions and form different biochemical pathways.

The major kinds of biochemical pathways/networks are — signal transduction, gene regulatory and metabolic networks. Signal transduction networks are the pathways of molecular interactions that mediate sensing and stimuli by detecting, amplifying, and integrating diverse external signals to communicate between the cell membrane and intracellular end-points. A gene regulatory network (GRN) consists of a set of genes, proteins and the regulatory interactions present between them. GRN determines the expression of a particular gene at a time. Metabolic pathways are the series of biochemical reactions, mostly catalyzed by enzymes. It involves the step-by-step modification of an initial molecule (substrate) to form a product. In the present article, we are mainly concerned with the analysis methodologies for metabolic networks.

Through the use of high-throughput technologies and novel analytical tools, biology has shifted from a descriptive science to a predictive one. Biological pathway analysis has become an invaluable aid to understand the data generated from various 'omics' technologies. Information contained in primary databases and in the literature is so extensive and rapidly growing that it has become difficult to integrate and infer from them. Thus, it is convenient to infer from these vast repositories into knowledge bases that consists of annotated representations of biological pathways. Better understanding of these pathways helps biochemists and biotechnologists to draw an entire metabolic map of a cell and reconstruct it by rational metabolic engineering. A general introductory figure has been added as Fig. 1 to give an idea over metabolic pathway modeling and analysis. A number of specific databases and computational approaches have been developed, which makes it possible to perform these tasks. Using experimental data from laboratory, one may wish to systematically conduct his/her analyses with computational system. One major task is to implement molecular information systems that will allow to integrate different molecular database systems, and to design analysis tools. The task involves three key problems, like, 1) reconstruction of metabolic pathways leading to

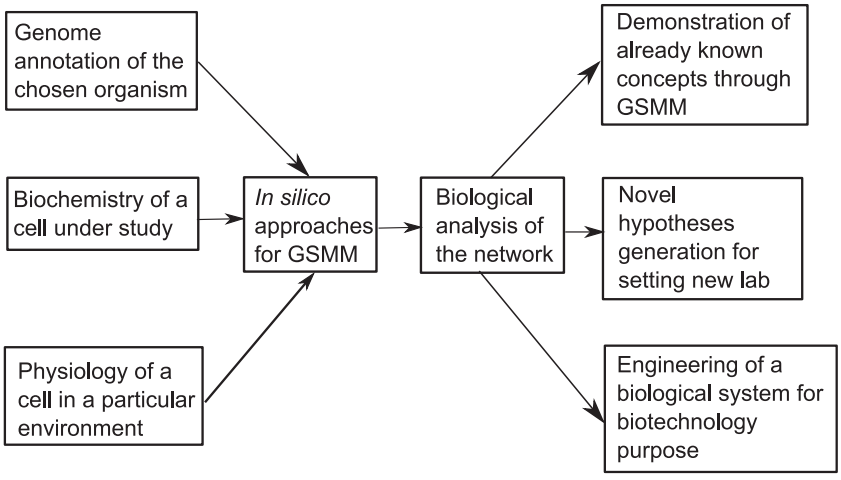

Fig. 1. A general idea over metabolic pathway modeling and analysis. GSMM-Genome scale metabolic modeling. The diagram follows the idea given in Papin et al., 2003

predict a function of an integrated system; 2) modeling and simulation; and 3) comparison of metabolism for getting the functional relationship between different metabolic pathways.

Current applications of metabolic network analysis include finding enzyme(s) that can produce a desired product; finding pathways of maximum yield, such as for antibiotic synthesis; finding non-redundant pathways that are important in drug designing; and genome comparisons by aligning metabolic pathways that help in identifying missing genes (Bower and Bolouri, 2001; Schuster et al., 2000). In the present article, we have surveyed extensively and picked up some relevant information regarding the tasks involved in a metabolic network analysis, like, determining optimal pathways from a substrate to target metabolites through which the amount of targets are maximum (Varma and Palsson, 1994); linking a cellular phenotype to a corresponding genotype (Schuster and Hilgetag, 1994); identifying essential genes and dominant metabolic processes (Edwards and Palsson, 2002; Schwartz and Kanehisa, 2006); analyzing redundancy (Papin et al., 2002; Price et al., 2002); gene knock out studies (Burgard et al., 2003; Segre et al., 2002; Shlomi et al., 2005); finding out alternative optimal solutions from which the same maximal objective can be fulfilled (Mahadevan and Schilling, 2003); and rational strain design (Carlson et al., 2002); finding the active reactions in a metabolic pathway (Beasley and Planes, 2007). The article also provides a detailed comparative analysis among various paradigms, approaches and methods.

The article is organized as follows. We start with a brief description of the existing databases that include various information related to metabolic networks. It is followed by the categorization of existing metabolic network methods in Section 3, which provides a brief description of various approaches. Sections 4 includes a detailed description of comparison of metabolic network analysis approaches under different categories, along with a brief discussion on genome scale metabolic models (GSMMs), its limitations and on the importance of automation of modeling process. Section 5 provides a brief account on metabolic engineering, describes and compares various methods for rational strain design of efficient microbial host strains. Section 6 concludes the article.

2. Sources of data on metabolic pathways

Here we briefly describe some gene, enzyme, reaction and pathway databases that are crucial for metabolic pathway reconstruction and analysis.

2.1. Kyoto Encyclopedia of Genes and Genomes (KEGG)

This database (http://www.kegg.jp) contains information on genes, proteins, reactions and pathways, and useful for building the associations among enzymes, reactions and genes (Kanehisa et al., 2012). Information on human diseases, drugs and other health-related substances is also integrated as perturbations of the KEGG molecular networks. It can be used as a reference knowledge base for integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental technologies. The 'KEGG Organisms' section, which is divided into eukaryotes and prokaryotes, encompasses many organisms for which gene and DNA information can be searched by typing the enzyme of choice. KEGG can be queried through a language based on XML, called KEGG Markup Language (KGML).

2.2. BioCyc

BioCyc (http://BioCyc.org) is a collection of 1690 pathway/genome databases,[1] where each database is dedicated to genome and metabolic pathways of one organism. The BioCyc site contains many tools for navigating and analyzing these databases, and for analyzing omics data. It includes a component called EcoCyc (Keseler et al., 2011) (http://biocyc.org/ecocyc/index.shtml), which is a highly detailed bioinformatics database on the genome and metabolic reconstruction of *Escherichia coli*, including description of *E. coli* signaling pathways and regulatory networks. Another component, called, MetaCyc (Caspi et al., 2010), Encyclopedia of Metabolic Pathways, (http://metacyc.org), contains information on 1787 metabolic pathways and 9609 metabolic reactions (as of December, 2011). It encodes pathways from the experimental literature, and labels each pathway with the organism(s) occurring therein. HumanCyc (http://biocyc.org/HUMAN/organism-summary?object=HUMAN) contains 255 metabolic pathways and 1886 Enzymatic Reactions including other information (as of December, 2011). It places many human genes in a pathway context and in this way, it facilitates analysis of gene expression, proteomics, and metabolomics datasets through a publicly available online tool called the Omics Viewer (Romero et al., 2004).

2.3. ERGO™

ERGO™ (http://www.integratedgenomics.com/ergo.html) integrates genomic and biochemical data from literature and high-throughput analyses into a comprehensive user friendly network of metabolic and nonmetabolic pathways. It is applied to data-mining of target gene discovery, and in silico metabolic engineering and strain improvement. It includes Enzymes and Metabolic Pathways (EMP) (http://www.ergo-light.com/EMP/indexing.html) and its current volume contains around 2.5 million data elements derived from 11,000 publications.[2]

2.4. metaTIGER

It is a collection of metabolic profiles and phylogenomic information on a taxonomically diverse range of eukaryotes (www.bioinformatics.leeds.ac.uk/metatiger/) (Whitaker et al., 2009). Phylogenomic information is provided by 2,257[3] of December, 2011 large phylogenetic trees which can be interactively explored. It also provides novel facilities for viewing and comparing the metabolic profiles.

2.5. ENZYME

It is an enzyme nomenclature database (http://www.expasy.ch/enzyme), a part of the ExPASY proteomics server of the Swiss Institute of Bioinformatics. The database provides reactions, which are catalyzed by certain enzymes. It is primarily based on the recommendations of International Union of Biochemistry and Molecular Biology (IUBMB).

2.6. BRaunschweig ENzyme Database (BRENDA)

A comprehensive enzyme database, BRENDA (http://www.brenda-enzymes.org) (Scheer et al., 2011), allows one to search for an enzyme by its name or Enzyme Commission (EC) number. The majority of the data are manually extracted from the primary literature. It provides detailed enzyme characteristics and kinetic parameters.

2.7. BioCarta

BioCarta (http://www.biocarta.com/) enables one to view and understand how genes interact in dynamic graphical models. It also catalogs and summarizes important resources providing information for over 120,000[4] genes from multiple species. However, BioCarta provides over 18,000 products, including antibodies, proteins, cells, cell-based assays, and detection kits.

[1] As of December, 2011.

[2] As of December, 2011.

[3] As of December, 2011.

[4] As of December, 2011.

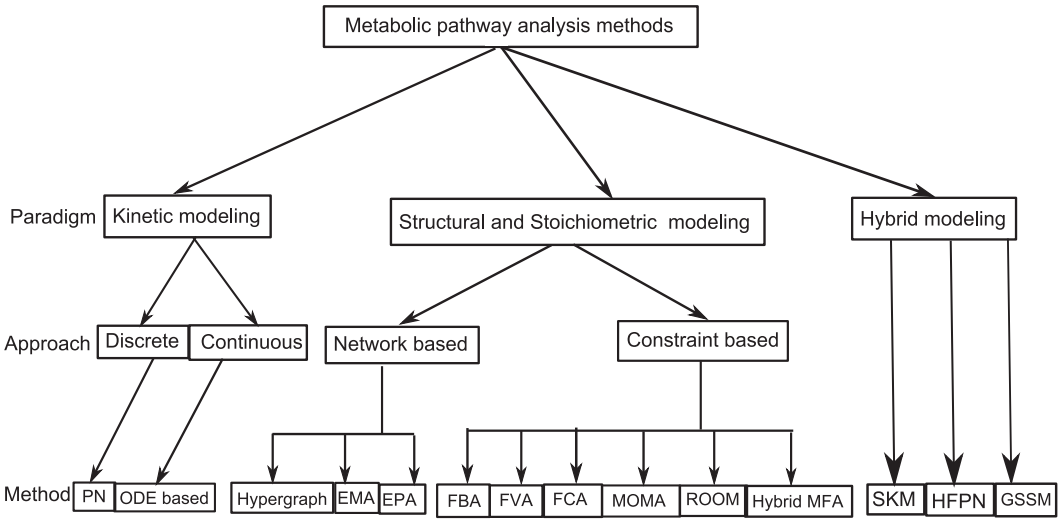

Fig. 2. Categorization of the existing metabolic network analysis methods.

2.8. PUBCHEM

PubChem (http://pubchem.ncbi.nlm.nih.gov/) is a database of chemical molecules and their activities against biological assays. PubChem is organized as three linked databases, viz., PubChem Substance, PubChem Compound, and PubChem BioAssay, within the NCBI's Entrez information retrieval system. It also provides a fast chemical structure similarity search tool.

2.9. Universal Protein Resource (UNIPROT)

Uniprot (www.uniprot.org/) (UniProt Consortium, 2011) provides repository of protein sequences with information on protein functional annotation. It is comprised of four major components, viz., the UniProt Archive, the UniProt Knowledgebase, the UniProt Reference Clusters and the UniProt Metagenomic and Environmental Sequence Database.

2.10. REACTOME

Reactome (http://www.reactome.org/ReactomeGWT/entrypoint.html) navigates pathway knowledge and a suite of data analysis tools to support the pathway-based analysis. It contains species-specific pathways, with each pathway step supported by literature citations (Joshi-Tope et al., 2005). Reactome data model includes reactions and entities, such as, nucleic acids, proteins and small molecules, participating in reactions to form a network of biological interactions and are grouped into pathways. Reactome includes classical intermediary metabolism, signaling, immune system, transcriptional regulation, translation among others.

2.11. Chemical Entities of Biological Interest (CHEBI)

ChEBI (www.ebi.ac.uk/chebi/) (Degtyarenko et al., 2008) is a database and ontology of molecular entities focused on 'small' chemical compounds. ChEBI uses nomenclature, symbolism and terminology endorsed by the International Union of Pure and Applied Chemistry (IUPAC) and Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB).

2.12. ExplorENZ

This enzyme database (http://www.enzyme-database.org/) (McDonald et al., 2009) provides a way to access the data of IUBMB Enzyme Nomenclature List. It uses the regular-expression based pattern-matching facility of MySQL. An associated form-based curatorial application has been developed to facilitate the curation of enzyme data.

2.13. Integrated relational Enzyme database (IntEnz)

IntEnz (http://www.ebi.ac.uk/intenz/) is a freely available resource for enzyme nomenclature (Fleischmann et al., 2004). IntEnz is created in collaboration with the Swiss Institute of Bioinformatics (SIB) for the production of the ENZYME resource.

2.14. Protein ANalysis THrough Evolutionary Relationships (PANTHER)

It is designed for high-throughput functional analysis of large sets of protein sequences, and has been used to annotate the human genome (Thomas et al., 2003). It uses published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence. It has major metabolic pathways, viz., glycolysis, TCA cycle and pentose phosphate pathway.

3. Existing metabolic network analysis methods: A Categorization

The methods for analysis of metabolic systems are mainly based on kinetic data and structural/stoichiometric modeling, however, it also includes hybrid modeling techniques. A categorization of the methods is given in Fig. 2.

3.1. Kinetic modeling

Quantitative description of the changes in concentrations of various molecules is a challenging task as biological processes are time dependent, which identify the flow of mass between cellular components. A kinetic reaction network consists of biochemical reactions that can be traditionally described by nonlinear ordinary or partial differential equations (ODEs or PDEs). Kinetic models can easily demonstrate oscillations and bi-stability. Under kinetic modeling, there are two kinds of approaches: discrete and continuous.

3.1.1. Discrete approach

Experimental data are obtained by sampling the continuous biochemical reactions at discrete time points. One can get pathway reaction dynamics through discrete time models and generates predictions which rules out the expensive and time consuming lab experiments. Discrete time models act as an interface between

experiments and computational simulations. Petri Net (PN) modeling, developed by CA Petri in 1960 (Petri, 1962), is an example of discrete time modeling approach. In PN modeling approach, concentrations of molecules at an instant of time are expressed as the discrete number of tokens, assigned to a place (state of a metabolite). A transition is said to fire when its input places have the minimum number of tokens specified in the corresponding arc weights. A metabolic process or kinetic law of a reaction can be demonstrated by transition rate in PN terms. Substrate and product of a metabolic reaction are placed as input and output arcs respectively. Number of tokens on places are concentrations of metabolites, enzymes or compounds in PN terms. Manipulation and analysis of functionality of a cell can be done through PN modeling of a complex metabolic network (Baldan et al., 2010; Chaouiya, 2007). Structural validation of metabolic networks through qualitative analysis (Koch et al., 2005) has been done through PN theory, where sucrose breakdown metabolism in a potato tuber has been studied. One can use Colored PNs (CPNs) to simulate enzymatic reaction chains, where, associated color is a pair encompassing the name and the concentration of the related substrate (Hartmann Genrich and Voss, 2001). A systematic approach for modeling regulated metabolic pathways with PNs has been developed (Simao et al., 2005).

3.1.2. Continuous approach

Continuous approach simulates the continuous time dependent behavior of a pathway system quantitatively using differential equations. Differential equations specify the concentrations of various molecular species evolve over time, however, complex numerical computation is one of its drawback. The nonlinearity in the behavior of an enzyme is commonly shown by mass action or Michaelis–Menten like kinetic models. A model has been developed to study the diauxic growth of *E. coli* in glucose and lactose with 13 state variables (Wong et al., 1997). A Cellerator language extension, kMech has been introduced that describes enzyme mechanisms (Yang et al., 2005). Continuous approach requires kinetic measurements, however, only a few parameters are known experimentally. Experimental data is limited and noisy, and also estimating parameters of continuous ODEs is computationally costly.

3.2. Structural and stoichiometric modeling

Underlying biological knowledge present in a metabolic network can be translated into mathematical terms through stoichiometric matrix. Analysis of structural invariants of a biochemical network has been first applied to a system of chemical reactions (Milner, 1964). Stoichiometric network analysis (SNA), based on convex analysis, was the pioneering work that identifies unique pathways. It has also been used for investigating steady states and stabilities of dynamical systems (Clarke, 1981; Feinberg and Horn, 1974). One can study the qualitative, nonlinear dynamics of chemical reaction mechanisms and other systems with stoichiometry through SNA (Clarke, 1988). One can avoid the problems of developing the kinetic models that involve a lot of intracellular experimental measurements (Llaner and Pico, 2008) through adapting structural and stoichiometric modeling. Stoichiometric modeling can be either network-based or constraint-based techniques.

3.2.1. Network based approaches

Network based approaches describe systemic functions mathematically through the existing knowledge of the cellular components and their connectivities. The three basic methods under this approach are — hypergraph based, elementary flux mode analysis (EMA) and extreme pathway analysis (EPA).

Graph representation of a biological pathway has been used in studying robustness, modularity, connectivity, finding reaction clusters, network structure analysis, shortest path finding, path enumeration, and reaction clustering. Graph-theoretic concepts are being used to predict the structural and dynamical properties of a metabolic network (Aittokallio and Schwikowski, 2006; Albert, 2007). Detailed information on the application of hypergraphs for biochemical pathways has been given in Klamt et al. (2009). Elementary Mode Analysis, introduced by Schuster and Hilgetag (1994), is useful to identify the structure of a metabolic network. An Elementary Flux Mode is a minimal set of enzymes that can operate at steady-state such that all irreversible reactions involved proceed in the thermodynamically favored directions. All metabolic capabilities under steady states are composed of elementary flux modes (Schuster et al., 1999; Trinh et al., 2009) and in this way, it links cellular phenotype to the corresponding genotype. Some of the publicly available softwares for calculating elementary modes are METATOOL (Pfeiffer et al., 1999; von Kamp and Schuster, 2006), GEPASI (Mendes, 1993), COPASI (Hoops et al., 2006) and FluxAnalyzer (Klamt et al., 2003). Extreme pathways consider all the necessary reaction steps of a network that must be used to complete the synthesis process. The internal reversible reactions are decomposed into two irreversible reactions but reversible exchange reactions are not decomposed for the calculation of extreme pathways. Extreme pathways represent the edges of the steady-state flux cone derived from convex analysis, and they can be used to represent any flux distribution obtained by a metabolic network (Schilling et al., 2000). Extreme pathways can be characterized by their length and reaction participation. These properties have been computed for the production of individual amino acids and protein production in *Helicobacter pylori* and individual amino acid production in *Haemophilus influenzae* (Papin et al., 2002). Moreover, RBC metabolism and its metabolic physiology have been interpreted as an application of extreme pathway (Wiback and Palsson, 2002).

3.2.2. Constraints based approaches

The other approach under structural and stoichiometric modeling is constraint based that applies a set of constraints on a metabolic pathway to characterize its possible behaviors. There exist many constraint based methods, viz., FBA, FVA, FCA, MOMA, ROOM, MFA and Hybrid MFA. Some toolboxes are developed under constraint based approaches, include, COBRA and FASIMU. Constraint-based reconstruction and analysis (COBRA) (http://gcrg.ucsd.edu/Downloads/CobraToolbox) toolbox is for quantitative prediction of cellular behavior using a constraint-based approach (Feist et al., 2007). It allows predictive computations of both steady-state and dynamic optimal growth behavior, the effects of gene deletions; and robustness analyses. Another command line oriented software is Flux-balance Analysis based SIMUlations (FASIMU) (http://www.bioinformatics.org/fasimu), which is suitable for a wide variety of FBA algorithms and can handle batch series of flux-balance optimizations (Hoppe et al., 2011). It computes flux distributions using a variety of FBA algorithms, including the first available implementation of weighted flux minimization, fitness maximization for partially inhibited enzymes, and using the concentration-based thermodynamic feasibility constraint.

The classical starting point of constraint-based modeling is flux balance analysis (FBA), a mathematical method for analyzing the metabolic capacity of a cell. The objective of FBA is to find out the set of metabolic fluxes that maximizes the growth rate of a target metabolite, given some known available nutrients. The earliest work in FBA includes that of Papoutsakis (1984) that demonstrated the way to construct flux balance equations using a metabolic map. However, Watson (1984, 1986) has first introduced the use of linear programming (LP) and an objective function to determine an optimal path. The success of FBA can be seen in the ability to accurately predict the growth rate of the prokaryote *E. coli* when cultured in different growth media (Edwards et al., 2001), and to define precise minimal media for the culture of *S. typhimurium* (Raghunathan et al., 2009).

Minimization of metabolic adjustment (MOMA) (Segre et al., 2002) is a quadratic programming (QP) based algorithm that addresses the issue of mutants/knockouts in the cases, where assumption of optimality is not justifiable. It identifies a point in flux space,

which is closest to a wild-type point, compatible with the gene deletion constraint. The outcome of the method has been supported by experimental data for *E. coli* knockout growth rates. A series of transient metabolic changes occurs due to a regulatory system. Thus predicting the metabolic state of an organism after a gene knockout is a challenging task. Regulatory on–off minimization (ROOM) (Shlomi et al., 2005), has been introduced for predicting the metabolic steady state after gene knockouts. This can be taken care of through implementing ROOM, which aims at minimizing the number of significant flux changes, hence on–off, with respect to the wild type.

A key issue of constraint based models is the existence of alternate optimal solutions for the same maximal objective, which can be achieved through different flux distributions. Flux Variability Analysis (FVA), an LP-based method, has been described to calculate the range of flux variability that achieves optimal as well as suboptimal objective states (Mahadevan and Schilling, 2003). The tool fastFVA, an open source implementation of FVA, has been developed (Gudmundsson and Thiele) to study flux distributions under suboptimal growth, network redundancy, and is used for optimal strain design and optimization of process feed formulation for antibiotic production.

Flux Coupling Analysis (FCA) classifies reactions into subsets of coupled reaction sets in which activity of one reaction initiates activity of another reaction. Flux Coupling Finder (FCF), variation of FCA, solves LPs to find out coupled reactions (directionally, partially or fully coupled) and blocked reactions in genome-scale metabolic (GSM) systems. It elucidates the topological and flux connectivity features of genome-scale metabolic networks (Burgard et al., 2004). The method, Minimal Metabolic Behaviors (MMB-FCA) (Larhlimi and Bockmayr, 2009), computes a minimal set of generating vectors of the flux cone. It infers the coupling relation for any pair of reactions based on the co-appearance of nonzero fluxes in the generating vectors. Feasibility-based Flux Coupling Analysis (FFCA) (http://www.bioinformatics.org/ffca/) (David et al., 2011) finds a feasible solution, which is faster than computing an optimal solution. Flux patterns only contain the information about the activity or inactivity of the fluxes, but not regarding the flux values, thus the method does not distinguish between partial and full coupling. In FFCA, finding the first feasible solution is sufficient, while the LPs are solved for optimality in the case of FCF. FCF decomposes every reversible reaction into two (forward and backward) irreversible reactions, which slows down the procedure and increases the size of the LPs to be solved.

Metabolic Flux Analysis assumes the pseudo-steady state, in which the net balance between the consumption and production fluxes should be equal. It provides a flux distribution, which is a constant vector obtained by solving a linear system constructed from the stoichiometric matrix. As stoichiometric models constitute the basic framework for fluxome quantification, a hybrid method has been developed combining classical metabolic flux analysis (MFA) (Provost and Bastin, 2006) and projection to latent structures (PLS) (Svante and Sjostroma, 2001). Hybrid metabolic flux analysis (Hybrid MFA) (Carinhas et al., 2011) extends the study of Bernal et al. (2009), where rational strategies for Baculovirus production optimization in insect cell cultures have been developed. It is based on classical Metabolic Flux Analysis (MFA). Hybrid MFA is a suitable tool for metabolic identification and quantification in incomplete metabolic networks. There also exists ^{13}C metabolic flux analysis that utilizes ^{13}C labeling patterns of metabolic products (A. MR et al., 2007) and computes in vivo metabolic fluxes by carbon isotopomer balances.

3.3. Hybrid modeling approaches

There exist certain complex biological systems that show both continuous and discrete dynamic behaviors. Hybrid system based modeling approaches are commonly used to model dynamic systems showing both of these behaviors as it deals with different time-scales to account for dynamic modes (Lincoln and Tiwari, 2004). There

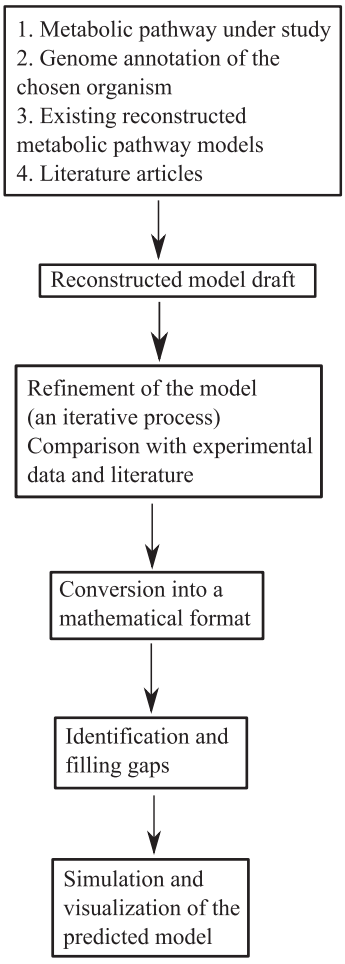

Fig. 3. GSMM steps.

are two methods under hybrid modeling, like, Structural kinetic modeling (SKM), Hybrid Functional Petri Nets (HFPNs) and Genome scale metabolic modeling (GSMM). However, we have discussed Genome-scale metabolic modeling and importance of automation of modeling process in this Section only.

Structural kinetic modeling (SKM) incorporates both the stoichiometric information and the dynamic aspects of a biological system (Steuer and Junker, 2009; Steuer et al., 2006). SKM represents a metabolic system, in the form of a Jacobian matrix, at each possible point in parameter space. Each element can be accessed without explicit knowledge of the functional form of the rate equations. It evaluates an ensemble of possible models, rather than evaluating a single model. It aims at the statistical evaluation of the Jacobian matrix, where each element of the matrix represents the available experimental information.

Discrete transitions can be set to fire with an associated time-delay. Thus the idea of using Hybrid Functional Petri Nets (HFPNs) is well suited that includes continuous places and transitions; and different time-scales and describes the systems dynamics (Matsuno et al., 2003). The regulation of urea cycle in liver is studied using HFPN (Chen and Hofestadt, 2003), which shows that the defects of the enzymes in the urea cycle can be treated by limiting the input of ammonia or by replacing the missing intermediates from the cycle, by supplementing with arginine or citrulline.

Genome-scale metabolic models (GSMM): Reconstruction of genome-scale metabolic (GSM) networks is being done through the availability of sequenced and annotated genome. There are five steps in the reconstruction of a GSMM, viz., (1) creation of a draft model, (2) reconstruction of the model, (3) conversion into a mathematical framework, (4) identification and filling of gaps,

and (5) simulation and visualization of the predicted models (Thiele and Palsson, 2010). It is shown in Fig. 3. Stoichiometry, constraint based approaches and in silico simulation have enabled the phenotypic functional analysis of GSMMs. GSMMs can be used in the investigation of the effects of gene deletion (Imielinski et al., 2005; Doerr, 2010), and finding out the essential genes (Gerdes et al., 2006; Hillyard and Redd, 2007) and the metabolic concentrations (Imielinski et al., 2006). The differences in metabolite concentrations under known environmental conditions are mapped onto GSM networks and then integrated with transcriptomic data to study the effects of metabolic regulation.

A remarkable regulatory-metabolic network of *Saccharomyces cerevisiae* GSMM was reconstructed, which possessed 55 regulatory transcription factors regulating 348 metabolic genes (Herrgard et al., 2006). Based on metabolic flux analysis, global transcription regulation impact by global regulatory proteins on glucose catabolism in *E. coli* was also investigated (Perrenoud and Sauer, 2005). A GSM network iSB619, of *S. aureus* strain N315, has been reconstructed, which consists of 619 genes (enzymes) that catalyze 640 metabolic reactions, 537 proteins, 571 metabolites. Analysis of model capabilities has been done to form hypotheses, e.g., the efficiency of growth on different carbon sources, and potential drug targets (Becker and Palsson, 2005). Metabolite coupling is the term for a pair of metabolites that appear together in the same biochemical transformation, like ATP, which couples with ADP, Pi and H + in reactions that transfer the phosphate moiety. Becker et al. (2006) has given a measure for topological analysis of metabolite coupling in GSM networks.

The process of reconstruction of GSMM has been shown to follow by 96 steps (Thiele and Palsson, 2010), the model SEED (http://www.theseed.org/models/), a web-based resource designed to automate the creation of new metabolic models (Henry et al., 2010). It generates functional draft metabolic models of an organism from a genome sequence. SEED pipeline is integrated with SEED framework to couple genome annotation and metabolic reconstruction. The assembled genome sequence is annotated by the RAST server (http://rast.nmpdr.org/) and imported into the SEED analysis system, where a preliminary model is generated. Additional intracellular and transport reactions are added to create an analysis-ready model in the auto-completion step of the pipeline. The final three steps of the pipeline involve the removal and addition of reactions to better fit any existing experimental growth phenotype data.

The stoichiometric matrix can also be linked explicitly to the gene expression data through connecting each chemical or transport reaction to the corresponding proteins and genes (Terzer et al., 2009). Delineating the gene–protein–reaction relationships enables the comparison of computational network analyses with experimentally determined gene knockout phenotypes. Gene expression data and gene-transcription factor (TF) relationships should be explored more to get the metabolic behavior. There can be inconsistencies present between experimental and model predicted fluxes. In order to find the in vivo active reactions in a GSM network, optimal metabolic network identification (OMNI) method has been introduced (Herrgard et al., 2006). Sequenced genome and high-throughput data also give understanding of microbial physiology at the systems level, which can be further explored for increasing the production of fuels, chemicals, biotechnological and pharmaceuticals products.

Metabolic expression placement (MEP) method (Kharchenko et al., 2004) is based on the coexpression properties of the metabolic network to identify genes encoding missing metabolic enzymes in a partially reconstructed metabolic network. A method combining a local structure of metabolic network and functional association evidence, like, clustering of genes on the chromosome, similarity in phylogenetic profiles, gene expression and protein fusion events, among others, to predict genes encoding metabolic enzymes (Kharchenko et al.). The authors have claimed that this method does not rely on direct sequence homology and can be used with homology-based metabolic reconstruction methods. We have described optimization based approaches in Section 5.1.2 of this article. Optimization-based procedure was shown to identify and eliminate gaps in metabolic network reconstructions, introduced new reactions, and modified existing reactions (Kumar et al., 2007). They have first identified the metabolites that cannot be produced under any uptake conditions, which was followed by the identification of the reactions from a customized multi-organism database and thus connects these metabolites to the parent network. However, most of the existing algorithms do not consider the impact of global changes on the quality of models.

3.3.1. Limitations of GSMM

Enzyme kinetics, transcriptional and metabolic regulations have a significant effect on reaction fluxes. Enzyme kinetics is a function of metabolite concentrations and enzymatic activities, which show changes as a result of transcriptional and metabolic regulation (Smallbone et al., 2007). Here, many parameters are unknown or difficult to measure, thus model prediction for large GSM systems is a very challenging task, specially, when one works with kinetic approach. Constraint-based models have shown to overcome these problems as they consider average reaction rates achievable by cells grown in steady or slowly varying environmental conditions. Here, the law of conservation mass followed, where the production and consumption rates of metabolites need to be balanced. These approaches are mainly based on reaction fluxes and do not consider metabolite concentration or regulation. However, there are certain efforts to take care of these limitations. They include thermodynamics (Beard et al., 2004; Henry et al., 2007) and regulatory interactions (Covert et al., 2001). After completing the modeling process, the model is refined with experimental data, however, it should be done in limit as extensive use of experimental data to predict a model makes it more descriptive instead of a predictive one.

3.3.2. Automated metabolic network model reconstruction

We have discussed the idea of automation of metabolic network model reconstruction as a current challenge in metabolic pathway model development and analysis. Network model reconstruction was done manually earlier. After that, several pathway structure, reaction and molecule interaction databases and tools are being developed to automate the process of reconstruction. Traditionally, a functional description is assigned to a given gene in gene annotation, whereas the missing gene problem assigns a gene to a specific metabolic function (Osterman and Overbeek, 2003). Therefore, we can say that the identification of an enzyme catalyzing a metabolic reaction in a well/partially characterized metabolic network is the solution of the missing genes problem. Function to a gene can be assigned using gene finding algorithms, sequence homology searches, and non-homology based algorithms for previously unknown genes. However, there are some drawbacks of these approaches as information is being retrieved from databases and similarities exist in chosen organisms, but it is also possible that similar enzymes can have different functionalities in different organisms (Blazeck and Alper, 2010). Moreover, other challenges are: assigning a function to an enzyme that may catalyze several reactions, enzymes with broad substrate specificity, among others. There are organisms' genomes that are yet to be sequenced, and their pathways, enzymes, or metabolites are uncharacterized. Ultimately, experimental/biochemical validation of enzyme is a must as minor differences in sequence can alter enzyme function.

4. Comparison of metabolic network analysis approaches under different categories

In this section, we throw light on comparative aspects of the above mentioned paradigm, approaches and methods. Moreover, it clearly articulates the differences and relationships among them, which can be considered appropriately by users.

4.1. Comparison among paradigm

A metabolic network can be interpreted as a bipartite graph for a structural view, consisting of two sets of nodes that represent metabolites and biochemical inter conversions, respectively. This bipartite graph may either be collapsed into a substrate graph (Jeong et al., 2000), or into a reaction graph (Wagner and Fell, 2001). As a metabolic network is a network of biochemical inter conversions, several aspects of metabolic networks differ fundamentally from the networks of cellular interactions. Stoichiometric analysis has gone beyond merely applying topological approach and taken the physicochemical properties of metabolic networks into account, based on the null space of complex reaction networks.

If the network structure and certain information about input and output fluxes are available, the intracellular steady state fluxes can be estimated utilizing constraint based approach, like, flux balance analysis. Although it is has been applied widely, it has some drawbacks. The reactions have defined stoichiometries, and consequently, certain flux distributions are not allowed. FBA may select the optimal solution that might be highly undesirable regarding its dynamic property. FBA itself is not sufficient to uniquely determine intracellular fluxes and is incapable in solving parallel metabolic routes, reversible reactions, cyclic fluxes and futile cycles, in particular.

While stoichiometric analysis has proven effective to analyze the functional capabilities of large metabolic networks, it does not work on dynamic aspects of the system. The most profound shortcoming of stoichiometric network analysis, is the assumption of steady state balance equation that does not give any idea about the stability or possible instability of a metabolic state. This is the well known fact that metabolism is a complex dynamical system, where multiple steady states, feedback loops, bifurcations and temporal oscillations are present to add their dynamical behavior. Another important aspect is fluctuations in the concentrations of biomolecules that may require special treatment in modeling the biological networks. It also involves the transcriptional regulation of gene expression. These factors have been advocated for the quantitative analysis of biological networks through kinetic modeling. It helps in modeling the time evolution of molecular species in a reaction system and the traditional way is to use a set of ordinary differential equations (ODEs). It also relies on the enzymatic rate equations and their associated parameter values. However, these dynamic models encompass a number of limitations. Some of the parameter values can be available from the literature, but some of them may also depend on many other factors, like, tissue type or experimental conditions. It has also been found that the most enzyme-kinetic rate laws have been determined in vitro, but it may not be described well if a particular rate law is appropriate in vivo. We can say that it is still hampered by inadequate knowledge of the enzyme-kinetic rate laws and their associated parameter values due to formidable progress in experimental accessibility of system variables.

Thus there is a necessity for hybridization of existing pathway analysis paradigms. Continuous models of biological systems can be too large and complex to be simulated and analyzed. On the other hand, a fully discrete approximation of the model can sometimes loses crucial information (Lincoln and Tiwari, 2004). Hybridization of continuous and discrete modeling approaches helps here, providing desired levels of abstraction, approximation, and simplification of biological pathways. Hybrid systems are mathematical models that combine continuous dynamical systems with discrete transition systems. In a hybrid system, the continuous dynamics of time-varying variables are given using differential equations. In models from biology, the differential equations specify how the concentrations of various molecular species evolve over time. Hybrid discrete–continuous models provide more complete and scalable analysis.

4.2. Comparison among approaches

We have categorized metabolic network analysis approaches into network based, constraint based, discrete and continuous approaches. This subsection mainly deals with the requirement of using certain approaches for metabolic network analysis. Network based metabolic pathway approaches are simply a mathematical description of systemic functions of the cellular components and connectivities among them. It works on an idea to decompose a metabolic network into basic functional and structural units but does not deal with time scales at which certain biological events occur. It only assesses the inherent topological and structural properties in biochemical reaction networks. This problem can be solved through applying constraint based approaches, which deal with the basic constraint, especially, mass balance conservation within the system. The core feature of constraint based approaches is a stoichiometric representation of all metabolic reactions that imposes constraints on the flow of metabolites through the network. It assumes that a system should be in steady state, where the matrix of stoichiometries imposes flux balance constraints on the system.

To know the dynamics of how the constituent components of a biological system interact is crucial as it defines the spatiotemporal response of the system to a stimulus. Thus the kinetics of biological processes has been modeled that represent a biophysical system. For example, discrete elements can be explained as a control system in the form of transcription factors that can turn on/off the gene expression (Lincoln and Tiwari, 2004). The delay concept of the discrete transition can describe the transcription which occurs after a certain time. On the other hand, transcription, translation, and enzymatic and metabolic reactions are those biological events whose conditions change continuously. Another aspect is the way of experimentation. Although, biological reactions occur in a continuous time process, sampling for experiments is done at discrete time points. A major shortcoming for logical modeling under discrete approach is that it is a time-consuming process and thus infeasible for large systems. Continuous modeling approach can simulate the behavior of a system in a quantitative way through time. Differential equations have been used to specify how the concentrations of various molecular species evolve over time. On the other hand, discrete components of biological systems' models arise from state transitions (e.g. from healthy to abnormal states). This is the reason, one should look for hybrid modeling approaches to understand all possible behaviors of a biological system. Once a hybrid model is obtained, analysis techniques can be applied to uncover the dynamics of the biological system of interest. As mentioned before, Steuer et al. (2006) shows the importance of hybridization of different approaches exploring study the dynamical capabilities of a metabolic system in a quantitative way, without requiring functional form of the rate equations.

4.3. Comparison among methods

Here we compare the feature of various methods/tools under different approaches. There are three methods under network based approaches, viz., hypergraph based, elementary mode and extreme pathway analyses.

4.3.1. Comparison of EPA and EMA

Elementary modes represent basic and non decomposable subnetworks of system that can operate at steady-state, and are the edges of a convex polyhedral cone in the space of flux distributions. Extreme pathways are conceptually closely related to elementary modes, but they rely on a network reconfiguration, where reversible reactions are decoupled into the corresponding forward and backward reactions. The sets of elementary modes and extreme pathways may coincide in certain network topologies. The features of these two methods are well described in Klamt et al. (2003). Extreme pathways form a subset of elementary modes. The sets of extreme pathways and elementary modes can be identical if all the reactions (including both internal and exchange reactions) are irreversible in a metabolic network. Therefore, the identification of extreme pathways depends on the configuration of the metabolic network while the identification of elementary modes does not. Some elementary modes are not systematically independent, but they are genetically independent as

it has direct implementation of the non-decomposability constraint. These two methods give good results for metabolic pathway analysis but they are associated with the problem of combinatorial complexity and thus cannot be applied to medium or large scale networks. In other words, the algorithms for identification of elementary modes and extreme pathways do not scale well for genome-scale models of complex microorganisms, due to combinatorial explosion (Klamt and Stelling, 2002).

Comparison of the methods under constraint based approach: We have briefly described several methods under constraint based approach and comparison among them is as follows. If we compare ROOM with FBA, we can find that FBA explicitly maximizes the growth rate, while ROOM implicitly favors flux distributions having high growth rates. MOMA does not assume optimality of growth or of any other metabolic function, in contrast to FBA. Instead, MOMA approximates metabolic phenotype by performing distance minimization in flux space under perturbations. In case of lethality of some *E. coli* gene deletions, MOMA, unlike FBA, has been able to predict correctly.

Elementary flux patterns can be applied to the computation of minimal media, the development of knockout strategies, and the analysis of combined genome-scale networks. In this way, one can say that elementary flux pattern is an application derived from the concept of elementary mode analysis to genome-scale metabolic networks. Moreover, it does not have the drawbacks that arise due to exchange fluxes. Comparing FBA and elementary flux pattern analysis, we can say that both can be used to find a pathway producing a certain metabolite. However, elementary flux pattern analysis is better for the exhaustive enumeration of pathways in a subsystem. Elementary flux pattern also uses a linear programming method to determine a global pathway using the reactions of a flux pattern in a subsystem. Thus, elementary flux pattern analysis can be considered as a combination of elementary mode analysis and FBA.

Taking the case of Flux Coupling Analysis (FCA), MMB-FCA, a constraint based approach, uses an outer description of the steady state flux cone, based on sets of irreversible reactions while the other methods, such as elementary mode analysis or extreme pathway analysis, uses an inner description, based on sets of generating vectors.

Other important methods, like analysis through applying Petri Nets (PNs) comes under kinetic modeling paradigm. Standard PNs have a disadvantage of incapability to deal with continuous type of biological processes. Here concentrations of mRNA and protein should be stored in continuous places, while rate of reactions of transcription and translation should be assigned at continuous transitions. In case of modeling metabolic reaction through PNs, e.g., Michaelis–Mentens equations can be modeled through assigning concentrations of substrate and product to continuous places and the formula for the reaction to the continuous transition between these two places. To deal with this problem, hybrid functional petri nets (HFPNs) has been introduced, which have both discrete and continuous parts to represent discrete elements (discrete place and discrete transition) and continuous elements (continuous place and continuous transition), respectively, of a biological system. Another characteristic of HFPN formalism is that the quantification of consumed resources can be different from the quantification of the resources produced.

5. Metabolic engineering

Metabolic engineering combines systematic analysis of metabolic and other pathways with molecular biological techniques to improve cellular functionalities by designing and implementing rational genetic modifications (Koffas et al., 1999). In general, we can say that metabolic engineering measures metabolic fluxes and elucidates their control as determinants of metabolic function and target product yield. Conventionally, the “rate-limiting step” is identified in the first place in a given metabolic process. Moreover, to overcome this shortcoming, either the heterologous gene(s) responsible for affecting the rate-limiting step(s) is overexpressed or the inefficient pathway(s) that contributes to by-product formation is inactivated (Martin et al., 2003; Sauer et al., 1997). However, more complex aspects of metabolism cannot be altered by manipulating the metabolic

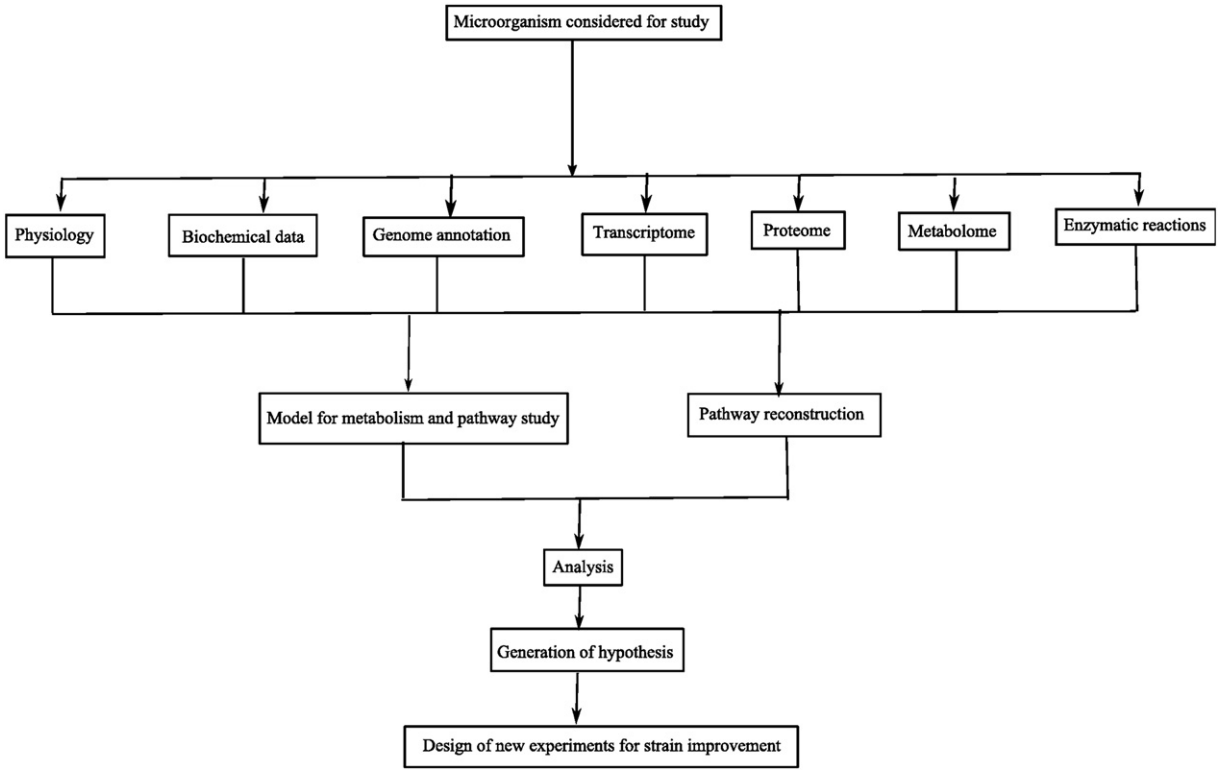

Fig. 4. Work flow to design efficient microbial strains.

gene(s) due to the complex network of regulatory mechanisms of the cell (Vemuri and Aristidou, 2005). Thus it is justified that now the focus is shifted towards engineering the regulatory control mechanisms in a desired way so that microbial cellular functions for certain purposes can be improved. It can exploit genome-scale metabolic pathways of microbes to replace chemical production processes with biotechnological routes based on microbial fermentations. It can be done through integrating rational design procedures, mathematical tools and experimental techniques. Mathematical models have an important role in solving optimal metabolic path problem and suggesting changes in the genotype of the microorganism for an improved metabolic network. Consequently, molecular biology techniques can be used for experimental validation.

Metabolic engineering has broadened its design methodologies in the post-genomic era through incorporating various kind of biological data, viz., biochemical, genome-scale metabolic, transcriptomic and metabolomic data. Genome-scale stoichiometric models of microorganisms represent a first step in this direction (Patil et al., 2004). Moreover, metabolic flux analysis using ^{13}C-labeled substrates has become an important tool in metabolic engineering. It allows the detailed quantification of all intracellular fluxes in the central metabolism of a microorganism (Wiechert, 2001). Fig. 4 shows a possible way to design of efficient microbial host strains.

There are assumptions and constraints decided to build a specific model (Blazeck and Alper, 2010). Initially, these in silico models are used to assess whether the model can predict biomass production in general. The first metabolic reconstructed model is assumed to be neither sufficiently accurate nor complete, thus in silico approaches like, FBA, filling gaps, comparative gene knockout data are applied to improve the model prediction. This is an iterative process that contains validation and assumption correction steps. There are some challenges for models in metabolic engineering: sometimes non-suitable hosts, lesser concentration of substrates, a shortage/accumulation of intermediates, competing pathways, and/or restricted storage capacity for end products hamper the introduction of foreign pathways into a host. However, this results in poor or no growth (Klein-Marcuschamer et al., 2007). We now describe rational design of efficient microbial host strains under metabolic engineering.

Table 1
Methods for strain optimization and corresponding description.

Methods	Description
Elementary modes and Extreme pathways	Network based approaches
OptKnock	Based on gene knockouts aiming at maximizing target product yield
OptStrain	Removes competing functionalities that divert flux away from targeted product
OptGene	Evolutionary algorithm based gene deletion approach
OptReg	Based on gene knockouts, over-and down regulations of reactions
GDLS	Based on local search with multiple search paths
OptORF	Identifies the optimal metabolic and regulatory gene deletions
OptFORCE	Identifies all possible engineering interventions for a wild-type strain
CASOP	Finds relative contribution of a reaction derived from weighted elementary modes
OptFlux	User-friendly, open-source modular software for metabolic engineering
EMILIO	Uses successive linear programming (SLP)

5.1. Rational design of efficient microbial host strains: An application to metabolic engineering

Microbial strains can be manipulated in order to improve their product yield and/or improved growth characteristics. One needs to have all unique pathways existing in a metabolic network to design efficient host strains with specialized metabolic functionalities. This will lead to the elimination of inefficient pathways and force host strains to function only according to efficient pathways. Three steps have been described in Trinh et al. (2009) to design the novel host strains. The first step is to identify all unique pathways, which can be done applying elementary mode analysis (EMA). The second step is pathway selection, in which a strain can be designed such that growth and operation of the efficient pathways can be coupled. The third step is pathway evolution (optional) in which the already designed strain can be utilized. This approach has been used to design and construct minimal cells with efficient metabolic functionalities for biomass production. The most efficient biomass producing *E. coli* TCS062 has been designed with 6 genetic knockouts (Trinh et al., 2006). Similar strategy has been used to design a minimal *E. coli* cell for ethanol production (Trinh et al., 2008). The designed mutant TCS083/pLOI297 has 8 genetic knockouts, and has been able to efficiently convert hexose and pentose to ethanol. It has also been closer to the theoretical limit without glucose catabolite repression.

Metabolic reaction fluxes are a fundamental determinant of the cell physiology. Computational procedures for predicting metabolic interventions leading to the overproduction of biochemicals in microbial strains are widely in use. Existing approaches can be distinguished between pathway based and optimization based approaches. Following is the brief description of the available tools for phenotypic simulation of microorganisms. Table 1 gives a one line description of existing tools for design of microbial strain through metabolic engineering.

5.1.1. Pathway based design

Rational metabolic engineering requires powerful theoretical methods for pathway analysis, in which the topology of metabolic networks is considered. There exist a few studies to illustrate the applicability of extreme pathways and elementary modes in metabolic engineering. All metabolic capabilities in steady states are composed of elementary flux modes, which are minimal sets of enzymes that can generate valid steady states.

Minimal medium requirements have been defined for *H. influenzae* and *H. pylori* using extreme pathways (Schilling et al., 2000, 2002) and this has been found to be highly consistent with experimental results. Dien and Lidstrom (2002) demonstrated both computationally and experimentally that *M. extorquens* AM1 had several redundant pathways that one pathway could compensate the role of other. They also reconstructed the C3–C4 pathways based on genome sequence data and simulation results. Elementary mode analysis has been applied to a intermediary metabolism of *S. cerevisiae*, based on experimental data from a strain of *S. cerevisiae*, and elementary modes have been used to study the effects of gene additions and deletions on the yield values for production of poly-β-hydroxybutyrate in a recombinant strain (Nalecz et al., 1991). It has also been used in a framework for assigning functions to orphan genes using metabolomic data for a model of core metabolism in *S. cerevisiae* (Boles et al., 1998). The modes of the fructose-2,6-bisphosphate cycle, the combined tricarboxylic-acid-cycle–glyoxylate-shunt system and tryptophan synthesis have shown their applicability in metabolic engineering (Schustera et al., 1999). This approach can be used in many biotechnological applications such as increasing the yield of a product, channeling a product into desired pathways and for a functional pathway reconstruction from genomic data. This concept can also be used in further refinement of poorly characterized biological networks or finding the missing links by characterizing the function of mutants with knocking out the desired gene.

5.1.2. Optimization based design

An alternative approach for metabolic engineering includes optimization-based methods, where models use the same stoichiometric and thermodynamic constraints, but solutions are identified through optimization (through maximizing or minimizing the objective function). Following is a brief description of available optimization based tools and algorithms for rational design of efficient microbial host strains.

5.1.2.1. OptKnock. OptKnock (Burgard et al., 2003) is a computational procedure, which selects the reactions to remove from a metabolic network. It can be done by deleting the gene(s) associated with the identified functionality. The basis of OptKnock is to ensure the drain towards growth resources, i.e., carbon, redox potential and energy that must be accompanied by the production of a desired product. The procedure was demonstrated based on gene deletions for succinate, lactate and 1,3-propanediol (PDO) production.

In an extended application of OptKnock (Pharkya et al., 2003), the transport rates of carbon dioxide, ammonia and oxygen, as well as the secretion pathways for key metabolites, are introduced as optimization variables in the framework, in addition to gene deletions. These strategies range from the central metabolic network genes, and the amino acid biosynthesis and degradation pathways. They have well demonstrated the importance of manipulating energy-producing/consuming pathways, controlling the uptake of nitrogen and oxygen, and blocking the secretion pathways of key competing metabolites. The identified pathway modifications include elimination of competing reactions and also a number of nonintuitive knockouts quite distant from the amino acid-producing pathways. According to the analysis of the results, one can find that OptKnock specifically suggests three reactions deletion (i.e., reactions catalyzed by pyruvate kinase, phosphotransacetylase, and ATPase), and also the elimination of 2-ketoglutarate dehydrogenase, to generate a glutamate-overproducing mutant.

5.1.2.2. OptStrain. A comprehensive database of biotransformations, referred to as the Universal database, with more than 5700 reactions, (http://fenske.che.psu.edu/Faculty/CMaranas/pubs.html.) is compiled and regularly updated by downloading and curating reactions from multiple biopathway database sources. OptStrain (Pharkya et al., 2004) allows knock-ins of non-native functionalities from this comprehensive Universal database of reactions to maximize the yield of desired biochemicals. It can be referred to as hierarchical computational framework that aims at guiding pathway modifications through reaction additions and deletions. Combinatorial optimization is then used to elucidate the set(s) of non-native functionalities, extracted from the Universal database. It removes competing functionalities that divert flux away from the targeted product.

5.1.2.3. OptGene. Determining an optimal gene deletion is a combinatorial problem and thus needs faster algorithms. In this regard, OptGene (Patil et al., 2005), an evolutionary programming based method, has been introduced to rapidly identify gene deletion strategies for optimization of a desired phenotypic objective function. Using a genome-scale model of *S. cerevisiae*, potential metabolic engineering targets for improved production of succinic acid, glycerol and vanillin have been identified.

5.1.2.4. OptReg. Sometimes gene addition/deletion does not describe the entire range of genetic manipulation strategies available. OptReg (Pharkya and Maranas, 2006) is an extended form of OptKnock, which considers knockout as well as up/down regulation of various reactions by using stoichiometric metabolic models. A reaction is upregulated/downregulated if it is constrained to assume flux values significantly above or below its steady-state before the genetic manipulations. It has been demonstrated by studying the overproduction of ethanol in *E. coli*. Computational results reveal that the simultaneous application of both types of genetic manipulations, i.e., reaction deletions and modulations, provide the most desired output. One of the good examples in this regard is the downregulation of phosphoglucomutase in conjunction with the deletion of oxygen uptake and pyruvate formate lyase, thereby yielding 99.8% of the maximum theoretical ethanol. The approach separates the reversible reactions into forward and backward reactions.

5.1.2.5. Genetic Design through Local Search (GDLS). Lun et al. (2009) developed GDLS, based on local search with multiple search paths, results in effective, low-complexity search of the space of genetic manipulations. The authors showed that the MILP-based GDLS predicted complex genetic designs with higher in silico production rates than methods based on evolutionary algorithms. However, it suffers from an exponential increase in complexity with increasing scope of each local search.

5.1.2.6. OptORF. It has been seen that the computational strain designs based on reaction deletions can sometimes result in strategies that are genetically complicated or infeasible due to the presence of multi-functional enzymes, isozymes and regulatory restrictions. Thus a method, called OptORF (Kim and Reed, 2010), has been developed which systematically integrates transcriptional regulatory networks (TRNs) and metabolic networks. It formulates a linear optimization problem that searches for metabolic and/or regulatory perturbations. The method identifies the optimal metabolic and regulatory gene deletions as well as gene overexpression, which maximizes biochemical production at the maximum cellular growth under transcriptional regulatory constraints. Here Boolean logic has been applied to model gene–protein–reaction (GPR) associations. Interactions between the regulatory and metabolic networks have been modeled by turning on/off metabolic gene expression in response to the status of transcriptional factor (TF).

5.1.2.7. OptFORCE. We have seen in the above methods that they do not actively make use of flux measurements for the wild-type and/or an engineered strain to identify the fluxes that need to be actively engineered for achieving the production target. In order to overcome these limitations, OptForce (Ranganathan et al., 2010) has been developed to identify all possible engineering interventions for a wild-type strain characterized by specific metabolic flux data consistent with getting the desired target. It identifies a sufficient and non-redundant set of fluxes, called MUST set, that must change to get overproduced target. This procedure identifies all coordinated reaction modifications that force the pathway to have overproduction of target. The method identifies and refers the reaction fluxes that must increase in case of meeting the requirements to $MUST^U$, whereas ones that must decrease, to $MUST^L$. Next, it identifies the ways to impose the collective set of changes present in the MUST set on the wild-type metabolic network with the minimal number of perturbations, viz., knock-outs/up regulation/down regulation. The collective set of minimal network modifications is referred to as a FORCE set. Most of the reactions in FORCE set are also the members of various MUST sets. OptForce has been demonstrated to predict metabolic interventions for succinate overproduction in *E. coli iAF1260*.

5.1.2.8. Computational Approach for Strain Optimization aiming at high Productivity (CASOP). CASOP (Hadicke and Klamt, 2010) is based on the relative contribution of a reaction derived from weighted elementary modes. It aims at shifting the natural flux distribution to synthesize the desired product with high production rates. Two major determinants of specific productivity, like, product yield and network capacity, have been taken explicitly into account. Thus, we can say that CASOP finds a trade-off between high yield and high network flexibility, aiming at achieving maximum production rates. The method allows consideration of regulatory/operational constraints. The relative contribution of each reaction to yield and network capacity is estimated, and thereby productivity is calculated by analyzing the spectrum of available conversion routes, i.e., elementary modes. Its

application is demonstrated with case studies on overproduction of succinate and lactate in *E. coli*. A very useful application of this approach is the assessment of cofactor/co-metabolite requirements in conjunction with product synthesis, which may help in identifying non-intuitive metabolic limitations.

5.1.2.9. OptFlux. OptFlux (http://www.optflux.org) (Rocha et al., 2010), an open-source user-friendly modular software for metabolic engineering applications, has been introduced, which is compatible with Systems Biology Markup Language (SBML) (http://sbml.org) and Cell Designer (www.celldesigner.org/) (Kitano et al., 2005). The current version OptFlux 2.0 provides several tools and algorithms for manipulation of metabolic models: methods for Metabolic Flux Analysis; methods for phenotypic simulation through FBA, MOMA and ROOM; pathway analysis through calculating elementary mode analysis; strain optimization algorithms, viz., OptKnock, evolutionary algorithms (EAs) and simulated annealing (SA).

5.1.2.10. Enhancing Metabolism with Iterative Linear Optimization (EMILiO). EMILiO (Yang et al., 2011) generates complex strain designs using genome-scale models with unprecedented speed through the use of successive linear programming (SLP) (Baker and Lasdon, 1985). They demonstrated the capability of EMILiO through generating numerous alternate strain designs producing succinate, L-glutamate and L-serine. It identifies a subset of reactions with the potential to improve growth-coupled biochemical production if their fluxes are optimized and predict the optimal flux ranges quantitatively that maximize production.

5.2. Comparison among different methods for metabolic engineering

Well studied biochemical networks are convenient for metabolic engineering applications, otherwise a huge amount of efforts would be required in terms of time and money that would be spent on culturing and cloning of microorganisms. There is certainly an improvement in the functional aspects of the available tools for metabolic engineering. It incorporates additional constraints along with reaction deletion, genetic manipulation and regulation. Here, we discuss comparative aspects of available tools and algorithms for design of efficient microbial strains.

Identification of the relevant pathways of a metabolic network is essential for finding effective metabolic engineering strategies. These pathways can also help in deriving minimal media requirements for an organism and assessing the robustness and redundancy of key metabolic pathways. These criteria are easily taken care through elementary mode and extreme pathway analyses, although, both of them do not consider the optimization of objective function. However, applications of elementary modes to rational metabolic engineering approaches have now enabled the development of strain design algorithms, as used in OptFlux.

Turning the gene expression upward or downward, the development of and consequently enzyme levels and corresponding flux rates is of great importance in metabolic engineering, as a gene deletion can be lethal in many cases whereas its downregulation is not. OptKnock (Burgard et al., 2003), one of the earliest efforts, suggested only removal of the reactions associated with the target gene from the metabolic networks. However, genes and reactions do not always have a one-to-one relationship. Thus, in contrast to OptKnock, OptReg (Pharkya and Maranas, 2006), an extended version of it, considers overexpression and down regulation of various reactions, along with deletion by using stoichiometric metabolic models. Actually, it uses the OptKnock formulation as a starting point. However, both of them rely on biomass maximization to perform flux allocation in a metabolic network. OptKnock has certain limitations. For example, it gives certain reactions as its output for deletion that do not have any gene association. This hampers the reconstruction of strains experimentally. OptStrain (Pharkya et al., 2004) also extends OptKnock as it gives minimal reaction set recombination tasks to confer a desired non-native biochemical production capability on a microbial host. OptGene (Patil et al., 2005) extends the applicability of OptKnock approach using a Genetic Algorithm. It demands relatively less computational time and thus it can be applied to larger networks. The important aspect of OptGene formulation is the optimization of non-linear objective functions, incorporating non-linear constraints through which it provides a set of solutions closer to optimal ones.

Regulatory constraints, presence of multi-functional enzymes and isozymes, infeasible computational designs based on reaction deletions sometimes make the strategies genetically complicated. OptORF (Kim and Reed, 2010) overcomes these limitations by considering both gene deletion and transcription regulation simultaneously, which have not been applied to above mentioned approaches. This is an effective method to systematically integrate transcriptional regulatory and metabolic networks that searches for metabolic and/or regulatory perturbations that couple biomass and biochemical productions. OptKnock requires more gene deletions unlike those using OptORF, as OptORF can get the desired high production yield after making only the smallest number of gene deletions. Another important aspect should be noted is that only reaction deletions are considered in OptKnock, which sometimes results in reduced yields or lethal phenotypes. On the other hand, OptORF (Ranganathan et al., 2010) proposes the overexpression of metabolic genes or deletion of transcription factors that may lead to faster evolutionary trajectories.

OptKnock and other methods do not use kinetic descriptions, but rely on the maximization of surrogate biological fitness functions, viz., maximization of biomass yield or MOMA (Segre et al., 2002) to estimate flux redirection upon strain engineering. These estimates may predict inaccurate representation of metabolic response to genetic or environmental perturbations. In contrast to these existing methods, OptForce (Ranganathan et al., 2010), a more recent bilevel approach, identifies all possible engineering interventions for a wild-type strain characterized by specific metabolic flux data. The authors have compared the yields predicted for succinate overproduction through applying OptKnock, OptReg and OptForce, and found OptForce to be superior to OptKnock and OptReg. Both OptKnock and OptReg rely on biomass maximization to perform flux allocation in a metabolic network, whereas OptForce reports the most conservative value for succinate production allowed by the stoichiometry and conditions. The major difference is that OptFORCE uses optimization to identify how metabolic fluxes must change to improve metabolite production, and is independent of any assumptions about what functions are used to predict cellular behavior. Moreover, it is also found that bi-level optimization techniques are limited in applicability due to the required computational time and resources scale poorly with the increase in the size of the metabolic system and in the increase in the number of genetic manipulations. In this case, GDLS is developed to genetic designs with greater in silico production of desired metabolites and shown to perform favorably in comparison with heuristic searches based on evolutionary algorithms and simulated annealing. However, Yang et al. (2011) applied GDLS to OptReg problem and found that it still suffers from an exponential increase in complexity with increasing scope of each local search. Here, authors found smaller local search scopes proved insufficient for escaping local optima.

6. Conclusions

The structure and dynamics of cellular function of an organism is examined to understand biology at system level. Analysis of biochemical pathways is one of the key topics in the post-genomic era. A goal of Systems Biology is to formulate initial working models for biological networks that are predictive of both the dynamic and equilibrium behavior of the system in question. In order to understand the cellular mechanisms, to automatically retrieve metabolic information from

the predicted metabolic pathways, we have to develop and implement useful methodologies, algorithms and tools. Well described mathematical pathway models can be served as a virtual laboratory which signifies the system and give insights into basic principles of cellular functions, like, robustness and adaptability of a cell in different environments. Advances in network topology theory and visualization tools might enable biologists to assemble data into network models that better present the kinetics of molecular interactions.

Certain metabolic pathway analysis approaches focus on only the final product-synthesis reaction, however, metabolic engineering emphasizes the metabolic pathway as a whole. Metabolic engineering applies an integrated, systems-level approach for optimizing a desired cellular phenotype. The integration of cellular metabolism and regulation often gives a complex picture. The availability of complete genome sequences for several microorganisms has provided an opportunity to develop genomic scale metabolic models. The key question is how to extract all the relevant information present in experimental genome-scale omics databases for designing efficient industrial processes. We have included a comparative analysis of most of the methodologies for metabolic networks in this article, along with an emphasis on metabolic engineering. We have also included a brief discussion on GSMM, its limitations and the necessity of automation of modeling process. There should be enough development of the techniques in the future, which can search and extract the information and biochemical reactions that can be incorporated into standard systems. Moreover, work on automation of modeling process should be enhanced as it will increase the practicality of metabolic engineering technique (Copeland et al., 2012).

Acknowledgments

Ms. Namrata Tomar, one of the authors, gratefully acknowledges CSIR, India for providing her a Senior Research Fellowship (9/93 (0145)/12, EMR-I).

References

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2007. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab. Eng. 9, 68–86.
Aittokallio, T., Schwikowski, B., 2006. Graph-based methods for analysing networks in cell biology. Briefings in Bioinformatics 7, 243–255.
Albert, R., 2007. Network inference, analysis, and modeling in systems biology. Plant Cell 19, 3327–3338.
Baker, T., Lasdon, L., 1985. Successive linear programming at Exxon. Manag. Sci. 31, 264–274.
Baldan, P., Cocco, N., Marin, A., Simeoni, M., 2010. Petri nets for modelling metabolic pathways: a survey. Nat. Comput. 9, 955–989.
Beard, D., Babson, E., Curtis, E., Qian, H., 2004. Thermodynamic constraints for biochemical networks. J. Theor. Biol. 228, 327–333.
Beasley, J.E., Planes, F.J., 2007. Recovering metabolic pathways via optimization. Bioinformatics 23, 92–98.
Becker, S., Palsson, B., 2005. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. BMC Microbiol. (5).
Becker, S., Price, N., Palsson, B., 2006. Metabolite coupling in genome-scale metabolic networks. BMC Bioinformatics (7).
Bernal, V., Carinhas, N., Yokomizo, A., Carrondo, M., Alves, P., 2009. Cell density effect in the baculovirus-insect cells system: a quantitative analysis of energetic metabolism. Biotechnol. Bioeng. 104, 162–180.
Blazeck, J., Alper, H., 2010. Systems metabolic engineering: genome-scale models and beyond. Biotechnol. J. (5).
Boles, E., de Jong-Gubbels, P., Pronk, J.T., 1998. Identification and characterization of MApE1, the *Saccharomyces cerevisiae* structural gene encoding mitochondrial malic enzyme. J. Bacteriol. 180, 2875–2882.
Bower, J.M., Bolouri, H., 2001. Computational Modeling of Genetic and Biochemical Networks. Massachusetts Institute of Technology.
Burgard, A., Pharkya, P., Maranas, C., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657.
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.
Carinhas, N., Bernal, V., Teixeira, A.P., Carrondo, M.J., Alves, P.M., Oliveira, R., 2011. Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products. BMC Syst. Biol. (5).
Carlson, R., Fell, D., Srienc, F., 2002. Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol. Bioeng. 79, 121–134.
Caspi, R., et al., 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/Genome Databases. Nucleic Acids Res. 38, D473–D479.
Chaouiya, C., 2007. Petri net modelling of biological networks. Brief. Bioinform. 8, 210–219.
Chen, M., Hofestadt, R., 2003. Quantitative Petri Net model of gene regulated metabolic networks in the cell. In Silico Biol. 3, 347–365.
Clarke, B.L., 1981. Complete set of steady states for the general stoichiometric dynamical system. J. Chem. Phys. 75, 4970–4979.
Clarke, B.L., 1988. Stoichiometric network analysis. Cell Biophys. 237, 237–253.
Copeland, W.B., et al., 2012. Computational tools for metabolic engineering. Metab. Eng 14, 270–280.
Covert, M., Schilling, C., Palsson, B., 2001. Regulation of gene expression in flux balance models of metabolism. J. Theor. Biol. 213, 73–88.
David, L., Marashi, S.-A., Larhlimi, A., Mieth, B., Bockmayr, A., 2011. FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. BMC Bioinformatics (12).
Degtyarenko, K., et al., 2008. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 36, D344–D350.
Dien, S.V., Lidstrom, M., 2002. 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.
Doerr, A., 2010. A global map of gene function. Nat. Methods 7, 176.
Edwards, J., Palsson, B., 2000. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli K-12* gene deletions. BMC Bioinformatics (1).
Edwards, J., Ibarra, Palsson, B., 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19, 125–130.
Feinberg, M., Horn, F.J.M., 1974. Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. Chem. Eng. Sci. 29, 775–787.
Feist, A., Mo, M., Hannum, G., Palsson, B., Herrgard, M., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat. Protoc. 2, 727–738.
Fleischmann, A., et al., 2004. IntEnz, the integrated relational enzyme database. Nucleic Acids Res. 32, D434–D437.
Gerdes, S., Edwards, R., Kubal, M., Fonstein, M., Stevens, R., Osterman, A., 2006. Essential genes on metabolic maps. Curr. Opin. Biotechnol. 17, 448–456.
Gudmundsson, S., Thiele, I., 2010. Computationally efficient flux variability analysis. BMC Bioinformatics (11).
Hadicke, O., Klamt, S., 2010. CASOP: a computational approach for strain optimization aiming at high productivity. J. Biotechnol. 147, 88–101.
Hartmann Genrich, R.K., Voss, K., 2001. Executable Petri net models for the analysis of metabolic pathways. Int. J Software Tools Technol. Transfer 3, 394–404.
Henry, C., Broadbelt, L., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. Biophys. J. 92, 1792–1805.
Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977–982.
Herrgard, M., Lee, B., Portnoy, V., Palsson, B., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. Genome Res. 16, 627–635.
Hillyard, D., Redd, M., 2007. Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engineering. Elsevier, San Diego.
Hoops, S., et al., 2006. COPASI — a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074.
Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C., Holzhütter, H.-G., 2011. FASIMU: flexible software for flux-balance computation series in large metabolic networks. BMC Bioinformatics (12).
Imielinski, M., Belta, C., Halasz, A., Rubin, H., 2005. Investigating metabolite essentiality through genome-scale analysis of *Escherichia coli* production capabilities. Bioinformatics 21, 2008–2016.
Imielinski, M., Belta, C., Rubin, H., Halasz, A., 2006. Systematic analysis of conservation relations in *Escherichia coli* genome-scale metabolic network reveals novel growth media. Biophys. J. 90, 2659–2672.
Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabasi, A.L., 2000. The large scale organization of metabolic networks. Nature 407, 651–654.
Joshi-Tope, G., et al., 2005. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 33, D428–D432.
Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D109–D114.
Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 39, D583–D590.
Kharchenko, P., Vitkup, D., Church, G., 2004. Filling gaps in a metabolic network using expression information. Bioinformatics 20, i178–i185.
Kharchenko, P., Chen, L., Freund, Y., Vitkup, D., Church, G.M., 2006. Identifying metabolic enzymes with multiple types of association evidence. BMC Bioinformatics (7).
Kim, J., Reed, J.L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Sys. Biol. (4).
Kitano, H., Funahashi, A., Matsuoka, Y., Oda, K., 2005. Using process diagrams for the graphical representation of biological networks. Nat. Biotechnol. 23, 961–966.
Klamt, S., Stelling, J., 2002. Combinatorial complexity of pathway analysis in metabolic networks. Mol. Biol. Rep. 29, 233–236.
Klamt, S., Stelling, J., Ginkel, M., Gilles, E.D., 2003. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. Bioinformatics 22, 26–269.

Klamt, S., Haus, U.-U., Theis, F., 2009. Hypergraphs and cellular networks. PloS Comput. Biol. 5, 1–6.
Klein-Marcuschamer, D., Ajikumar, P., Stephanopoulos, G., 2007. Engineering microb cell fact for biosynthesis of isoprenoid molecules: beyond lycopene. Trends Biotechnol. 25, 417–424.
Koch, I., Junker, B., Heiner, M., 2005. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. Bioinformatics 21, 1219–1226.
Koffas, M., Roberge, C., Lee, K., Stephanopoulos, G., 1999. Metabolic engineering. Annu. Rev. Biomed. Eng. 1, 535–557.
Larhlimi, A., Bockmayr, A., 2009. A new constraint-based description of the steady-state flux cone of metabolic networks. Discret. Appl. Math. 157, 2257–2266.
Lincoln, P., Tiwari, A., 2004. Symbolic systems biology: hybrid modeling and analysis of biological networks. In: Alur, R., Pappas, G. (Eds.), Proceedings of Hybrid Systems: Computation and Control. : LNCS, Vol. 2993. Springer, pp. 660–672.
Llaner, F., Pico, J., 2008. Stoichiometric modelling of cell metabolism. J. Biosci. Bioeng. 105, 1–11.
Lun, D.S., et al., 2009. Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5, 1–8.
Kumar, S.V., Dasika, M.S., Maranas, C.D., 2007. Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics (8).
Mahadevan, R., Schilling, C., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276.
Martin, V., Pitera, D., Withers, S., Newman, J., Keasling, J., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. 21, 796–802.
Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., Miyano, S., 2003. Biopathways representation and simulation on hybrid functional Petri Net. In Silico Biol. 3, 389–404.
McDonald, A.G., Boyce, S., Tipton, K.F., 2009. ExplorEnz: the primary source of the IUBMB enzyme list. Nucleic Acids Res. 37, D593–D597.
Mendes, P., 1993. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. Comput. Appl. Biosci. 9, 563–571.
Milner, P., 1964. The possible mechanisms of complex reactions involving consecutive steps. J. Electrochem. Soc. 111, 228–232.
Nalecz, M.J., Nalecz, K.A., Azzi, A., 1991. Purification and functional characterisation of the pyruvate (monocarboxylate) carrier from baker's yeast mitochondria (*Saccharomyces cerevisiae*). Biochim. Biophy. Acta (BBA) — Protein Struct. Mol. Enzymol. 1079, 87–95.
Osterman, A., Overbeek, R., 2003. Missing genes in metabolic pathways: a comparative genomics approach. Curr. Opin. Chem. Biol. 7, 238–251.
Papin, J., Price, N., Edwards, J., Palsson, B., 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J. Theor. Biol. 215, 67–82.
Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., Palsson, B.O., 2002. Metabolic pathways in the post-genome era. Trends Biochem. Sci. (28).
Papoutsakis, E., 1984. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol. Bioeng. 26, 174–187.
Patil, K.R., Akesson, M., Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. Curr. Opin. Biotechnol. 15, 64–69.
Patil, K.R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinforma. 6.
Perrenoud, A., Sauer, U., 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. J. Bacteriol. 187, 3171–3179.
Petri, C.A. Kommunikation mit Automaten, Ph.D. thesis, Institut fur Instrumentelle Mathematik, Bonn (1962).
Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15, 251–257.
Pharkya, P., Maranas, C.D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8, 1–13.
Pharkya, P., Burgard, A.P., Maranas, C.D., 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol. Bioeng. 84, 887–899.
Pharkya, P., Burgard, A., Maranas, C., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.
Price, N., Papin, J., Palsson, B., 2002. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res. 12, 760–769.
Provost, A., Bastin, G., 2006. Metabolic Flux Analysis: an approach for solving non-stationary undetermined systems. Proceedings 5th MATHMOD, Vienna, pp. 48–55.
Raghunathan, A., Reed, J., Shin, S., Palsson, B., Daefler, S., 2009. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Syst. Biol. (3).
Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6, e1000744.
Rocha, I., et al., 2010. OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst. Biol. 4.
Romero, P., Wagg, J., Green, M.L., Kaiser, D., Krummenacker, M., Karp, P.D., 2004. Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 6, R2.
Sauer, U., Hatzimanikatis, V., Bailey, J., Hochuli, M., Szyperski, T., Wuthrich, K., 1997. Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. Nat. Biotechnol. 15, 448–452.
Scheer, M., et al., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676.
Schilling, C., Letscher, D., Palsson, B., 2000. 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.
Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., Palsson, B.O., 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. J. Bacteriol. 184, 4582–4593.
Schuster, S., Hilgetag, S., 1994. On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst. 2, 165–182.
Schuster, S., Dandekar, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol. 17, 53–60.
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.
Schustera, S., Dandekara, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechonol. 17, 53–60.
Schwartz, J., Kanehisa, M., 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics (7).
Segre, D., Vitkup, D., Church, G., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99, 15112–15117.
Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. 102, 7695–7700.
Simao, E., Remy, E., Thieffry, D., Chaouiya, C., 2005. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. coli*. Bioinformatics 21, ii190–ii196.
Smallbone, K., Simeonidis, E., Broomhead, D., Kell, D., 2007. Something from nothing — bridging the gap between constraint-based and kinetic modelling. FEBS J. 274, 5576–5585.
Steuer, R., Junker, B., 2009. Computational Models of Metabolism: Stability and Regulation in Metabolic Networks, vol. 142. John Wiley and Sons.
Steuer, R., Gross, T., Selbig, J., Blasius, B., 2006. Structural kinetic modeling of metabolic networks. Proc. Natl. Acad. Sci. 103, 11868–11873.
Svante, M., Sjostroma, L., 2001. Eriksson, PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.
Terzer, M., Maynard, N.D., Covert, M.W., Stelling, J., 2009. Genome-scale metabolic networks. Wiley Interdiscip. Rev. Syst. Biol. Med. 1, 285–297.
Thiele, I., Palsson, B., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.
Thomas, P.D., et al., 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 31, 334–341.
Trinh, C., Carlson, R., Wlaschin, A., Srienc, F., 2006. Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab. Eng. 8, 628–638.
Trinh, C., Unrean, P., Srienc, F., 2008. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl. Environ. Microbiol. 74, 3634–3643.
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.
UniProt Consortium, T., 2011. Ongoing and future developments at the universal protein resource. Nucleic Acids Res. 39, D214–D219.
Varma, A., Palsson, B., 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Biotechnology 12, 994–998.
Vemuri, G.N., Aristidou, A.A., 2005. Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. Microbiol. Mol. Biol. Rev. 69, 197–216.
von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22, 1930–1931.
Wagner, A., Fell, D.A., 2001. The small world inside large metabolic networks. Proc. R. Soc. Lond. B 268, 1803–1810.
Watson, M.R., 1984. Metabolic maps for the Apple II. Biochem. Soc. Trans. 12, 1093–1094.
Watson, M.R., 1986. A discrete model of bacterial metabolism. Comput. Appl. Biosci. 2, 23–27.
Whitaker, J.W., Letunic, I., McConkey, G.A., Westhead, D.R., 2009. metaTIGER: a metabolic evolution resource. Nucleic Acids Res. 37, D531–D538.
Wiback, S.J., Palsson, B.O., 2002. Extreme pathway analysis of human red blood cell metabolism. Biophys. J. 83, 808–818.
Wiechert, W., 2001. ^{13}C metabolic flux analysis. Metab. Eng. 3, 195–206.
Wong, P., Gladney, S., Keasling, J., 1997. Mathematical model of the lac Operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotechnol. Prog. 13, 132–143.
Yang, C.-R., Shapiro, B.E., Mjolsness, E.D., Hatfield, G.W., 2005. An enzyme mechanism language for the mathematical modeling of metabolic pathways. Bioinformatics 21, 774–780.
Yang, L., Cluett, W.R., Mahadevan, R., 2011. EMILiO: A fast algorithm for genome-scale strain design. Metabol. Eng. 13, 272–281.