NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
Appl Microbiol Biotechnol. Author manuscript; available in PMC 2010 July 23.

Published in final edited form as:
Appl Microbiol Biotechnol. 2009 January ; 81(5): 813–826. doi:10.1007/s00253-008-1770-1.

Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism

Cong T. Trinh[1,2], **Aaron Wlaschin**[1,2,#], and **Friedrich Srienc**[1,2,*]

[1] Department of Chemical Engineering and Materials Science, University of Minnesota, 240 Gortner Laboratory, 1479 Gortner Ave., St. Paul, MN 55108, USA

[2] BioTechnology Institute, University of Minnesota, 240 Gortner Laboratory, 1479 Gortner Ave., St. Paul, MN 55108, USA

Abstract

Elementary Mode Analysis is a useful Metabolic Pathway Analysis tool to identify the structure of a metabolic network that links the cellular phenotype to the corresponding genotype. The analysis can decompose the intricate metabolic network comprised of highly interconnected reactions into uniquely organized pathways. These pathways consisting of a minimal set of enzymes that can support steady state operation of cellular metabolism represent independent cellular physiological states. Such pathway definition provides a rigorous basis to systematically characterize cellular phenotypes, metabolic network regulation, robustness, and fragility that facilitate understanding of cell physiology and implementation of metabolic engineering strategies. This mini-review aims to overview the development and application of elementary mode analysis as a metabolic pathway analysis tool in studying cell physiology and as a basis of metabolic engineering.

Keywords

Metabolic Pathway Analysis; Metabolic Engineering; Elementary Mode; Extreme Pathway; Weighting Factors; Rational Strain Design; Genetic Knockout Analysis; Metabolic Flux Ratio; Minimal Cut Set; Control Effective Flux; Robustness; Minimal Cell

Introduction

Recent advances in sequencing technologies have provided gene sequences of many different organisms (Benson et al. 2008). Bioinformatics tools have facilitated the reconstruction of cellular metabolism of these organisms based on information encoded in their genomes (Karp et al. 2007, Caspi et al. 2006, Kanehisa et al. 2008, Duarte et al. 2007). To understand the phenotypic capabilities of these organisms, it is useful to characterize cellular metabolism through quantitative analysis of pathway operations.

Computational tools to analyze cellular metabolism have been developed for more than two decades. Such analysis involves the determination of metabolic fluxes defined by the rates of enzyme-catalyzed reactions participating in a metabolic network. A metabolic flux vector, also known as metabolic flux distribution, defines cellular phenotype under a given growth condition. Depending on the goals of the analysis, these tools can be grouped into three categories including (i) Metabolic Flux Analysis (Stephanopoulos et al. 1998, Wiechert

*Corresponding author: Friedrich Srienc. srienc@umn.edu. Tel: 612 624 9776. Fax: 612 625 1700.
#Current address: General Mills, Inc., Minnesota, USA.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2001), (ii) Flux Balance Analysis (Kauffman et al. 2003, Price et al. 2004, Edwards et al. 1999) and (iii) Metabolic Pathway Analysis (Klamt and Stelling 2003, Schuster and Hilgetag 1994, Schilling et al. 2000). Even though differences exist in the problem formulation, these tools are related. They are developed from the same mathematical principle discussed in detail later in the text. In this mini-review, we focus primarily on elementary mode analysis which is one of the metabolic pathway analysis tools.

In general, metabolic pathway analysis identifies the topology of cellular metabolism based on only the stoichiometric structure and thermodynamic constraints of reactions where kinetic parameters are not explicitly revealed and/or required for the calculations (Reder 1988, Schuster et al. 2002, Clarke 1988). This type of analysis has been successfully applied to various organisms to investigate metabolic network structure, robustness, fragility, regulation, metabolic flux vector, and rational strain design (Carlson and Srienc 2004, Carlson and Srienc 2004, Stelling et al. 2002, Poolman et al. 2004, Price et al. 2003, Schuster et al. 2000, Wiback et al. 2004, Trinh et al. 2006, Wlaschin et al. 2006, Klamt and Gilles 2004, Trinh et al. 2008).

This review aims to provide an overview of the history and theory behind metabolic pathway analysis and in particular elementary mode analysis. After a brief survey of progress and advances in software and algorithm development for elementary mode analysis, the interpretations and applications of elementary mode analysis relevant to genomics, cellular physiology, and metabolic engineering with an emphasis on rational strain design is highlighted.

History of Metabolic Pathway Analysis

Analysis of structural invariants of a (bio)chemical network has been first applied to systematically postulate the reaction mechanisms for a system of chemical reactions (Milner 1964). This approach was further generalized for investigating the steady states and stabilities of the general relationships of dynamical systems that are interconnected through the reaction stoichiometry (Feinberg and Horn 1974, Clarke 1981). The Stoichiometric Network Analysis (SNA) based on convex analysis was the pioneering work in this field to identify “extreme currents” or unique pathways for a system of chemical reactions (Clarke 1988). Since the introduction of this concept, different related approaches with modifications in problem formulation and with improvements of algorithm implementation have been proposed and developed for biological systems over the past two decades. One of the first modified approaches is Elementary Mode Analysis introduced by Schuster in 1994 (Schuster and Hilgetag 1994). Different from the original approach, elementary mode analysis does not decompose the reversible reactions into two irreversible reactions in calculating elementary modes (EMs) and introduces a systematic way of extracting biologically meaningful pathways from an intricate metabolic network. An alternative approach is Extreme Pathway Analysis (Schilling et al. 2000). This analysis can be considered as a hybrid between stoichiometric network analysis and elementary mode analysis. In calculating extreme pathways (ExPas), the analysis splits only the internal reversible reactions into two irreversible reactions while not decomposing reversible exchange reactions. A detailed discussion of differences between these two techniques will be presented in the following theory section. In addition to convex analysis, other approaches have also been proposed for metabolic pathway analysis to analyze and synthesize metabolic pathways. For instance, database searching algorithms were developed to identify pathways to link carbon-carrying metabolites (Seressiotis and Bailey 1988, Mavrovouniotis et al. 1990). A Petri net theory was also applied to compute possible pathways that account for the inter-conversion of metabolites in a metabolic network (Koch

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

et al. 2005, Mavrovouniotis et al. 1996). A comparison of algorithm differences used in these techniques has been summarized by Schuster et al. (Schuster et al. 2002).

Theory of analyzing a metabolic network

The theory applied to analyze metabolic networks is developed based on the first principle of mass conservation of internal metabolites within a system (Reder 1988, Clarke 1981, Schuster and Schuster 1993). A biological system consists usually of a single cell or a cell compartment that contains metabolites. These metabolites can be transformed to others through an intricate metabolic network of enzyme-catalyzed reactions (Roels 1983). For classification purposes, reactions that transform metabolites within the system can be considered internal reactions while reactions involving the transport of metabolites in and out of the system can be considered to be exchange reactions (Schuster and Hilgetag 1994, Schilling et al. 2000). The general equation to describe the mass conservation of metabolites in a system of defined volume can be written as

$$\frac{d}{dt}\underline{C}=\underline{\underline{S}}\cdot\underline{r}-\mu\cdot\underline{C}, \tag{1}$$

where $\underline{C}$ (mol/L) is the concentration vector of m internal metabolites; $\underline{r}$(mol/L/hr) is the reaction rate (flux) vector of n reactions that convert metabolites; $\underline{\underline{S}}$ is the stoichiometry matrix of dimension $m_x n$ whose elements s_{ij} represents the stoichiometry coefficient of the element i involved in reaction j; μ (1/hr) is the specific dilution rate associated with the change in volume of the system. For a biological system such as a single cell, the dilution rate is much slower than the reaction rates that transform metabolites. Therefore, the contribution of volume change to the concentration changes of metabolites within the system is considered to be negligible. At steady state, there is no accumulation of internal metabolites in the system and Equation (1) can be simplified to

$$\underline{\underline{S}}\cdot\underline{r}=0. \tag{2}$$

Due to thermodynamic constraints, reactions have to proceed in the appropriate direction. Some reactions are irreversible and require additional constraints on positive flux values, that is,

$$r_i \geq 0. \tag{3}$$

Figure 1A demonstrates an example of how to formulate the problem for a simple metabolic network. The network consists of 9 reactions, two of which are reversible, and of 9 metabolites, five of which are internal.

For a cellular metabolism, Equation (2) is typically an underdetermined system where the number of metabolites is far fewer than the number of reactions. The number of metabolites defines the number of balance equations in (2) while the number of reactions represents the number of unknowns in (2). Depending on the invariant structure of the stoichiometric matrix $\underline{\underline{S}}$ and the knowledge of some experimentally measured fluxes (reactions), three main techniques have been proposed to solve the system of linear equations (2) together with inequality constraints (3) for metabolic flux vector $\underline{r}$. These techniques include Metabolic

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Flux Analysis, Flux Balance Analysis, and Metabolic Pathway Analysis. Their differences are briefly discussed next and demonstrated in Figure 1.

Metabolic Flux Analysis

Traditional Metabolic Flux Analysis calculates a metabolic flux vector $\underline{r}$ by solving equation (2) as follows:

$$\underline{\underline{S}}_u \cdot \underline{r}_u = -\underline{\underline{S}}_m \cdot \underline{r}_m. \tag{4}$$

Equation (4) is derived from Equation (2) where the flux vector is partitioned into the unmeasurable flux vector $\underline{r}_u$ and the measurable flux vector $\underline{r}_m$. Typically, the stoichiometric network is simplified or more fluxes are experimentally determined so that the partitioned matrix $\underline{\underline{S}}_u$ is invertible so that $\underline{r}_u = -\underline{\underline{S}}_u^{-1} \cdot \underline{\underline{S}}_m \cdot \underline{r}_m$ can be solved (Stephanopoulos et al. 1998). This is demonstrated in Figure 1B. Advanced metabolic flux analysis based on ^{13}C labeling experiments solve Equation (4) iteratively in a more complex computational scheme (Wiechert 2001) (also see references therein). In general, metabolic flux analysis relies on extensive experimental data to increase the number of measurable fluxes such that the unmeasurable flux vector can be calculated. It should be noted that metabolic flux analysis computes only a metabolic flux vector $\underline{r}$ for a particular growth condition. A change in measured fluxes $\underline{r}_m$ in different growth conditions will result in a different metabolic flux vector.

Flux Balance Analysis

Flux Balance Analysis is also a tool to determine a metabolic flux vector $\underline{r}$ of a cellular physiological state when knowledge of $\underline{r}_m$ is limited and $\underline{\underline{S}}_u$ can not be inverted to provide a unique solution. The approach is based on convex analysis by imposing an objective function to determine the metabolic flux vector (Kauffman et al. 2003, Price et al. 2004) (also see references herein) subject to several constraints such as substrate uptake rates, and/or product secretion rates, thermodynamic constraints, metabolic regulation and so on. For instance, Figure 1C demonstrates the determination of the metabolic flux vector of a simple network for the case of maximizing the product P from the sole supply of substrate A. The key of this approach is to figure out what objective functions likely represent the cellular metabolism under a given growth condition (Schuetz et al. 2007). Flux balance analysis has been successfully applied to predict specific growth rates by imposing the constraint that cells function by maximizing their specific growth rates for given substrate uptake rates (Edwards et al. 2001). Other frameworks that are also based on optimization strategies such as MOMA have been proposed to predict metabolic flux vectors of gene knockout mutants by imposing the constraint that mutants operate by minimizing their metabolic adjustment with respect to the wildtype (Segre et al. 2002). It should be mentioned that flux balance analysis identifies only one optimal solution while alternative optimal solutions or suboptimal solutions can exist. In general, flux balance analysis can calculate metabolic flux vectors based on limited experimental data and requires specification of objective functions for cellular metabolism. The more fluxes can be measured, the more accurately the flux vector can be computationally determined. However, the metabolic flux vector may not be unique. The approach depends very much on the validity that the formulated objective function indeed correctly represents the working system. Similar to metabolic flux analysis, flux balance analysis identifies a single metabolic flux vector under a given growth condition.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Metabolic Pathway Analysis

In contrast to the above techniques, metabolic pathway analysis can identify all metabolic flux vectors that exist in a metabolic network without requiring knowledge of any fixed flux rates or imposing any objective function for cellular metabolism. The solutions of Equation (2) together with the inequality (3) constitute the admissible flux space which is also known as the convex polyhedral cone (Rockafellar 1970). The inequality (3) represents the thermodynamic feasibility constraint or sign restriction constraint (Klamt and Stelling 2003, Schuster and Hilgetag 1994, Schuster et al. 2002, Gagneur and Klamt 2004). The number of these solutions is infinite. However, additional constraints on the admissible flux space such as non-decomposability and systematic independence can form a finite set of solutions. The application of these additional constraints results in different, but closely related techniques for metabolic pathway analysis including elementary mode analysis (Schuster and Hilgetag 1994) and extreme pathway analysis (Schilling et al. 2000).

Elementary mode analysis calculates all solutions in the admissible flux space by solving Equation (2) in conjunction with the thermodynamic constraint (3) and an additional non-decomposability constraint. Each solution presents an elementary (flux) mode. The non-decomposability constraint ensures that each elementary mode is unique up to a positive scalar factor because removal of any reaction in an elementary mode will automatically disrupt the entire pathway. Therefore, each elementary mode can be defined as a unique, minimal set of enzymes (participating reactions) to support steady state operation of a metabolic network with irreversible reactions to proceed in appropriate directions (Schuster and Hilgetag 1994, Schuster et al. 2002, Pfeiffer et al. 1999).

Different from elementary mode analysis, extreme pathway analysis contains one additional constraint to make all extreme pathways systematically independent (Klamt and Stelling 2003, Schilling et al. 2000, Papin et al. 2003, Papin et al. 2004). Systematic independence implies that none of the extreme pathways can be expressed as a nonnegative combination of at least two other extreme pathways. Even though some elementary modes are not systematically independent, they are genetically independent due to the direct implementation of the non-decomposability constraint. In addition, extreme pathways are a subset of elementary modes. The two sets of extreme pathways and elementary modes are identical when all reactions including both internal and exchange reactions are irreversible in a metabolic network. Therefore, the identification of extreme pathways depends on the reconfiguration of the metabolic network analyzed while the identification of elementary modes does not. For instance, extreme pathways that are identified in a metabolic network with each reversible exchange reaction split into two irreversible reactions may not be extreme pathways anymore in the original metabolic network with reversible exchange reactions not split (Klamt and Stelling 2003).

Due to the close relatedness in computing both elementary modes and extreme pathways, Boley et al. have recently developed a simple rank/nullity test to distinguish extreme pathways from elementary modes only by using the invariant stoichiometric matrix (Boley et al. 2008). Furthermore, the test can also determine whether a pathway is an extreme pathway or an elementary mode even before computing both sets of elementary modes or extreme pathways separately. This rank/nullity test is a very useful tool for researchers to assess the complete set of elementary modes and to be able to extract from it the subset of extreme pathways (Figure 1D). Knowledge of types of pathways allows one to choose the appropriate metabolic pathway analysis tool for specific applications.

Differences between elementary mode analysis and extreme pathway analysis are presented in the example of a simple network shown in Figure 1. Figure 1D, 1F shows the complete set of 8 elementary modes (EMs), 4 of which are also extreme pathways (ExPas). While

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

metabolic flux analysis and flux balance analysis only calculate, in general, a single metabolic flux vector under a given growth condition (Figure 1C, 1E), metabolic pathway analysis such as elementary mode analysis can identify all metabolic flux vectors that exist in a cellular metabolism without requiring any knowledge of some measured fluxes (Figure 1D, 1E, 1F). This characteristic allows a systematic and objective evaluation of metabolism capabilities in terms of cellular robustness, fragility and regulation.

Algorithm and software development for elementary mode analysis

Since the concept of elementary mode analysis was introduced in 1994 (Schuster and Hilgetag 1994), there has been an ongoing effort to develop more efficient algorithms during the last two decades. One of the improvements of the algorithm involves the reduction of the size of the stoichiometric matrix used for calculating elementary modes. Examination of the null space of the stoichiometric matrix can detect reactions that produce or consume unbalanced metabolites having neither sinks nor sources. These reactions can be automatically removed from the network for further calculation since fluxes though these reactions are always null (Reder 1988, Schuster and Schuster 1991). The null space also allows identification of reactions that operate in a linear pathway without branches with a fixed flux ratio. These reactions can be lumped into one reaction and hence reduce the size of the stoichiometric matrix used for calculating elementary modes. Some earlier versions of the publicly available software developed for calculating elementary modes are METATOOL (Pfeiffer et al. 1999), GEPASI and its successor COPASI (Hoops et al. 2006) and FluxAnalyzer (Klamt et al. 2003). A similar software with a closely related algorithm has been developed for calculating extreme pathways (ExPas) (Bell and Palsson 2005). In all of these software, the core of computing elementary modes is written in the C language. However, the FluxAnalyzer is also developed with a user friendly interface based on the MATLAB environment (The Mathworks, Inc., USA) and contains additional features to analyze the metabolic network (Klamt et al. 2003). Recent improvements in algorithm implementation have further advanced software development to compute elementary modes for larger metabolic networks. The original approach is to solve the equality (2) and the inequality (3) simultaneously while the latter approach known as Null Space approach first solves the equality (2) and then satisfies the inequality (3) while still in the null space (Wagner 2004, Urbanczik and Wagner 2005b). The recently developed software SNA uses the null space approach and is written in MATHEMATICA (Wolfram Research, Inc., USA) (Urbanczik 2006). Recent METATOOL software has also incorporated this algorithm together with an efficient rank test to check the elementarity of a mode (von Kamp and Schuster 2006). The computation of elementary modes by the latest METATOOL has codes written in either C language or MATLAB. YANA is also a recently developed software based on METATOOL that has a user-friendly interface with additional built-in tools for metabolic network analysis (Schwarz et al. 2005). A recent version, YANAsquare, further introduces the software capability to automatically import reconstructed metabolic networks of different microorganisms from the KEGG database (Schwarz et al. 2007). Gagneur and Klamt further suggest that the implementation of the binary approach during computation of elementary modes can decrease the memory demand up to 96% (Gagneur and Klamt 2004). This approach was included in FluxAnalyzer 5.1, and later the software name was changed to CellAnalyzer after including the analysis of signal transduction pathways (Klamt et al. 2007).

Computing all elementary modes is expensive, especially for a large metabolic network (Klamt and Stelling 2002). However, the extremely useful feature of knowing all pathway possibilities will continue to stimulate the development of more efficient algorithms, computational techniques, and problem formulations that will enable also handling of networks at a genome scale level. Encouraging results have been obtained by parallel

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

computing (Klamt et al. 2005, Samatova 2002), network decomposition (Schuster et al. 2002, Schilling and Palsson 2000, Schwartz et al. 2007), examination of only a functional conversion flux cone encapsulated in the parent flux cone (Urbanczik and Wagner 2005a, Song and Ramkrishna 2008), and advanced algorithm for enumerating elementary modes using bit pattern trees (Terzer and Stelling 2008).

Yields for all genetically independent pathways

Metabolic pathway analysis based on elementary mode analysis can rigorously identify all genetically independent pathways that are inherent in a metabolic network (Schuster et al. 2002). Since all elementary modes are unique up to scalar multiples, the fluxes in each mode represent only relative values. The most meaningful values are fluxes of an entire pathway that are normalized with respect to a flux of interest in a reaction such as a substrate flux or a product flux. This pathway definition allows a systematic approach to accurately compare molar yields of a metabolite with respect to another in multiple pathways (Pfeiffer et al. 1999). Knowledge of all possible pathways inherent in a metabolic network allows the assessment of pathways of interest based on their molar yields and hence direct metabolic engineering strategies (Schuster et al. 2000, Schuster et al. 1999). For instance, as shown Figure 1D, 1F, among the complete set of 8 EMs only 6 EMs including EM_1*, EM_3*, EM_4, EM_6, EM_7, and EM_8 utilize the substrate A. The asterisks indicate elementary modes that are also extreme pathways. Within the subset of these 6 EMs, only 4 elementary modes EM_4, EM_6, EM_7, and EM_8 can convert the substrate A into the product P. From these 4 EMs, it is straightforward to calculate the molar yields of P on A as follows: $Y_{P/A} = r_4/r_1$. The result shows that EM_6 and EM_7 achieve the highest molar yield of 2 while EM_4 and EM_8 achieve the lowest molar yield of 1. Therefore, EM_6 and EM_7 are the most efficient pathways to convert A to P while EM_4 and EM_8 are not. It is interesting to observe in this simple network that there is no extreme pathway to convert the substrate A into the product P. This example demonstrates that extreme pathways have limitations for the direct interpretation of functional pathways.

The first reported application of elementary mode analysis to a real biological system for the purpose of metabolic engineering is the optimization of the production of 3-deoxy-D-arabino-heptulosonate-7-phoshpate (DAHP) *in Escherichia coli*. DHAP is a main precursor for the amino acid synthesis pathways of tyrosine, phenylalanine, and tryptophan. Through the examination of flux vector of individual pathways in a simplified metabolic network of *E. coli* central metabolism, the most efficient DAHP producing pathway can be identified and optimized by over-expressing enzymes of key reaction steps, that are likely rate-limiting, to achieve *in vivo* a molar yield of DHAP close to the theoretical value (Liao et al. 1996). With improvements in software development for elementary mode analysis (Pfeiffer et al. 1999, Klamt et al. 2003, von Kamp and Schuster 2006), more complicated metabolic networks of different microorganisms have been analyzed *in silico* to design efficient and robust strains to produce desired products and followed by *in vivo* experiments to confirm prediction. For instance, elementary mode analysis has also been applied to predict anaerobic synthesis of poly-β-hydroxybutyrate (PHB) in transgenic *Saccharomyces cerevisiae* (Carlson et al. 2002) and *E. coli* (Wlaschin et al. 2006, Carlson et al. 2005), the optimal synthesis of recombinant proteins in *E. coli* (Vijayasankaran et al. 2005), the optimal production of L-methionine in *E. coli* and *Corynebacterium glutamicum* (Kromer et al. 2006), and the optimal production of cyanophycin in recombinant strains *Psedomonas putida* and *Ralstonia eutropha* (Diniz et al. 2006). Elementary mode analysis was also applied to study the physiology of the photosynthate metabolism of the chloroplast stroma (Poolman et al. 2003) and the metabolic capabilities of *Methylobacterium exorequens* AM1 (Van Dien and Lidstrom 2002) and purple nonsulfur bacteria (Klamt et al. 2002).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Similar to elementary mode analysis, extreme pathways have also been used to examine the metabolic capabilities of Human Red Blood Cell metabolism (Wiback and Palsson 2002), *Haemophilus influenzae* Rd metabolism (Schilling and Palsson 2000, Papin et al. 2002a) and *Helicobacter pylori* metabolism (Price et al. 2002).

Determination of metabolic flux vector

A metabolic flux vector is defined by the cellular physiological state of a cell under a given growth condition and consists of a weighted average of all elementary modes that are present. It shows what participating reactions are active and how fluxes through these reactions describe the physiological state. Knowledge of the metabolic flux vector helps understanding the cell physiology when perturbations such as genetic modifications and growth conditions are imposed on cell growth. Since metabolic pathway analysis can identify all genetically independent pathways inherent in a metabolic network, any pathway or a non-negative linear combination of pathways such as elementary modes or extreme pathways can describe the physiological states of cellular metabolism under different growth conditions. However, the challenging tasks are to figure out how to assign weighting factors to elementary modes or extreme pathways to describe a physiological state of interest and how to determine these weighting factors when they change from one physiological state to another in response to growth perturbations. Several different approaches using metabolic pathway analysis have been reported with encouraging results.

By using elementary mode analysis, Carlson and Srienc identify the four most efficient physiological states that can map the overall flux states of *E. coli* central metabolism as a function of oxygen consumption rates. These flux states consist of elementary modes that efficiently convert glucose into biomass and maintenance energy under different oxygen supply (Carlson and Srienc 2004). By using only the experimentally determined specific glucose uptake rates, flux vector patterns under different oxygen limitation can be calculated and agree with experimental data (Carlson and Srienc 2004). This result implies that a highly evolved organism evidently functions according to pathways that are highly efficient.

Another method proposed to predict the metabolic flux vector of a cellular physiological state is to use the concept of the α-spectrum (Wiback et al. 2004, Wiback et al. 2003). The α-spectrum defines which extreme pathways can constitute the metabolic flux vector of a physiological state and the range of weighting factors for the corresponding extreme pathways. The α-spectrum is considered to be a conservative technique since it only determines the range of metabolic flux vectors rather than the exact values. The range is calculated by minimizing and maximizing the weighting factors of extreme pathways constrained by some experimentally determined fluxes. The range or width of the α-spectrum becomes narrower if more experimentally determined fluxes or transcriptional regulatory constraints (Covert and Palsson 2003) are available. It should be also noted that due to the possible existence of multiple optimal solutions obtained from linear programming optimization, the number and values of weighting factors assigned for extreme pathways may not be unique. The α-spectrum approach has been applied to investigate the *E. coli* central metabolism (Wiback et al. 2004) and the Human Red Blood Cell metabolism (Wiback et al. 2003). Based on the concept of the α-spectrum, Kurata et al. further introduced a heuristic, non-mechanistic model to determine the range of flux vectors of mutants by incorporating into the model the enzymatic activities of the mutant relative to the wild type and its metabolic flux vectors (Kurata et al. 2007). A closely related optimization method based on linear programming has also been proposed to determine flux vectors of cellular metabolism. With the objective function of maximizing the number of elementary modes constrained by some experimentally determined exchange fluxes, Nookaew et al. could estimate the weighting factors, also called flux regulation coefficients, of elementary

NIH-PA Author Manuscript

modes. This technique has been applied to several *Saccharomyces* species under various growth conditions (Nookaew et al. 2007).

Poolman et al. proposed an alternative method based the Moore-Penrose generalized inverse to determine weighting factors for elementary modes such that the sum of all weighting factors squared is minimized. The technique also helps to evaluate the measurement errors for experimentally determined fluxes. The approach was applied to calculate the metabolic flux vectors of *Lactobacillus rhamnosus* during batch growth and to examine the fluctuations of weighting factors for elementary modes during the fermentation (Poolman et al. 2004). A similar approach was also applied to identify weighting factors of elementary modes for yeast glycolysis (Schwartz and Kanehisa 2005). Schwartz and Kenehisa further investigated the effect of change in enzymatic kinetics on the steady state metabolic flux vector and hence on the weighting factors of elementary modes by incorporating kinetic modeling (Schwartz and Kanehisa 2006). First, the kinetic model was simulated to obtain steady state metabolic flux vectors for different random perturbations of enzyme kinetics. These metabolic flux vectors were then decomposed to identify the weighting factors of elementary modes. This approach facilitates the identification of the most dominant subset of biologically active elementary modes among the complete set of elementary modes that can be very large.

For a large metabolic network, the total number of elementary modes can reach several millions. Wang et al. proposed a quadratic programming approach to assign weighting factors for a smaller subset of elementary modes that are randomly chosen by minimizing the difference between the calculated fluxes and measured fluxes. This approach can potentially identify the dominant modes representing cellular metabolism under a given growth condition (Wang et al. 2007). It will be interesting to see whether the subset of identified elementary modes coincides with that of the extreme pathways since the set of extreme pathways is typically several orders of magnitude smaller than the set of elementary modes. This task can be easily examined by using the rank/nullity test developed by Boley et al. (Boley et al. 2008).

Another interesting approach to calculate weighting factors comes from classical thermodynamics (Wlaschin et al. 2006). The individual weighting factors can be formally viewed as probabilities that a given elementary mode is used within a functioning metabolism. In a metabolic network, as substrates are consumed the energy contained within their chemical bonds is used by the cellular metabolism. Some of this energy is used to build new bonds between metabolites to build larger macromolecules. Usable energy is also lost as heat. This constant flux of products and substrates can be modeled in classical thermodynamics as an open system in near equilibrium steady state (Qian and Beard 2005, Qian et al. 2003). Elementary modes with the same overall stoichiometry have been grouped into families of EMs. Each family stoichiometry is viewed as a “reaction” in classical thermodynamics. This effectively reduced the number of weighting factors that could be estimated from a completely determined system (Wlaschin et al. 2006). It was found that the weighting factors correlated inversely with the corresponding entropies of reactions. The reaction generating the least amount of entropy had the largest weighting factor. This result is a direct validation of Prigogine’s principle of minimal entropy production in an open irreversible system operating at steady state (Prigogine 1945).

Analysis of metabolic network properties

Network structure

Determination of the complete set of all minimal and unique pathways by metabolic pathway analysis allows a systematic evaluation of basic structures of a metabolic network

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

such as pathway lengths, reaction participation in a pathway, and the correlated subsets of reactions. Papin et al. reported a simple implementation of linear algebra on the extreme pathway matrix to form the symmetric Pathway Length Matrix and Reaction Participation Matrix and to extract from these matrices useful characteristics demonstrated for cellular metabolisms of *Helicobacter pylori* and *Haemophilus influenza* (Papin et al. 2002b). The diagonal elements of the pathway length matrix define how many reactions participate in an extreme pathway while the off-diagonal elements represent the number of reactions shared by two different extreme pathways. Similarly, the diagonal elements of the reaction participation matrix represent how many extreme pathways utilize a particular reaction while the off-diagonal elements define the number of extreme pathways shared by two different reactions. This analysis is useful to classify groups of reactions that always, sometimes, or never contribute to the synthesis of a desired product. Even though the technique is demonstrated for the extreme pathway matrix, it is completely applicable also for the elementary mode matrix. It should be noted that the method of detecting correlated reaction subsets can also be implemented by analyzing the null space of the stoichiometric matrix without using the extreme pathway matrix or elementary mode matrix (Pfeiffer et al. 1999, Poolman et al. 2007).

Network robustness

Decomposition of a complex cellular metabolism by metabolic pathway analysis into a complete set of genetically independent pathways reveals detailed information of interrelationships between metabolic network structure and cellular functionality such as cellular robustness and regulation. In the context of cellular metabolism, robustness is defined as the ability of cells to achieve the optimal performance even under perturbations imposed by a gene knockout (Stelling et al. 2002, Barkai and Leibler 1997, Edwards and Palsson 2000, Edwards and Palsson 2000, Stelling et al. 2004). By analyzing the entire set of elementary modes of *E. coli* metabolism, Stelling et al. reported that the network is highly robust and maintains high biomass yield in cells containing single gene knockouts (Stelling et al. 2002). The robustness of cellular metabolism is mainly due to the redundancy of pathway options that the wild type can choose from to function to achieve similar performance. Such a high redundancy is also observed in *Haemophilus influenza* (Papin et al. 2002a) and *Helicobacter pylori* (Price et al. 2002).

Network regulation

Metabolic pathway analysis can also be applied to reveal the regulation pattern based on the structure of a metabolic network. The functional operation of pathways under a given growth condition is a result of complex interactions of cellular phenotype and genotype that can be quantified though metabolic fluxes and expression transcripts. To approach the problem, Stelling et al. introduced a technique to calculate the control-effective fluxes from the set of elementary modes. The resulting fluxes are weight-averaged values that are taken into account for both network flexibility and network efficiency (Stelling et al. 2002). In the paper, Stelling et al. defined that the network flexibility reflects the ability of cellular metabolism to adjust with different growth environments while network efficiency represents the ability of cellular metabolism to achieve maximal outcomes such maximal growth yield by utilizing minimal resources such as investments to produce enzymes. The technique allows direct calculation of flux ratios for each reaction in a pathway and provides a basis for direct comparison with expression transcripts under perturbations. By analyzing the central metabolism of *E. coli*, Stelling et al. found that there is strong correlation of control-effective flux ratios and corresponding expression transcript ratios for growth on glucose and on acetate (Stelling et al. 2002). Cakir et al. also used the same approach to investigate in yeast *Saccharomyces* the correlation of control-effective flux ratios and corresponding expression transcript ratios for growth on glucose and on galactose (Cakir et

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

al. 2004) and for shifts from a fermentative carbon source to a non-fermentative carbon source such as ethanol, acetate, and lactate during fermentation (Cakir et al. 2007).

Price et al. suggested a very interesting approach that the single value decomposition on the extreme pathway matrix can be applied to identify the key branch points of the metabolic network. This information is useful for designing strategies to redirect carbon flow to desired pathways (Price et al. 2002).

Carlson has recently reported a very interesting approach based on elementary mode analysis to investigate the principal design of cellular metabolism through cost-benefit analysis (Carlson 2007). This analysis reveals an interesting correlation between synthesis requirements for cellular operation (investment cost) and thermodynamic efficiency (operating cost) under growth conditions of nutrient sufficiency, scarcity, and excess. Due to the high robustness of cellular metabolism afforded by isozymes and parallel pathways, cells can switch on and off pathways to modulate their metabolism to accommodate the nutrient availability. Interestingly, the study reported that the efficient pathways require high cost of cellular synthesis for thermodynamic efficiency, but cells trade off to utilize the inefficient pathways under growth conditions of nutrient scarcity or excess such as “overflow metabolism”.

Network fragility

In complete contrast to the structural robustness of the cellular metabolism, fragility is also an important concept to help understand the network functionality. The concept of a Minimal Cut Set has been introduced to determine the minimal set of reactions whose deletion completely blocks a target (Klamt and Gilles 2004, Klamt 2006). The identification of a minimal cut set is useful in metabolic engineering and in the identification of drug targets (Klamt and Gilles 2004). Several techniques have been proposed to measure the fragility of the metabolic network (Klamt and Gilles 2004, Wilhelm et al. 2004, Behre et al. 2008).

Rational design of efficient host strains

Knowledge of all unique pathways existing in a metabolic network allows a rational *in silico* design of efficient host strains with specialized metabolic functionalities. By eliminating inefficient pathways, host strains can be forced to function only according to efficient pathways. Furthermore, host strains can be constrained such that the operation of efficient pathways and cell growth are coupled. With this strategy of strain design, the host strains can be used also to select and evolve foreign pathways of interest that has related function to the native efficient pathway.

Three main steps are involved in designing the novel host strains as shown in Figure 2. Step 1 is pathway identification. By applying elementary mode analysis, all unique pathways existing in a metabolic network can be identified. It should be emphasized that this step of pathway identification may not be done by using extreme pathway analysis. The main reason is that extreme pathways are only a subset of elementary modes. A simple example in Figure 1F demonstrates that no extreme pathway can be identified to convert A to P at the highest molar yield of 2. From Figure 2, the efficient pathway shown in blue can be easily selected among other inefficient pathways shown in green since knowledge of entire sets of possible pathways can be achieved by elementary mode analysis. Under a given growth condition, cells can select any combination of pathways to operate to maximize their fitness. Due to the existence of multiple pathway options, deletion of the efficient pathway will not affect cell growth since cells can use other pathways to function. The second step is pathway selection. In order to enforce cells to function only according to the efficient pathway by

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

deleting inefficient ones, the strain can be designed such that growth and operation of the efficient pathway are complemented. The third step is pathway evolution. This step is optional. The strain designed in step 2 can be further utilized for pathway evolution. Due to the complementation of cell growth and operation of the efficient pathway, it is possible to replace an enzymatic reaction of the efficient pathway with a potential candidate that has similar enzymatic activities from a library of mutated enzymes and evolve this enzyme.

The concept of strain design can be demonstrated through the analysis of a simple metabolic network shown in Figure 1. For the sake of simplicity, cell growth is not incorporated into the analysis of this simple network (see Trinh et al. for this topic (Trinh et al. 2008)). In this example, it is of particular interest to identify the most efficient pathway that produces the product P from the substrate A and to enforce the operation of this pathway. We start by applying elementary mode analysis on the simple network. We can identify 6 EMs (EM_1*, EM_3*, EM_4, EM_6, EM_7, and EM_8) out of 8 EMs that can utilize the substrate A and only 4 EMs (EM_4, EM_6, EM_7, and EM_8) that can convert the substrate A into the product P (Figure 1D, 1F). EM_7 and EM_8 are the two most efficient pathways that can convert the substrate A into the product P at the maximal molar yield of 2 (Cmole/Cmole). In step 2 of pathway selection, we are interested in identifying multiple reaction deletions such that only the operation of either EM_7 or EM_8 is feasible. The trivial solution is to remove all null fluxes in either EM_7 or EM_8. However, due to the coupling of reactions in a pathway and of pathways in the network, it is not necessary to remove the complete set of null fluxes in the efficient pathways of interest. Instead, only a minimal subset of the null fluxes can accomplish this. This approach requires multiple sequential rounds of reaction deletions that follow three basic rules (Trinh et al. 2008). First, to identify a selection target, the effect of elimination of individual reactions on the number of EMs is evaluated. For instance, Figure 3A–D demonstrates the effect of single reaction deletion on the fraction of remaining elementary modes. In these figures, the deleted reactions are sorted according to the increasing number of fraction of remaining elementary modes. Second, the maximum and minimum yields of desired products are also evaluated among the set of remaining elementary modes when a reaction is deleted as shown in Figure 3A–D. Third, the reaction with the smallest fraction of remaining elementary modes that still support maximum yields of desired products is chosen for elimination. For example, the first round of reaction deletion is demonstrated in Figure 3A. The choice of deleting R7 is not optimal because it reduces the maximal possible yield of P while deletion of either R1 or R4 is obviously detrimental to the feasible operation of the efficient pathways of interest. The most efficient possible choice is any of reactions R2, R5, and R6r. In the next rounds of reaction deletion as shown in Figure 3B–D, we choose R2 as the first reaction deletion for the purpose of demonstration. Other choices of R5 and R6r are presented in Figure 3F. We continue sequential steps of single reaction deletion until no further improvement can be achieved. As shown in Figure 3D, since only one EM remains and operates according to the most efficient pathway to convert the substrate A to the product P, no further deletion is necessary. However, in Figure 2B, if there is no continuation of sequential reaction deletion after the first round of eliminating R2, inefficient pathways for converting the substrate A into the product P still exist. In Figure 3B–D, the deletion of a reaction that results in the mode fractions of 1 indicates that all elementary modes analyzed at the round of deletion have null flux through this particular reaction. For instance, in the round 3 of deletion as shown in Figure 3C, deletion of R2 and R3 automatically eliminates R6r and R9. Figure 3E summarizes the effect of sequential reaction deletion on the total number of elementary modes as well as the yield range of desired product P. The approach here ensures identification of a minimal set of reactions that should be deleted to enforce the operation of pathways of interest. However, such a set is not unique. As shown in Figure 3F, alternative sets of deleted reactions are {R2, R6r, R8r} and {R3, R5, R8r}. The sets {R2, R3, R8r} and {R2, R6r, R8r} support the operation of EM6 while {R3, R5, R8r} supports the functioning of EM7.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

This approach has been successfully applied to design and construct minimal cells with efficient metabolic functionalities for biomass production (Trinh et al. 2006) and for ethanol production (Trinh et al. 2008). The characterization of the most efficient biomass producing *E. coli* TCS062 containing 6 genetic knockouts was conducted in chemostats under different dilution rates. The result showed that the mutant TCS062 achieved approximately 30% improvement in growth yield when directly compared to the wildtype under identical growth conditions (Trinh et al. 2006). The mutant performance closely matched the theoretical prediction. Trinh et al. also applied a similar strategy to design a minimal *E. coli* cell with efficient functionality for ethanol production. The designed mutant TCS083/pLOI297 containing 8 genetic knockouts was constructed and characterized in controlled batch bioreactors. The mutant was able to efficiently convert hexoses and pentoses to ethanol close to the theoretical limit in a simultaneous manner without glucose catabolite repression (Trinh et al. 2008). In addition, in this strain cell growth and ethanol production are highly coupled as predicted by the model. These two examples further demonstrate the feasibility of the approach for rational design of most efficient organisms with minimized metabolic functionalities tailored for specific applications.

Conclusion

Elementary mode analysis is a useful metabolic pathway analysis tool to characterize cellular metabolism. It provides insights into cellular physiology, robustness and regulation. The application covers many key aspects of metabolic engineering strategies such as identifying key regulatory branch points of a metabolic pathway, designing minimal cells with minimized but efficient metabolic functionalities tailored for specific applications, and identifying key enzyme targets for drug design. It can also be used as a systems biology tool to elucidate the interaction of cellular genotypes and phenotypes by the analysis of flux ratios and transcript ratios and interpreting the principal design of cellular metabolism. Since metabolic pathway analysis is the analysis of structural invariants without introduction of kinetic parameters to describe dynamic interaction of cellular metabolisms, it can only predict discrete states of cellular metabolisms. Therefore, introduction of dynamic features into structural modeling can offer a potential adventure to understand the dynamics of cellular metabolism. Furthermore, current techniques of determining the exact physiological state are still semi empirical because some fluxes need to be measured experimentally and used for calculation. Therefore, developing more robust techniques to accurately extract cellular physiological states with the least amount of experimental data is useful to predict cellular metabolism. This approach promises to become practical if intelligent design of operational rules based on the transcriptomic, proteomic, and metabolomic data can be developed and incorporated into the model.

Acknowledgments

We thank NIH for supporting this work (GM077529).

References

Barkai N, Leibler S. Robustness in simple biochemical networks. Nature 1997;387:913–917. [PubMed: 9202124]

Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S. Structural robustness of metabolic networks with respect to multiple knockouts. J Theor Biol 2008;252:433–441. [PubMed: 18023456]

Bell SL, Palsson BO. Expa: a program for calculating extreme pathways in biochemical reaction networks. Bioinformatics 2005;21:1739–1740. [PubMed: 15613397]

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2008;36:D25–30. [PubMed: 18073190]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Boley, D.; Jevremovic, D.; Srienc, F.; Trinh, CT. A Simple Rank Test to Distinguish Extreme Pathways from Elementary Flux Modes in Metabolic Networks. 2008. (submitted)

Cakir T, Kirdar B, Onsan ZI, Ulgen KO, Nielsen J. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in Saccharomyces cerevisiae. BMC Syst Biol 2007;1:18. [PubMed: 17408508]

Cakir T, Kirdar B, Ulgen KO. Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. Biotechnol Bioeng 2004;86:251–260. [PubMed: 15083505]

Carlson R, Srienc F. Fundamental Escherichia coli Biochemical Pathways for Biomass and Energy Production: Identification of Reactions. Biotech Bioeng 2004;85:1–18.

Carlson R, Fell D, Srienc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol Bioeng 2002;79:121–134. [PubMed: 12115428]

Carlson R, Srienc F. Fundamental Escherichia coli biochemical pathways for biomass and energy production: creation of overall flux states. Biotechnol Bioeng 2004;86:149–162. [PubMed: 15052634]

Carlson R, Wlaschin A, Srienc F. Kinetic studies and biochemical pathway analysis of anaerobic poly-(R)-3-hydroxybutyric acid synthesis in Escherichia coli. Appl Environ Microbiol 2005;71:713–720. [PubMed: 15691921]

Carlson RP. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics 2007;23:1258–1264. [PubMed: 17344237]

Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucl Acids Res 2006;34:D511–516. [PubMed: 16381923]

Clarke BL. Stoichiometric network analysis. Cell Biophys 1988;12:237. [PubMed: 2453282]

Clarke BL. Complete set of steady states for the general stoichiometric dynamical system. J Chem Phys 1981;75:4970–4979.

Covert MW, Palsson BO. Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space. Journal of Theoretical Biology 2003;221:309–325. [PubMed: 12642111]

Diniz SC, Voss IV, Steinbuchel A. Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design. Biotechnol Bioeng 2006;93:698–717. [PubMed: 16435401]

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A 2007;104:1777–1782. [PubMed: 17267599]

Edwards JS, Palsson BO. Robustness Analysis of the *Escherichia coli* Metabolic Network. Biotechnol Prog 2000;16:927. [PubMed: 11101318]

Edwards, JS.; Ramakrishna, R.; Schilling, CH.; Palsson, BO. Metabolic Flux Balance Analysis. Lee, SY.; Papoutsakis, ET., editors. Metabolic Engineering. Marcel Dekker, Inc; New York: 1999. p. 13-57.

Edwards JS, Palsson BO. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 2000;97:5528–5533. [PubMed: 10805808]

Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotech 2001;19:125–130.

Feinberg M, Horn FJM. Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. Chemical Engineering Science 1974;29:775–787.

Gagneur J, Klamt S. Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics 2004;5:175. [PubMed: 15527509]

Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U. COPASI--a COmplex PAthway SImulator. Bioinformatics 2006;22:3067–3074. [PubMed: 17032683]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36:D480–4. [PubMed: 18077471]

Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I, Collado-Vides J, Gama-Castro S, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Bonavides-Martinez C, Ingraham J. Multidimensional annotation of the Escherichia coli K-12 genome. Nucleic Acids Res 2007;35:7577–7590. [PubMed: 17940092]

Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. Curr Opin Biotechnol 2003;14:491–496. [PubMed: 14580578]

Klamt S, Schuster S, Gilles ED. Calculability Analysis in Underdetermined Metabolic Networks Illustrated by a Model of the Central Metabolism in Purple Nonsulfur Bacteria. Biotechnol Bioeng 2002;77:734–751. [PubMed: 11835134]

Klamt S. Generalized concept of minimal cut sets in biochemical networks. BioSystems 2006;83:233–247. [PubMed: 16303240]

Klamt S, Gagneur J, von Kamp A. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. Syst Biol (Stevenage) 2005;152:249–255. [PubMed: 16986267]

Klamt S, Saez-Rodriguez J, Gilles ED. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst Biol 2007;1:2. [PubMed: 17408509]

Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. Mol Biol Rep 2002;29:233–236. [PubMed: 12241063]

Klamt S, Gilles ED. Minimal cut sets in biochemical reaction networks. Bioinformatics 2004;20:226–234. [PubMed: 14734314]

Klamt S, Stelling J. Two approaches for metabolic pathway analysis? Trends Biotechnol 2003;21:64–69. [PubMed: 12573854]

Klamt S, Stelling J, Ginkel M, Gilles ED. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. Bioinformatics 2003;19:261–269. [PubMed: 12538248]

Koch I, Junker BH, Heiner M. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. Bioinformatics 2005;21:1219–1226. [PubMed: 15546934]

Kromer JO, Wittmann C, Schroder H, Heinzle E. Metabolic pathway analysis for rational design of L-methionine production by Escherichia coli and Corynebacterium glutamicum. Metab Eng 2006;8:353–369. [PubMed: 16621639]

Kurata H, Zhao Q, Okuda R, Shimizu K. Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. BMC Systems Biology 2007;1:31. [PubMed: 17640350]

Liao JC, Hou SY, Chao YP. Pathway Analysis, Engineering, and Physiological Considerations for Redirecting Central Metabolism. Biotechnol Bioeng 1996;52:129–140. [PubMed: 18629859]

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. Qualitative analysis of biochemical reaction systems. Computers in Biology and Medicine 1996;26:9–24. [PubMed: 8654057]

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. Computer-aided synthesis of biochemical pathways. Biotechnol Bioeng 1990;36:1119–1132. [PubMed: 18595053]

Milner PC. The Possible Mechanisms of Complex Reactions Involving Consecutive Steps. J Electrochem Soc 1964;111:228.

Nookaew I, Meechai A, Thammarongtham C, Laoteng K, Ruanglek V, Cheevadhanarak S, Nielsen J, Bhumiratana S. Identification of flux regulation coefficients from elementary flux modes: A systems biology tool for analysis of metabolic networks. Biotechnol Bioeng 2007;97:1535–1549. [PubMed: 17238207]

Papin JA, Price ND, Edwards JS, Palsson BBO. The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy. J Theor Biol 2002a; 215:67–82. [PubMed: 12051985]

Papin JA, Price ND, Palsson BO. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. Genome Res 2002b;12:1889–1900. [PubMed: 12466293]

NIH-PA Author Manuscript

Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO. Metabolic pathways in the post-genome era. Trends Biochem Sci 2003;28:250–258. [PubMed: 12765837]

Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. Trends Biotechnol 2004;22:400–405. [PubMed: 15283984]

Pfeiffer T, Sanchez-Valdenebro I, Nuno J, Montero F, Schuster S. METATOOL: for studying metabolic networks. Bioinformatics 1999;15:251–257. [PubMed: 10222413]

Poolman MG, Venkatesh KV, Pidcock MK, Fell DA. A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. Biotechnol Bioeng 2004;88:601–612. [PubMed: 15470705]

Poolman MG, Fell DA, Raines CA. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. Eur J Biochem 2003;270:430–439. [PubMed: 12542693]

Poolman MG, Sebu C, Pidcock MK, Fell DA. Modular decomposition of metabolic systems via null-space analysis. J Theor Biol 2007;249:691–705. [PubMed: 17949756]

Price ND, Papin JA, Palsson BO. Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis. Genome Res 2002;12:760–769. [PubMed: 11997342]

Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–897. [PubMed: 15494745]

Price ND, Reed JL, Papin JA, Famili I, Palsson BO. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. Biophys J 2003;84:794–804. [PubMed: 12547764]

Prigogine I. Modération et transformations irréversibles des systèmes ouverts. Bull Acad Roy Belg Cl Sci 1945;31:600–606.

Qian H, Beard DA. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. Biophys Chem 2005;114:213–220. [PubMed: 15829355]

Qian H, Beard DA, Liang S. Stoichiometric network theory for nonequilibrium biochemical systems. European Journal of Biochemistry 2003;270:415–421. [PubMed: 12542691]

Reder C. Metabolic control theory: a structural approach. J Theor Biol 1988;135:175–201. [PubMed: 3267767]

Rockafellar, RT. Convex analysis. Princeton University Press; Princeton, N.J: 1970.

Roels, JA. Energetics and kinetics in biotechnology. Elsevier Biomedical Press; Amsterdam; New York: 1983.

Samatova, NF. Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways. Parallel and Distributed Processing Symposium, Proceedings International, IPDPS; 2002; 2002. Abstracts and CD-ROM 185–192

Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 2000;203:229–248. [PubMed: 10716907]

Schilling CH, Palsson BO. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. J Theor Biol 2000;203:249–283. [PubMed: 10716908]

Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol 2007;3

Schuster R, Schuster S. Refined Algorithm and Computer Program Calculating All Non-Negative Fluxes Admissible in Steady States of Biochemical Reaction Systems with and without Some Fluxes Rates Fixed. CABIOS 1993;9:79–85. [PubMed: 8435772]

Schuster S, Fell DA, Dandekar T. A General Definition of Metabolic Pathways Useful for Systematic Organization and Analysis of Complex Metabolic Networks. Nat Biotechnol 2000;18:326–332. [PubMed: 10700151]

Schuster S, Hilgetag C, Woods JH, Fell DA. Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. J Math Biol 2002;45:153–181. [PubMed: 12181603]

Schuster S, Hilgetag S. On Elementary Flux Modes in Biochemical Reaction Systems At Steady State. J Biol Syst 1994;2:165–182.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 1999;17:53–60. [PubMed: 10087604]

Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae. Bioinformatics 2002;18:351–361. [PubMed: 11847093]

Schuster S, Schuster R. Detecting strictly detailed balanced subnetworks in open chemical reaction networks. Journal of Mathematical Chemistry 1991;6:17–40.

Schwartz JM, Gaugain C, Nacher JC, de Daruvar A, Kanehisa M. Observing metabolic functions at the genome scale. Genome Biol 2007;8:R123. [PubMed: 17594483]

Schwartz J, Kanehisa M. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 2006;7:186. [PubMed: 16584566]

Schwartz J, Kanehisa M. A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. Bioinformatics 2005;21:ii204–205. [PubMed: 16204104]

Schwarz R, Liang C, Kaleta C, Kuhnel M, Hoffmann E, Kuznetsov S, Hecker M, Griffiths G, Schuster S, Dandekar T. Integrated network reconstruction, visualization and analysis using YANAsquare. BMC Bioinformatics 2007;8:313. [PubMed: 17725829]

Schwarz R, Musch P, von Kamp A, Engels B, Schirmer H, Schuster S, Dandekar T. YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities. BMC Bioinformatics 2005;6:135. [PubMed: 15929789]

Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. PNAS 2002;99:15112–15117. [PubMed: 12415116]

Seressiotis A, Bailey JE. MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways. Biotechnol Bioeng 1988;31:587–602. [PubMed: 18584649]

Song HS, Ramkrishna D. Reduction of a set of elementary modes using yield analysis. Biotechnol Bioeng. 2008 (Epub ahead of print).

Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. Nature 2002;420:190–193. [PubMed: 12432396]

Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J. Robustness of cellular functions. Cell 2004;118:675–685. [PubMed: 15369668]

Stephanopoulos, G.; Aristidou, AA.; Nielsen, JH. Metabolic engineering: principles and methodologies. Academic Press; San Diego: 1998.

Terzer M, Stelling J. Large scale computation of elementary flux modes with bit pattern trees. Bioinformatics. 2008 (Epub ahead of print).

Trinh CT, Unrean P, Srienc F. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 2008;74:3634–3643. [PubMed: 18424547]

Trinh CT, Carlson R, Wlaschin A, Srienc F. Design, construction and performance of the most efficient biomass producing E. coli bacterium. Metab Eng 2006;8:628–638. [PubMed: 16997589]

Urbanczik R. SNA--a toolbox for the stoichiometric analysis of metabolic networks. BMC Bioinformatics 2006;7:129. [PubMed: 16533403]

Urbanczik R, Wagner C. Functional stoichiometric analysis of metabolic networks. Bioinformatics 2005a;21:4176–4180. [PubMed: 16188931]

Urbanczik R, Wagner C. An improved algorithm for stoichiometric network analysis: theory and applications. Bioinformatics 2005b;21:1203–1210. [PubMed: 15539452]

Van Dien SJ, Lidstrom ME. 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 2002;78:296–312. [PubMed: 11920446]

Vijayasankaran N, Carlson R, Srienc F. Metabolic pathway structures for recombinant protein synthesis in Escherichia coli. Appl Microbiol Biotechnol 2005;68:737–746. [PubMed: 15739064]

von Kamp A, Schuster S. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics. 2006

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Wagner C. Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems. J Phys Chem B 2004;108:2425–2431.

Wang Q, Yang Y, Ma H, Zhao X. Metabolic network properties help assign weights to elementary modes to understand physiological flux distributions. Bioinformatics 2007;23:1049–1052. [PubMed: 17341495]

Wiback SJ, Mahadevan R, Palsson BO. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum. Biotechnol Bioeng 2004;86:317–331. [PubMed: 15083512]

Wiback SJ, Mahadevan R, Palsson BO. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J Theor Biol 2003;224:313–324. [PubMed: 12941590]

Wiback SJ, Palsson BO. Extreme pathway analysis of human red blood cell metabolism. Biophys J 2002;83:808–818. [PubMed: 12124266]

Wiechert W. 13C Metabolic Flux Analysis. Metabolic Engineering 2001;3:195–206. [PubMed: 11461141]

Wilhelm T, Behre J, Schuster S. Analysis of structural robustness of metabolic networks. Syst Biol (Stevenage) 2004;1:114–120. [PubMed: 17052121]

Wlaschin AP, Trinh CT, Carlson R, Srienc F. The fractional contributions of elementary modes to the metabolism of Escherichia coli and their estimation from reaction entropies. Metab Eng 2006;8:338–352. [PubMed: 16581276]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

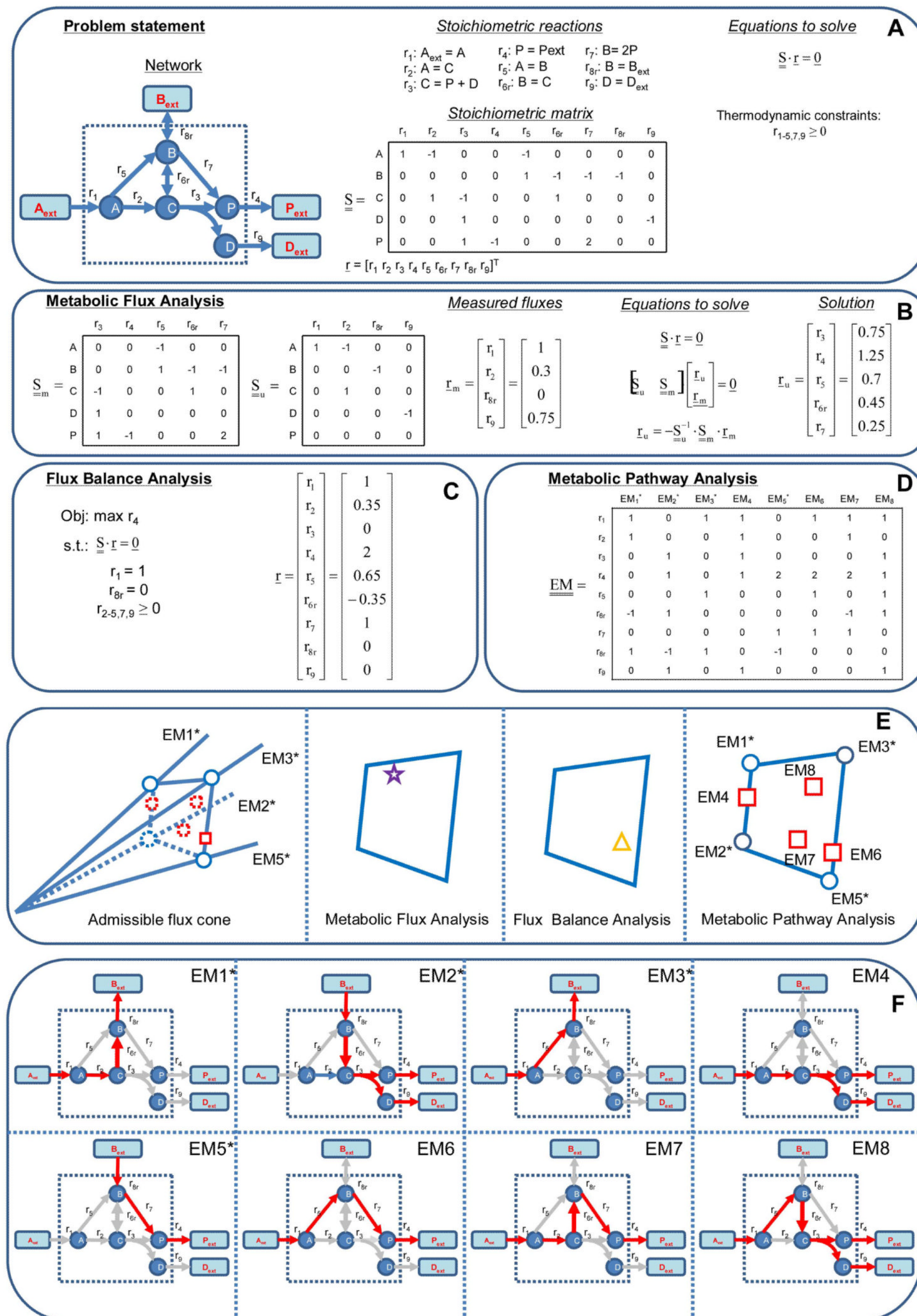

Figure 1.

Analysis of a simple metabolic network. **A. Problem statement**. The system is confined by the dashed line and consisted of internal metabolites A, B, C, D, and P that are linked through internal reactions r_2, r_3, r_5, r_{6r}, r_7. External metabolites A_{ext}, B_{ext}, D_{ext}, and P_{ext} can either enter or exit the system by exchange reactions r_1, r_{8r}, r_4, r_9, respectively. Two reversible reactions r_{6r} and r_{8r} allow the reactions to proceed in either forward or backward directions. By definition reversible reactions can have either positive or negative fluxes but irreversible reactions only have non-negative fluxes. From the stoichiometric reactions, a stoichiometric matrix $\underline{\underline{S}}$ for internal metabolites can be set up where rows correspond to

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

internal metabolites and columns represent reactions. Each element s_{ij} represents the stoichiometric coefficient of a metabolite i in reaction j. The coefficient is positive if the metabolite (product) is produced and negative if the metabolite (reactant) is consumed. **B. Metabolic Flux Analysis.** The stoichiometric matrix $\underline{\underline{S}}$ is partitioned into $\left[\begin{array}{cc}\underline{\underline{S}}_u & \underline{\underline{S}}_m\end{array}\right]$ and $\underline{r}$ is partitioned into $\left[\begin{array}{cc}\underline{r}_u & \underline{r}_m\end{array}\right]$ where subscripts u, m are referred to "unmeasured" and "measured", respectively. The calculation of $\underline{r}_u$ is feasible if and only if $\underline{r}_m$ is known. **C. Flux Balance Analysis.** The objective function is to maximize flux through the secretion of desired product P when only A is considered the only substrate with $r_1 = 1$ unit. **D. Metabolic Pathway Analysis.** By using METATOOL, elementary mode analysis identifies 8 unique elementary modes listed in the matrix form $\underline{\underline{EM}}$ where rows correspond to reactions and columns represent elementary modes. The asterisks indicate that these elementary modes are also extreme pathways. **E. Geometric interpretation.** The admissible flux cone represents all possible pathways that can exist. The cone is spanned by four extreme pathways that represent the edge of the cone. Some elementary modes lie on the face and inside the cone. Metabolic Flux Analysis identifies only a pathway that lies anywhere in the cone (star in purple). For instance, metabolic flux vector in **B** can be expressed as a nonnegative linear combination of extreme pathways or elementary modes in **D** as follows: $r = 0.3\ \underline{EM}^*_1 + 0.75\ \underline{EM}^*_2 + 0.7\ \underline{EM}^*_3 + 0.25\ \underline{EM}^*_5$ (the asterisks refer to elementary modes that are also extreme pathways) or $r = 0.2169\ \underline{EM}_4 + 0.1669\ \underline{EM}_6 + 0.0831\ \underline{EM}_7 + 0.5331\ EM_8$. Similarly, Flux Balance Analysis represents only a pathway that lies anywhere in the cone (triangle in orange) and satisfies the defined objective function. For instance, metabolic flux vector in **C** can be expressed as a non-negative linear combination of extreme pathways or elementary modes in **D** as follows: $\underline{r} = 0.35\ \underline{EM}^*_1 + 0.65\ \underline{EM}^*_3 + 1.0\ \underline{EM}^*_5$ or $\underline{r} = 0.65\ \underline{EM}_6 + 0.35\ \underline{EM}_7$. Metabolic Pathway Analysis identifies all genetically independent pathways with extreme pathways shown in blue circle and with elementary modes shown in red square. **F.** Graphical representation of extreme pathways and elementary modes for the simple metabolic network.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

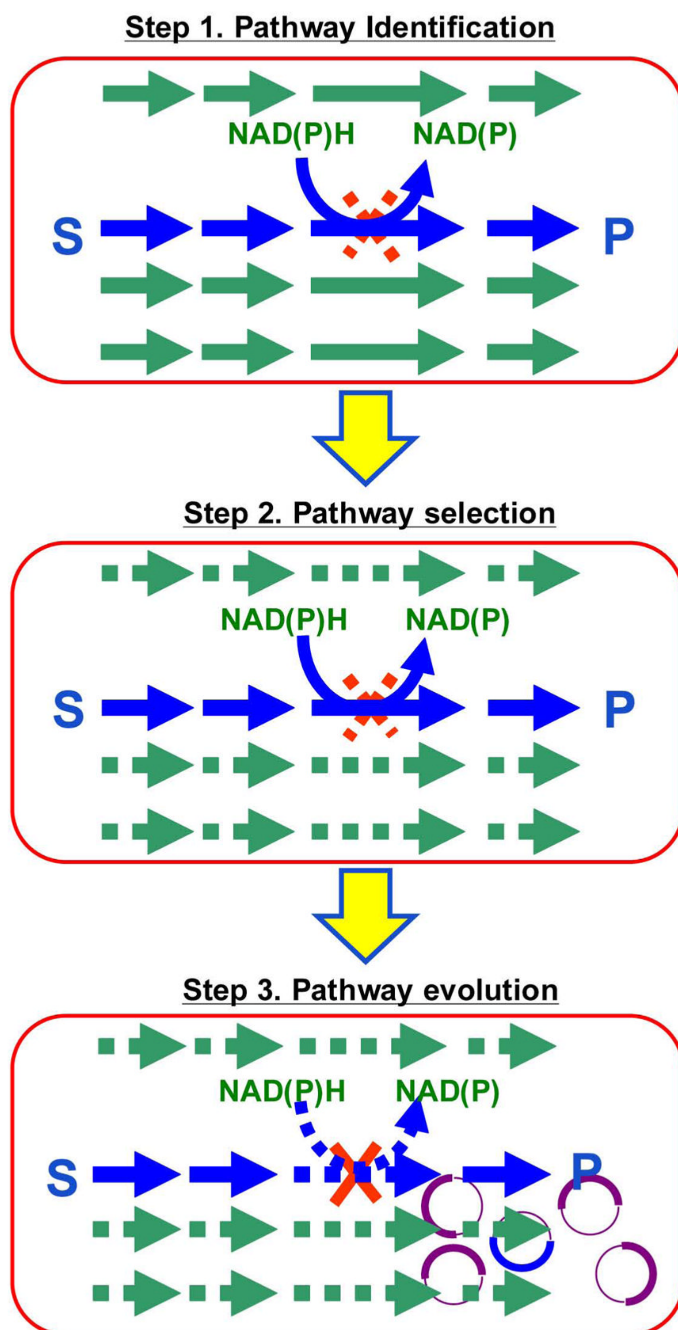

Figure 2.

Strategy for rational design of cells with the most efficient minimized metabolic functionalities. In **Step 1**, arrows shown in blue constitute the efficient pathway of interest to convert S to P while arrows shown in green represent the inefficient pathway for conversion of S to P. The dashed X implies the proposed disruption of a reaction in the efficient pathway to block its operation. The blockage of the efficient pathway does not affect cellular function since cells can choose other inefficient pathways to function. In **Step 2**, the dashed arrows of the inefficient pathways represent their inactivation. The blockage of the efficient pathway in this step inhibits cellular function because the strain is designed to couple cell growth and the operation of the efficient pathway. In **Step 3**, the circles represent the

plasmids carrying a library of mutated genes encoding corresponding mutated enzymes that may complement the enzyme activity of the reaction disrupted in the efficient pathway.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

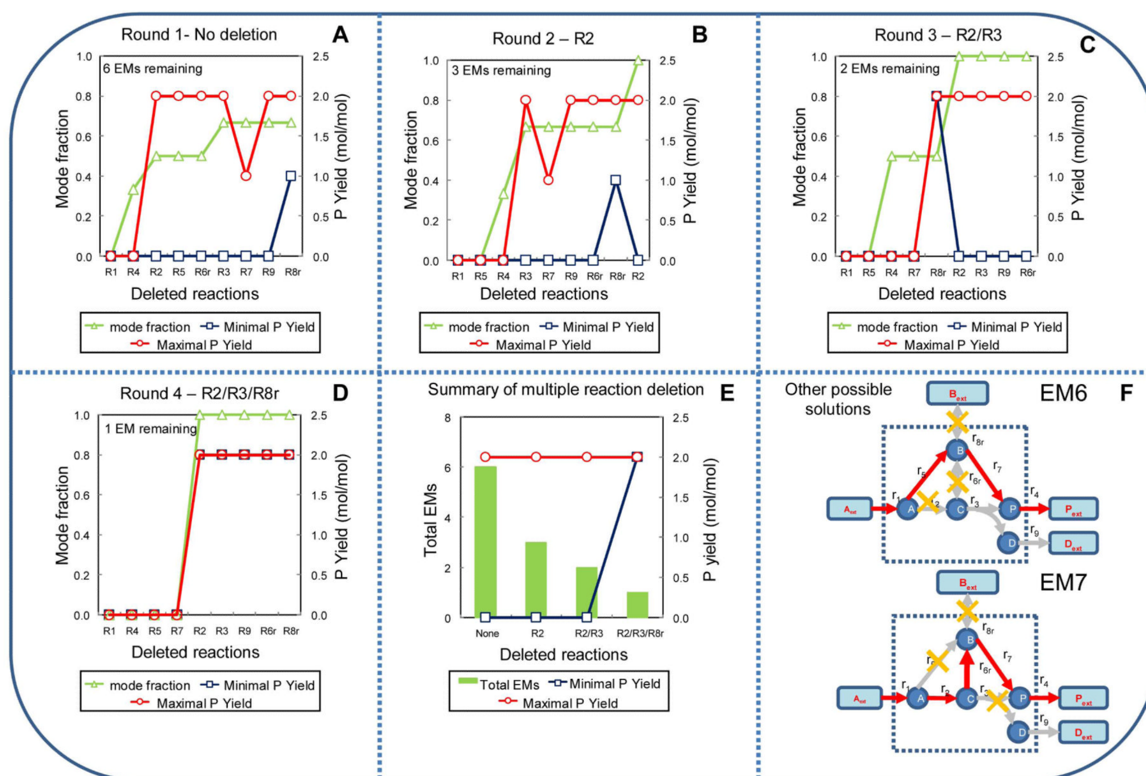

Figure 3.

Implementation of multiple reaction deletion to rationally design the most efficient pathways for converting the substrate A into the product P. **A–D.** Evaluation of sequential reaction deletion. **E**. Summary of the effect of multiple reaction deletion on the total number of elementary modes as well as the yield range of the desired product P. **F**. Alternative sets of reaction deletion to achieve the most efficient pathways to convert the substrate A into the product P.