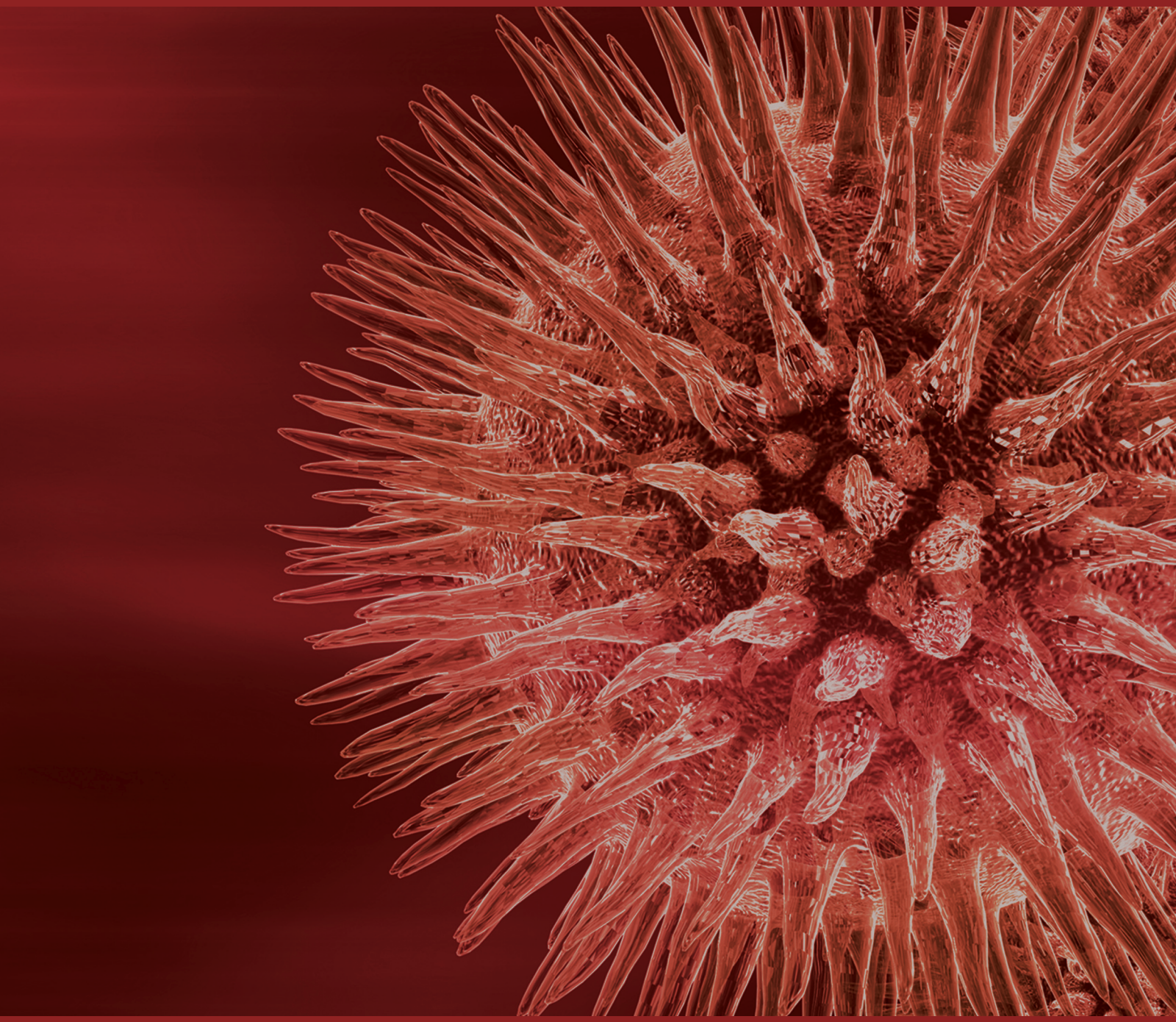
Journal of Biomedicine and Biotechnology

Computational Approaches in Metabolic Engineering

Guest Editors: Jennifer L. Reed, Ryan S. Senger,
Maciek R. Antoniewicz, and Jamey D. Young



Computational Approaches in Metabolic Engineering

Journal of Biomedicine and Biotechnology

Computational Approaches in Metabolic Engineering

Guest Editors: Jennifer L. Reed, Ryan S. Senger, Maciek R. Antoniewicz, and Jamey D. Young

Editorial Board

The editorial board of the journal is organized into sections that correspond to the subject areas covered by the journal.

Agricultural Biotechnology

Guihua H. Bai, USA
Christopher P. Chanway, Canada
Ravindra N. Chibbar, Canada
Ian Godwin, Australia
Hari B. Krishnan, USA
Carol A. Mallory-Smith, USA
Dennis P. Murr, Canada
Rodomiro Ortiz, Mexico
B. C. Saha, USA
Mariam B. Sticklen, USA
Chiu-Chung Young, Taiwan

Animal Biotechnology

E. S. Chang, USA
Hans H. Cheng, USA
Bhanu P. Chowdhary, USA
Noelle E. Cockett, USA
Peter Dovc, Slovenia
Scott C. Fahrenkrug, USA
Dorian J. Garrick, USA
Thomas A. Hoagland, USA
Tosso Leeb, Switzerland
James D. Murray, USA
Anita M. Oberbauer, USA
Jorge A. Piedrahita, USA
Daniel Pomp, USA
Kent M. Reed, USA
Lawrence Reynolds, USA
Lawrence B. Schook, USA
Mari A. Smits, The Netherlands
Leon Spicer, USA
J. Verstegen, USA
Matthew B. Wheeler, USA
Kenneth L. White, USA

Biochemistry

Robert Blumenthal, USA
David Ronald Brown, UK
Saulius Butenas, USA
Vittorio Calabrese, Italy
F. Castellino, USA
Roberta Chiaraluce, Italy
D. M. Clarke, Canada
Francesca Cutruzzolà, Italy
Paul W. Doetsch, USA
Hicham Fenniri, Canada
Nick V. Grishin, USA
J. Guy Guillemette, Canada
Paul W. Huber, USA
Chen-Hsiung Hung, Taiwan
Michael Kalafatis, USA
B. E. Kemp, Australia
Phillip E. Klebba, USA
Wen-Hwa Lee, USA
Richard D. Ludescher, USA
George Makhatadze, USA
Leonid Medved, USA
Susan A. Rotenberg, USA
Jason Shearer, USA
Andrei Surguchov, USA
John B. Vincent, USA
Yujun George Zheng, USA

Bioinformatics

T. Akutsu, Japan
Miguel A. Andrade, Germany
Mark Y. Borodovsky, USA
Rita Casadio, Italy
Artem Cherkasov, Canada
David Corne, UK
Sorin Draghici, USA
Stavros J. Hamodrakas, Greece
Paul Harrison, USA
George Karypis, USA
Jack A. Leunissen, The Netherlands
Guohui Lin, Canada
Satoru Miyano, Japan
Zoran Obradovic, USA
Florencio Pazos, Spain
Zhirong Sun, China
Ying Xu, USA
A. Zelikovsky, USA
Albert Zomaya, Australia

Biophysics

Miguel Castanho, Portugal
P. Bryant Chase, USA
Kuo-Chen Chou, USA
Rizwan Khan, India
Ali A. Khraibi, Saudi Arabia
Rumiana Koynova, USA
Serdar Kuyucak, Australia
Jianjie Ma, USA
S. B. Petersen, Denmark
Peter Schuck, USA
Claudio M. Soares, Portugal

Cell Biology

Omar Benzakour, France
Sanford I. Bernstein, USA
Phillip I. Bird, Australia
Eric Bouhassira, USA
Mohamed Boutjdir, USA
Chung-Liang Chien, Taiwan
Richard Gomer, USA
Paul J. Higgins, USA
Pavel Hozak, Czech Republic
Xudong Huang, USA
Anton M. Jetten, USA
Seamus J. Martin, Ireland
Manuela Martins-Green, USA
Shoichiro Ono, USA
George Perry, USA
M. Piacentini, Italy
George E. Plopper, USA
Lawrence Rothblum, USA
Michael Sheetz, USA
James L. Sherley, USA
G. S. Stein, USA
Richard Tucker, USA
Thomas van Groen, USA
Andre Van Wijnen, USA
Steve Winder, UK
Chuanyue Wu, USA
Bin-Xian Zhang, USA

Genetics

Adewale Adeyinka, USA
Claude Bagnis, France
J. Birchler, USA
Susan Blanton, USA
Barry J. Byrne, USA
R. Chakraborty, USA
Domenico Coviello, Italy
Sarah H. Elsea, USA
Celina Janion, Poland
J. Spencer Johnston, USA
M. Ilyas Kamboh, USA
Feige Kaplan, Canada
Manfred Kayser, The Netherlands
Brynn Levy, USA
Xiao Jiang Li, USA
Thomas Liehr, Germany
James M. Mason, USA
Mohammed Rachidi, France
Raj S. Ramesar, South Africa
Elliot D. Rosen, USA
Dharambir K. Sanghera, USA
Michael Schmid, Germany
Markus Schuelke, Germany
Wolfgang Arthur Schulz, Germany
Jorge Sequeiros, Portugal
Mouldy Sioud, Norway
Rongjia Zhou, China

Genomics

Vladimir Bajic, Saudi Arabia
Margit Burmeister, USA
Settara Chandrasekharappa, USA
Yataro Daigo, Japan
J. Spencer Johnston, USA
Vladimir Larionov, USA
Thomas Lufkin, Singapore
Joakim Lundeberg, Sweden
John L. McGregor, France
John V. Moran, USA
Yasushi Okazaki, Japan
Gopi K. Podila, USA
Momiao Xiong, USA

Immunology

Hassan Alizadeh, USA
Peter Bretscher, Canada
Robert E. Cone, USA
Terry L. Delovitch, Canada
Anthony L. DeVico, USA
Nick Di Girolamo, Australia
Don Mark Estes, USA
Soldano Ferrone, USA
Jeffrey A. Frelinger, USA
John Robert Gordon, Canada
James D. Gorham, USA
Silvia Gregori, Italy
Thomas Griffith, USA
Young S. Hahn, USA
Dorothy E. Lewis, USA
Bradley W. McIntyre, USA
R. Mosley, USA
Marija Mostarica-Stojković, Serbia
Hans Konrad Muller, Australia
Ali Ouaissi, France
Kanury V. S. Rao, India
Yair Reisner, Israel
Harry W. Schroeder, USA
Wilhelm Schwaeble, UK
Nilabh Shastri, USA
Yufang Shi, China
Piet Stinissen, Belgium
Hannes Stockinger, Austria
J. W. Tervaert, The Netherlands
Graham R. Wallace, UK

Microbial Biotechnology

Jozef Anné, Belgium
Yoav Bashan, Mexico
Marco Bazzicalupo, Italy
Nico Boon, Belgium
Luca Simone Cocolin, Italy
Peter Coloe, Australia
Daniele Daffonchio, Italy
Han de Winde, The Netherlands
Yanhe Ma, China
Bernd H. A. Rehm, New Zealand
Angela Sessitsch, Austria
Effie Tsakalidou, Greece
J. Wiegel, USA

Microbiology

D. Beighton, UK
Steven R. Blanke, USA
Stanley Brul, The Netherlands
Isaac K. O. Cann, USA
Peter Dimroth, Switzerland
Stephen K. Farrand, USA
Alain Filloux, UK
Gad Frankel, UK
Roy Gross, Germany
Hans-Peter Klenk, Germany
Tanya Parish, UK
Gopi K. Podila, USA
Frederick D. Quinn, USA
Didier A. Raoult, France
Isabel Sá-Correia, Portugal
P. L. C. Small, USA
Lori Snyder, UK
Michael Thomm, Germany
H. C. van der Mei, The Netherlands
Schwan William, USA

Molecular Biology

Rudi Beyaert, Belgium
Michael Bustin, USA
Douglas Cyr, USA
K. Iatrou, Greece
Lokesh Joshi, Ireland
David W. Litchfield, Canada
Noel F. Lowndes, Ireland
Wuyuan Lu, USA
Patrick Matthias, Switzerland
John L. McGregor, France
S. L. Mowbray, Sweden
Elena Orlova, UK
Yeon-Kyun Shin, USA
William S. Trimble, Canada
Lisa Wiesmuller, Germany
Masamitsu Yamaguchi, Japan

Oncology

Colin Cooper, UK
F. M. J. Debruyne, The Netherlands
Nathan Ames Ellis, USA
Dominic Fan, USA
Gary E. Gallick, USA
Daila S. Gridley, USA
Xin-yuan Guan, Hong Kong
Anne Hamburger, USA
Manoor Prakash Hande, Singapore
Beric Henderson, Australia
Steve B. Jiang, USA
Daehee Kang, Republic of Korea
Abdul R. Khokhar, USA
Rakesh Kumar, USA
Macus Tien Kuo, USA
Eric W. Lam, UK
Sue-Hwa Lin, USA
Kapil Mehta, USA
Orhan Nalcioglu, USA
Vincent C. O. Njar, USA
P. J. Oefner, Germany
Allal Ouhtit, USA
Frank Pajonk, USA
Waldemar Priebe, USA
F. C. Schmitt, Portugal
Sonshin Takao, Japan
Ana Maria Tari, USA
Henk G. Van Der Poel, The Netherlands
Haodong Xu, USA
David J. Yang, USA

Pharmacology

Abdel A. Abdel-Rahman, USA
M. Badr, USA
Stelvio M. Bandiera, Canada
Ronald E. Baynes, USA
R. Keith Campbell, USA
Hak-Kim Chan, Australia
Michael D. Coleman, UK
J. Descotes, France
Dobromir Dobrev, Germany
Ayman El-Kadi, Canada
Jeffrey Hughes, USA
Kazim Husain, USA
Farhad Kamali, UK
Michael Kassiou, Australia
Joseph J. McArdle, USA
Mark J. McKeage, New Zealand
Daniel T. Monaghan, USA
T. Narahashi, USA
Kennerly S. Patrick, USA
Vickram Ramkumar, USA
Michael J. Spinella, USA
Quadiri Timour, France
Todd W. Vanderah, USA
Val J. Watts, USA
David J. Waxman, USA

Plant Biotechnology

P. L. Bhalla, Australia
J. R. Botella, Australia
Elvira Gonzalez De Mejia, USA
H. M. Häggman, Finland
Liwen Jiang, Hong Kong
Pulugurtha Bharadwaja Kirti, India
Yong Pyo Lim, Republic of Korea
Gopi K. Podila, USA
Ralf Reski, Germany
Sudhir Kumar Sopory, India

Toxicology

M. Aschner, USA
Michael L. Cunningham, USA
Laurence D. Fechter, USA
Hartmut Jaeschke, USA
Youmin James Kang, USA
M. Firoze Khan, USA
Pascal Kintz, France
R. S. Tjeerdema, USA
Kenneth Turteltaub, USA
Brad Upham, USA

Virology

Contents

Computational Approaches in Metabolic Engineering, Jennifer L. Reed, Ryan S. Senger, Maciek R. Antoniewicz, and Jamey D. Young
Volume 2010, Article ID 207414, 7 pages

Dynamic Metabolic Flux Analysis Demonstrated on Cultures Where the Limiting Substrate Is Changed from Carbon to Nitrogen and *Vice Versa*, Gaspard Lequeux, Joeri Beauprez, Jo Maertens, Ellen Van Horen, Wim Soetaert, Erick Vandamme, and Peter A. Vanrolleghem
Volume 2010, Article ID 621645, 19 page

Which Metabolic Pathways Generate and Characterize the Flux Space? A Comparison among Elementary Modes, Extreme Pathways and Minimal Generators, Francisco Llaneras and Jesús Picó
Volume 2010, Article ID 753904, 13 page

Elementary Mode Analysis for the Rational Design of Efficient Succinate Conversion from Glycerol by *Escherichia coli*, Zhen Chen, Hongjuan Liu, Jianan Zhang, and Dehua Liu
Volume 2010, Article ID 518743, 14 page

Optimal Fluxes, Reaction Replaceability, and Response to Enzymopathies in the Human Red Blood Cell, A. De Martino, D. Granata, E. Marinari, C. Martelli, and V. Van Kerrebroeck
Volume 2010, Article ID 415148, 10 page

A Metabolic Model of Human Erythrocytes: Practical Application of the E-Cell Simulation Environment, Ayako Yachie-Kinoshita, Taiko Nishino, Hanae Shimo, Makoto Suematsu, and Masaru Tomita
Volume 2010, Article ID 642420, 14 page

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 207414, 7 pages
doi:10.1155/2010/207414

Editorial

Computational Approaches in Metabolic Engineering

Jennifer L. Reed,[1] Ryan S. Senger,[2] Maciek R. Antoniewicz,[3] and Jamey D. Young[4]

[1] *Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706-1607, USA*
[2] *Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA*
[3] *Department of Chemical Engineering, University of Delaware, Newark, DE 19716, USA*
[4] *Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37235-1604, USA*

Correspondence should be addressed to Jennifer L. Reed, reed@engr.wisc.edu

Received 31 December 2010; Accepted 31 December 2010

1. Introduction

Metabolic engineering involves the adjustment of metabolic and regulatory processes to improve desired cellular behaviors, such as the production of proteins and chemicals. Since cellular metabolic and regulatory networks are often large and complex, the construction and analysis of computational models of these networks can be useful for identifying current network states and evaluating the effects of network perturbations on desired phenotypes. This special issue includes papers that illustrate how computational approaches can be used in metabolic engineering. Here, we provide a brief overview of several established computational approaches that can be used to aid in the engineering of metabolic networks, while describing some of the exciting recent advances in these fields.

2. Descriptive Approaches for Identifying Intracellular Metabolic States

A variety of experimental measurements can be used to quantify the state of metabolic and regulatory networks, including flux analysis, gene expression, protein expression, metabolite concentration, enzyme activity, uptake and secretion rates, and transcription factor-DNA binding assays. Computational models and approaches can be useful for integrating and analyzing such datasets to quantify metabolic fluxes and uncover their regulatory properties.

2.1. ^{13}*C Labeling Experiments and Metabolic Flux Analysis.* The ability to quantitatively map intracellular fluxes using metabolic flux analysis (MFA) is critical for identifying pathway bottlenecks and elucidating network regulation in biological systems, especially in engineered cells with nonnative metabolic capacities [1, 2]. ^{13}C-MFA experiments involve feeding isotopically labeled substrates to cells, tissues, or whole organisms and subsequently measuring the patterns of isotope incorporation that occur in intracellular metabolites or secreted products [3]. Both mass spectrometry (MS) and nuclear magnetic resonance (NMR) can be used to quantify the relative abundance of different "isotopomers" (i.e., isotope isomers) associated with each measured biomolecule. Because different metabolic pathways often give rise to distinct isotope labeling patterns, isotopomer abundances can be used to infer the relative fluxes through these pathways [4]. As a result, isotopomer measurements obtained from MS or NMR, in combination with direct measurements of extracellular uptake or secretion rates, can be computationally analyzed to reconstruct comprehensive flux maps describing intracellular metabolism, which is the essence of MFA. A complete flux map is thus the phenotypic equivalent of the transcriptional map obtained from DNA microarrays, and tracking flux changes in response to targeted perturbations can provide important information about the distribution of kinetic and regulatory controls in metabolism [5].

The standard approach for computing a flux map involves a nonlinear least-squares regression to minimize the lack-of-fit between (i) experimentally measured and (ii) computationally simulated data. The latter are derived by solving a "forward problem", which involves solving the isotopomer and metabolite balance equations for a particular set of flux parameters in order to calculate the relative

abundance of all observable metabolite labeling patterns. Because the forward problem must be solved up to hundreds or even thousands of times to achieve an optimal fit, a great deal of effort has been placed on developing improved strategies to simulate the isotopic labeling produced by a particular network state. The Elementary Metabolite Unit (EMU) approach was developed by Antoniewicz et al. [6] to address precisely this problem. Through a novel decomposition of the isotopomer network, the algorithm systematically identifies the minimal set of variables required to simulate the available labeling measurements. The EMU approach has achieved 10-fold reductions in the system sizes required to simulate ^{13}C labeling in both medium- and large-scale networks [6, 7]. While these gains are impressive, the true power of the EMU approach is in its ability to unlock entirely new isotope labeling strategies that were previously outside the reach of computational tractability. For instance, the number of isotopomer variables required to simulate mixed ^{13}C, ^{18}O, and ^{2}H labeling in even a small-scale network of gluconeogenesis (24 reactions, 14 intracellular metabolites) climbs into the millions, while the number of EMU balances is in the mere hundreds [6].

Typically, MFA relies on the assumption of both metabolic and isotopic steady state. Achieving this situation experimentally involves (i) equilibrating the system in a stable metabolic state, (ii) introducing an isotopically labeled substrate without perturbing the metabolic steady state, (iii) allowing the system to establish a new isotopic steady state that reflects the underlying metabolic fluxes, and (iv) measuring isotopic labeling in the fully equilibrated system. Depending on the relative speed of metabolic and isotopic dynamics, however, other experimental scenarios can be envisioned. If the isotopic labeling responds quickly to any metabolic changes in the system, quasi-stationary MFA can be applied to obtain a series of snapshots that describe the variation in network fluxes over time [8, 9]. Conversely, if labeling occurs slowly but metabolism is maintained in a fixed state, isotopically nonstationary MFA (INST-MFA) can be used to estimate fluxes [10]. Finally, when measurements are obtained under both metabolically and isotopically nonstationary conditions, a fully dynamic modeling approach is required to estimate fluxes [11]. In the current issue, Lequeux et al. (G. Lequeux et al. "*Dynamic metabolic flux analysis demonstrated on cultures where the limiting substrate is changed from carbon to nitrogen and vice versa*") have taken a different dynamic MFA approach by obtaining transient measurements of 10 different extracellular metabolites during the shift of *E. coli* cells from nitrogen- to carbon-limitation, or vice versa. Numerically differentiating these extracellular measurements allowed the authors to estimate dynamically changing uptake and secretion fluxes, which provided the measurements necessary to estimate all intracellular fluxes in their model as a function of time. This has the advantage of avoiding the complications imposed by measuring and fitting transient isotope labeling data, but at the same time provides limited redundancy to validate assumptions on cofactor balancing and respiration efficiency that must be invoked in order to close the system of balance equations [12].

By applying the EMU approach, Young et al. [13] have recently developed computational routines that achieve more than 5,000-fold speedup relative to prior INST-MFA algorithms. This opens the door to several practical applications that were previously intractable due to the computational complexity of INST-MFA, where the isotopomer balances are described by differential rather than algebraic equations. INST-MFA is ideally suited to systems that label slowly due to the presence of large intermediate pools or pathway bottlenecks. This approach not only avoids the additional time and cost of feeding isotope tracers over extended periods [14], but may become absolutely necessary in cases where the system cannot be held in a fixed metabolic state long enough to allow isotopic labeling to fully equilibrate. As a result, INST-MFA is expected to become an indispensible tool for extending MFA approaches to studies of mammalian systems [15–17], industrial bioprocesses [10, 18], and other scenarios where obtaining a strict isotopic steady state may be impractical. Another emerging application of INST-MFA is its application to photoautotrophic metabolism, which is the process by which plants, algae, and cyanobacteria use light energy to fix carbon dioxide into complex organic molecules. Because photoautotrophs assimilate carbon solely from CO_2, feeding $^{13}CO_2$ will produce a uniform steady-state labeling pattern that is insensitive to fluxes. Thus, conventional steady-state ^{13}C-MFA is incapable of quantifying autotrophic metabolic fluxes [19]. However, transient measurements of isotope incorporation following a step change from unlabeled to labeled CO_2 can be used to estimate fluxes by applying INST-MFA [20]. This approach enables comprehensive flux analysis of photoautotrophic metabolism, complementing previous ^{13}C-MFA studies of plants [21] and cyanobacteria [22] that were limited to heterotrophic or mixotrophic culture conditions, with sugar as the major carbon source. Taken together, these advances illustrate how combined progress in both analytical capabilities and computational techniques are driving MFA applications toward larger, more dynamic, and more complex biochemical networks, which encompass a growing variety of plant, animal, and microbial systems.

2.2. Using Metabolite and Gene/Protein Expression Measurements. In addition to ^{13}C-MFA experiments, other large-scale measurements can be made which capture cellular metabolic states. These include measurements of metabolite concentrations and gene or protein expression. Metabolite concentrations can now be quantified using mass spectroscopy for hundreds of metabolites in a single condition. For example, Bennett and colleagues used LC-MS/MS to quantify over 100 metabolite concentrations in *E. coli* cells grown under three different conditions [23]. These metabolite concentrations can be analyzed to identify potential metabolic bottlenecks by evaluating enzyme saturation and estimating Gibbs-free energy changes of reactions. Bennett and colleagues compared measured metabolite concentrations to reported Michaelis-Menten kinetic parameters (K_m) to determine whether individual reaction rates are substrate (where [substrate] $\ll$ K_m) or enzyme limited (where

[substrate] $\gg K_m$). Their analysis found that most substrate concentrations were higher than the reported K_m values (83%) indicating that for many reactions the rates are limited by enzyme levels [23]. Metabolite concentrations can also be used to estimate thermodynamic properties of metabolic reactions since the change in Gibbs-free energy for a reaction ($\Delta_R G'$) is dependent on substrate and product concentrations [24, 25]. The $\Delta_R G'$ values can be estimated from measured metabolite concentrations and then used to distinguish between those reactions that are operating close to equilibrium ($\Delta_R G' \approx 0$), and those reactions that are far from equilibrium ($\Delta_R G' \ll 0$) whose rates may be limited by regulation via enzyme kinetics [25]. Thermodynamic analysis using metabolite concentrations has been preformed for both *E. coli* and *S. cerevisiae* to identify these types of reactions [23, 25, 26]. Together the analysis of metabolite concentrations using both kinetic (if available) and thermodynamic information can be useful to identify metabolic bottlenecks or rate limiting reactions.

In addition to metabolite concentrations, gene expression and protein expression measurements can also be used to help elucidate metabolic fluxes and their regulation. Recent modeling efforts have used these types of measurements to improve predictions of fluxes through metabolic reactions [5, 27–29]. With these approaches, gene expression data is used to place restrictions on flux values or flux changes. The GIMME method uses an expression threshold and prevents flux through reactions associated with genes whose expression is below the threshold. In this case, an inconsistency score is minimized, where penalties depend on the magnitude of the flux and how far the expression is below the threshold [27]. Another method, proposed by Shlomi and colleagues, instead groups reactions into high, medium, and low sets based on expression levels of associated genes [29]. A flux distribution is then identified that has flux through as many reactions in the high set as possible, and no flux through as many reactions as possible in the low set. In the third method (E-flux), the relative expression levels of genes from a given condition are used to place constraints on the upper limits flux values can take [28]. Another approach, developed by Moxley and colleagues, instead uses the changes in expression level between two conditions to estimate the changes in fluxes between two conditions [5]. In this case, the predicted flux change depends on the expression change (ΔmRNA) and some additional model parameters. All of these methods have been applied to gene expression measurements, but they likely can also be used with quantitative proteomics measurements.

3. Predictive Approaches for Improving Cellular Phenotypes

The modeling approaches described in the previous sections can provide descriptions of the metabolic fluxes by analyzing different types of experimental measurements. Once these metabolic states are known, other modeling approaches can be used to identify which environmental and/or genetic perturbations would improve cellular phenotypes, such as the production of desired chemicals. These predictive approaches include pathway-based and optimization-based methods, that take into account the structure and stoichiometry of metabolic networks, as well as, kinetic modeling approaches that also account for enzyme kinetics.

3.1. Pathway-Based Approaches. Identification of the relevant pathways of a metabolic network is essential for finding effective metabolic engineering strategies. These pathways can also help derive minimal media requirements for an organism and assess the robustness and redundancy of key metabolic pathways. In this special issue, F. Llaneras and J. Picó review and compare four established methods used to identify relevant metabolic pathways: extreme currents, elementary modes, extreme pathways, and minimal generators (F. Llaneras and J. Picó, "*Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators*"). The authors recommend elementary modes for determining the metabolic impact of gene knockouts and the required pathways of a network for producing desired chemicals. The calculation of elementary modes requires knowledge of reaction stoichiometry and reversibility, making the recent modeling advances in biochemical reaction thermodynamics important to this analysis. Several algorithms and software packages have been developed for calculating the elementary modes of a metabolic network, including Metatool [30, 31], FluxAnalyzer [32], and functionalities built into OptFlux [33]. However, the problem of combinatorial explosion when calculating elementary modes for large complex metabolic networks has been documented [34]. This has paved the way for clustering algorithms such as the Agglomeration of Common Motifs (ACoM) [35] in an attempt to give biological meaning to elementary modes and define relatedness between reactions. Application of elementary modes to rational metabolic engineering approaches has now enabled strain design algorithms such as those used by OptFlux [33] and the genetic algorithm-based method compiled by Boghigian et al. [36]. Elementary flux modes have been successfully used to engineer strains with a variety of desired phenotypes, including sugar coutilization [37], ethanol [37, 38], and carotenoid production [39]. Z. Chen et al., in this issue, use elementary modes to find strategies for improving the conversion of glycerol into succinate by considering the effects of oxygen utilization and genetic alterations (Z. Chen et al., "*Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by Escherichia coli*"). Random sampling of flux distributions provides another way of investigating possible flux distributions through metabolic networks [40, 41], and A. De Martino et al. in this issue use both structural analysis and sampling to explore the robustness of human red blood cell (RBC) metabolism (A. De Martino et al., "*Optimal fluxes, reaction replaceability, and response to enzymopathies in the human red blood cell*").

3.2. Optimization-Based Approaches. Alternatives to pathway-based approaches include optimization-based methods, which can also identify mutations that would improve

desired phenotypes (e.g., increased production yields). Here, the models use the same stoichiometric and thermodynamic constraints as those used in pathway-based approaches, but solutions are identified which maximize or minimize a stated objective. To predict how metabolic fluxes will change in response to a genetic perturbation, a number of different approaches have been developed, including flux balance analysis (FBA, reviewed in [42]), minimization of metabolic adjustment (MOMA) [43], regulatory on/off mechanism (ROOM) [44], regulated flux balance analysis (rFBA and SR-FBA) [45, 46], and probabilistic regulation of metabolism (PROM) [47]. Most of these approaches are used to predict the immediate behavior of knockout strains [43, 44], with the exception being FBA, whose predictions more closely resemble strain behavior after cells have undergone adaptive evolution [48]. MOMA has been successfully used to engineer strains with increased production of a variety of products including lycopene [49, 50], valine [51], threonine [52], and polylactic acid [53].

To identify those genetic strategies that are predicted to have the highest chemical production levels, a large number of possible strategies must be considered. Bilevel approaches can be used to identify which are the best production strategies given a maximum number of genetic alterations without having to generate and store predictions for all possible strategies. These bilevel methods can be solved using integer programming and/or genetic algorithms [54–57]. A variety of such bilevel methods have been proposed which use FBA, SR-FBA, and MOMA to predict mutant strain behaviors, and these approaches can consider genetic changes involving gene deletions, altered gene expression, or altered transcriptional regulation [54, 56, 58–61]. These methods have been used to design strains for a variety of chemicals [55, 62–64]. A more recent bilevel approach (OptFORCE) uses optimization to identify how metabolic fluxes must change to improve metabolite production, which is independent of any assumptions about what functions are used to predict cellular behavior [60].

3.3. Kinetic Modeling Approaches. Both types of models and methods described in the last two sections do not take enzyme kinetics into account. As a result, they are unable to predict how changes in kinetic properties, enzyme, and metabolite concentrations would affect fluxes through metabolic pathways. Kinetic models are needed to make these types of predictions since they capture the dependence of fluxes on metabolite and enzyme concentrations. These types of models can be analyzed to identify which changes are needed for improving cellular phenotypes. The classical framework for elucidating parameters responsible for the control of metabolic fluxes is metabolic control analysis (MCA), developed in the early 1970s independently by Kacser and Burns [65] and Heinrich and Rapoport [66]. Recently, Visser et al. developed an alternative approach called linlog kinetics [67, 68]. Here, all rate equations are modeled with the same basic mathematical structure in which the relationship between rates and enzyme levels is linear, while for metabolite levels, a linear combination of logarithmic terms is used. Young et al. [69] have recently expanded the cybernetic modeling framework of Ramkrishna [70] to incorporate metabolic pathway concepts derived from elementary mode analysis. This led to models that could predict both local and global control properties of metabolic networks in response to either dynamic environmental shifts or stable genetic manipulations. These models were applied to predict phenotypes of several recombinant *E. coli* strains, and they were found to provide good agreement with experimental data. Furthermore, because of the dynamic nature of these models, they were capable of simulating responses that are not readily addressed by purely stoichiometric models (e.g., allosteric or kinetic effects of intermediate metabolites, enzyme overexpressions and partial knockdowns, and time-dependent culture conditions). In this special issue, A. Yachie-Kinoshita et al. review the history of kinetic models for human red blood cells (RBCs) and describe an RBC metabolic model implemented in the E-cell simulation environment (A. Yachie-Kinoshita et al., "*A metabolic model of human erythrocytes: practical application of the E-Cell Simulation Environment*"). They discuss how this E-cell RBC model can be applied to predict RBC responses to hypoxic environments and long-term cold storage and identify enzymes whose altered activity could improve storage conditions for RBCs.

4. Concluding Remarks

A growing number of computational tools that facilitate the evaluation and improvement of strains for metabolic engineering are currently being developed and expanded. These methods can account for a wide range of experimental measurements to provide an improved understanding of metabolic states and current limitations, and they can be used to identify new engineering strategies for improved chemical production. The collection of papers in this special issue highlight several recent advances and underscore the emerging applications of these computational tools.

Jennifer L. Reed
Ryan S. Senger
Maciek R. Antoniewicz
Jamey D. Young

References

[1] U. Sauer, "Metabolic networks in motion: ^{13}C-based flux analysis," *Molecular Systems Biology*, vol. 2, article 62, 2006.

[2] J. Nielsen, "It is all about metabolic fluxes," *Journal of Bacteriology*, vol. 185, no. 24, pp. 7031–7035, 2003.

[3] W. Wiechert, "C metabolic flux analysis," *Metabolic Engineering*, vol. 3, no. 3, pp. 195–206, 2001.

[4] W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, and A. A. De Graaf, "Bidirectional reaction steps in metabolic networks—III. Explicit solution and analysis of isotopomer labeling systems," *Biotechnology and Bioengineering*, vol. 66, no. 2, pp. 69–85, 1999.

[5] J. F. Moxley, M. C. Jewett, M. R. Antoniewicz et al., "Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator

Gcn4p," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6477–6482, 2009.

[6] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions," *Metabolic Engineering*, vol. 9, no. 1, pp. 68–86, 2007.

[7] P. F. Suthers, Y. J. Chang, and C. D. Maranas, "Improved computational performance of MFA using elementary metabolite units and flux coupling," *Metabolic Engineering*, vol. 12, no. 2, pp. 123–128, 2010.

[8] M. R. Antoniewicz, D. F. Kraynie, L. A. Laffend, J. González-Lergier, J. K. Kelleher, and G. Stephanopoulos, "Metabolic flux analysis in a nonstationary systemml: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol," *Metabolic Engineering*, vol. 9, no. 3, pp. 277–292, 2007.

[9] Y. Noguchi, J. D. Young, J. O. Aleman, M. E. Hansen, J. K. Kelleher, and G. Stephanopoulos, "Effect of anaplerotic fluxes and amino acid availability on hepatic lipoapoptosis," *Journal of Biological Chemistry*, vol. 284, no. 48, pp. 33425–33436, 2009.

[10] K. Nöh, K. Grönke, B. Luo, R. Takors, M. Oldiges, and W. Wiechert, "Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments," *Journal of Biotechnology*, vol. 129, no. 2, pp. 249–267, 2007.

[11] S. A. Wahl, K. Nöh, and W. Wiechert, "^{13}C labeling experiments at metabolic nonstationary conditions: an exploratory study," *BMC Bioinformatics*, vol. 9, article 152, 2008.

[12] K. Schmidt, A. Marx, A. A. De Graaf et al., "^{13}C tracer experiments and metabolite balancing for metabolic flux analysis: comparing two approaches," *Biotechnology and Bioengineering*, vol. 58, no. 2-3, pp. 254–257, 1998.

[13] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, "An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis," *Biotechnology and Bioengineering*, vol. 99, no. 3, pp. 686–699, 2008.

[14] Z. Zhao, K. Kuijvenhoven, C. Ras et al., "Isotopic non-stationary ^{13}C gluconate tracer method for accurate determination of the pentose phosphate pathway split-ratio in Penicillium chrysogenum," *Metabolic Engineering*, vol. 10, no. 3-4, pp. 178–186, 2008.

[15] K. Maier, U. Hofmann, M. Reuss, and K. Mauch, "Identification of metabolic fluxes in hepatic cells from transient C-labeling experiments—part II. Flux estimation," *Biotechnology and Bioengineering*, vol. 100, no. 2, pp. 355–370, 2008.

[16] J. Munger, B. D. Bennett, A. Parikh et al., "Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy," *Nature Biotechnology*, vol. 26, no. 10, pp. 1179–1186, 2008.

[17] K. Maier, U. Hofmann, A. Bauer et al., "Quantification of statin effects on hepatic cholesterol synthesis by transient ^{13}C-flux analysis," *Metabolic Engineering*, vol. 11, no. 4-5, pp. 292–309, 2009.

[18] S. Iwatani, Y. Yamada, and Y. Usuda, "Metabolic flux analysis in biotechnology processes," *Biotechnology Letters*, vol. 30, no. 5, pp. 791–799, 2008.

[19] A. A. Shastri and J. A. Morgan, "A transient isotopic labeling methodology for C metabolic flux analysis of photoautotrophic microorganisms," *Phytochemistry*, vol. 68, no. 16-18, pp. 2302–2312, 2007.

[20] W. Wiechert and K. Nöh, "From stationary to instationary metabolic flux analysis," *Advances in Biochemical Engineering/Biotechnology*, vol. 92, pp. 145–172, 2005.

[21] J. Schwender, "Metabolic flux analysis as a tool in metabolic engineering of plants," *Current Opinion in Biotechnology*, vol. 19, no. 2, pp. 131–137, 2008.

[22] C. Yang, Q. Hua, and K. Shimizu, "Metabolic flux analysis in Synechocystis using isotope distribution from C-labeled glucose," *Metabolic Engineering*, vol. 4, no. 3, pp. 202–216, 2002.

[23] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, "Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*," *Nature Chemical Biology*, vol. 5, no. 8, pp. 593–599, 2009.

[24] C. S. Henry, M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis, "Genome-scale thermodynamic analysis of *Escherichia coli* metabolism," *Biophysical Journal*, vol. 90, no. 4, pp. 1453–1461, 2006.

[25] A. Kümmel, S. Panke, and M. Heinemann, "Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data," *Molecular Systems Biology*, vol. 2, article 2006.0034, 2006.

[26] M. Klimacek, S. Krahulec, U. Sauer, and B. Nidetzky, "Limitations in xylose-fermenting saccharomyces cerevisiae, made evident through comprehensive metabolite profiling and thermodynamic analysis," *Applied and Environmental Microbiology*, vol. 76, no. 22, pp. 7566–7574, 2010.

[27] S. A. Becker and B. O. Palsson, "Context-specific metabolic networks are consistent with experiments," *PLoS Computational Biology*, vol. 4, no. 5, Article ID e1000082, 2008.

[28] C. Colijn, A. Brandes, J. Zucker et al., "Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production," *PLoS Computational Biology*, vol. 5, no. 8, article e1000489, 2009.

[29] T. Shlomi, M. N. Cabili, M. J. Herrgård, B. Ø. Palsson, and E. Ruppin, "Network-based prediction of human tissue-specific metabolism," *Nature Biotechnology*, vol. 26, no. 9, pp. 1003–1010, 2008.

[30] A. von Kamp and S. Schuster, "Metatool 5.0: fast and flexible elementary modes analysis," *Bioinformatics*, vol. 22, no. 15, pp. 1930–1931, 2006.

[31] S. Klamt, J. Gagneur, and A. von Kamp, "Algorithmic approaches for computing elementary modes in large biochemical reaction networks," *IEE Proceedings Systems Biology*, vol. 152, no. 4, pp. 249–255, 2005.

[32] S. Klamt, J. Stelling, M. Ginkel, and E. D. Gilles, "FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps," *Bioinformatics*, vol. 19, no. 2, pp. 261–269, 2003.

[33] I. Rocha, P. Maia, P. Evangelista et al., "OptFlux: an open-source software platform for in silico metabolic engineering," *BMC Systems Biology*, vol. 4, article 45, 2010.

[34] S. Klamt and J. Stelling, "Combinatorial complexity of pathway analysis in metabolic networks," *Molecular Biology Reports*, vol. 29, no. 1-2, pp. 233–236, 2002.

[35] S. Pérès, F. Vallée, M. Beurton-Aimar, and J. P. Mazat, "AComml: a classification method for elementary flux modes based on motif finding," *BioSystems*, vol. 103, no. 3, pp. 410–419, 2011.

[36] B. A. Boghigian, H. Shi, K. Lee, and B. A. Pfeifer, "Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design," *BMC Systems Biology*, vol. 4, article 49, 2010.

[37] C. T. Trinh, P. Unrean, and F. Srienc, "Minimal Eschenchia coli cell for the most efficient production of ethanol from hexoses

and pentoses," *Applied and Environmental Microbiology*, vol. 74, no. 12, pp. 3634–3643, 2008.

[38] C. T. Trinh and F. Srienc, "Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol," *Applied and Environmental Microbiology*, vol. 75, no. 21, pp. 6696–6705, 2009.

[39] P. Unrean, C. T. Trinh, and F. Srienc, "Rational design and construction of an efficient E. coli for production of diapolycopendioic acid," *Metabolic Engineering*, vol. 12, no. 2, pp. 112–122, 2010.

[40] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A. L. Barabási, "Global organization of metabolic fluxes in the bacterium *Escherichia coli*," *Nature*, vol. 427, no. 6977, pp. 839–843, 2004.

[41] I. Thiele, N. D. Price, T. D. Vo, and B. Ø. Palsson, "Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet," *Journal of Biological Chemistry*, vol. 280, no. 12, pp. 11683–11695, 2005.

[42] N. D. Price, J. L. Reed, and B. Ø. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints," *Nature Reviews Microbiology*, vol. 2, no. 11, pp. 886–897, 2004.

[43] D. Segrè, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, pp. 15112–15117, 2002.

[44] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, pp. 7695–7700, 2005.

[45] M. W. Covert and B. Ø. Palsson, "Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*," *Journal of Biological Chemistry*, vol. 277, no. 31, pp. 28058–28064, 2002.

[46] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, "A genome-scale computational study of the interplay between transcriptional regulation and metabolism," *Molecular Systems Biology*, vol. 3, article 101, 2007.

[47] S. Chandrasekaran and N. D. Price, "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 41, pp. 17845–17850, 2010.

[48] S. S. Fong and B. Ø. Palsson, "Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes," *Nature Genetics*, vol. 36, no. 10, pp. 1056–1058, 2004.

[49] H. Alper, Y. S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*," *Metabolic Engineering*, vol. 7, no. 3, pp. 155–164, 2005.

[50] H. Alper, K. Miyaoku, and G. Stephanopoulos, "Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets," *Nature Biotechnology*, vol. 23, no. 5, pp. 612–616, 2005.

[51] J. H. Park, K. H. Lee, T. Y. Kim, and S. Y. Lee, "Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 19, pp. 7797–7802, 2007.

[52] K. H. Lee, J. H. Park, T. Y. Kim, H. U. Kim, and S. Y. Lee, "Systems metabolic engineering of *Escherichia coli* for L-threonine production," *Molecular Systems Biology*, vol. 3, article 149, 2007.

[53] Y. K. Jung, T. Y. Kim, S. J. Park, and S. Y. Lee, "Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers," *Biotechnology and Bioengineering*, vol. 105, no. 1, pp. 161–171, 2010.

[54] A. P. Burgard, P. Pharkya, and C. D. Maranas, "OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering*, vol. 84, no. 6, pp. 647–657, 2003.

[55] J. A. Chemler, Z. L. Fowler, K. P. McHugh, and M. A. G. Koffas, "Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering," *Metabolic Engineering*, vol. 12, no. 2, pp. 96–104, 2010.

[56] K. R. Patil, I. Rocha, J. Förster, and J. Nielsen, "Evolutionary programming as a platform for in silico metabolic engineering," *BMC Bioinformatics*, vol. 6, article 308, 2005.

[57] D. S. Lun, G. Rockwell, N. J. Guido et al., "Large-scale identification of genetic design strategies using local search," *Molecular Systems Biology*, vol. 5, article 296, 2009.

[58] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: a computational framework for redesign of microbial production systems," *Genome Research*, vol. 14, no. 11, pp. 2367–2376, 2004.

[59] P. Pharkya and C. D. Maranas, "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems," *Metabolic Engineering*, vol. 8, no. 1, pp. 1–13, 2006.

[60] S. Ranganathan, P. F. Suthers, and C. D. Maranas, "OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions," *PLoS Computational Biology*, vol. 6, no. 4, article e1000744, 2010.

[61] J. Kim and J. L. Reed, "OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains," *BMC Systems Biology*, vol. 4, article 53, 2010.

[62] M. A. Asadollahi, J. Maury, K. R. Patil, M. Schalk, A. Clark, and J. Nielsen, "Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering," *Metabolic Engineering*, vol. 11, no. 6, pp. 328–334, 2009.

[63] S. S. Fong, A. P. Burgard, C. D. Herring et al., "In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid," *Biotechnology and Bioengineering*, vol. 91, no. 5, pp. 643–648, 2005.

[64] A. R. Brochado, C. Matos, B. L. Møller, J. Hansen, U. H. Mortensen, and K. R. Patil, "Improved vanillin production in baker's yeast through in silico design," *Microbial Cell Factories*, vol. 9, article 84, 2010.

[65] H. Kacser and J. A. Burns, "The control of flux," *Symposia of the Society for Experimental Biology*, vol. 27, pp. 65–104, 1973.

[66] R. Heinrich and T. A. Rapoport, "A linear steady-state treatment of enzymatic chains. General properties, control and effector strength," *European Journal of Biochemistry*, vol. 42, no. 1, pp. 89–95, 1974.

[67] D. Visser and J. J. Heijnen, "Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics," *Metabolic Engineering*, vol. 5, no. 3, pp. 164–176, 2003.

[68] D. Visser, J. W. Schmid, K. Mauch, M. Reuss, and J. J. Heijnen, "Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics," *Metabolic Engineering*, vol. 6, no. 4, pp. 378–390, 2004.

[69] J. D. Young, K. L. Henne, J. A. Morgan, A. E. Konopka, and D. Ramkrishna, "Integrating cybernetic modeling with pathway

analysis provides a dynamic, systems-level description of metabolic control," *Biotechnology and Bioengineering*, vol. 100, no. 3, pp. 542–559, 2008.

[70] D. Ramkrishna, "A cybernetic perspective of microbial growth," in *Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems*, pp. 161–178, American Chemical Society, Washington, DC, USA, 1982.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 621645, 19 pages
doi:10.1155/2010/621645

Research Article

Dynamic Metabolic Flux Analysis Demonstrated on Cultures Where the Limiting Substrate Is Changed from Carbon to Nitrogen and *Vice Versa*

Gaspard Lequeux,[1] Joeri Beauprez,[2] Jo Maertens,[1] Ellen Van Horen,[2] Wim Soetaert,[2] Erick Vandamme,[2] and Peter A. Vanrolleghem[3]

[1] *Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, 9000 Gent, Belgium*
[2] *Laboratory of Industrial Microbiology and Biocatalysis, Ghent University, Coupure Links 653, 9000 Gent, Belgium*
[3] *Département de Génie Civil et Génie des Eaux, Université Laval, Québec, QC, Canada G1V OA6*

Correspondence should be addressed to Gaspard Lequeux, gaspard.lequeux@biomath.ugent.be

Received 23 November 2009; Revised 10 April 2010; Accepted 24 June 2010

Academic Editor: Maciek Antoniewicz

The main requirement for metabolic flux analysis (MFA) is that the cells are in a pseudo-steady state, that there is no accumulation or depletion of intracellular metabolites. In the past, the applications of MFA were limited to the analysis of continuous cultures. This contribution introduces the concept of dynamic MFA and extends MFA so that it is applicable to transient cultures. Time series of concentration measurements are transformed into flux values. This transformation involves differentiation, which typically increases the noisiness of the data. Therefore, a noise-reducing step is needed. In this work, polynomial smoothing was used. As a test case, dynamic MFA is applied on *Escherichia coli* cultivations shifting from carbon limitation to nitrogen limitation and vice versa. After switching the limiting substrate from N to C, a lag phase was observed accompanied with an increase in maintenance energy requirement. This lag phase did not occur in the C- to N-limitation case.

1. Introduction

Material balances of reactors describe how certain compounds go in the reactor and are transformed into other compounds. Metabolic models consider the cell as the reactor and, additionally to the exchange rates, also include information on the different steps (reactions) on how the input compounds are transformed to the output compounds. This should allow to gain insight in the internal cellular fluxes. Typically the number of equations (material balances of the different compounds) obtained is less than the number of unknowns (exchange rates and reaction rates) one wants to calculate.

Techniques for solving metabolic models can be divided into two main groups: data-driven techniques and knowledge/assumption-driven techniques.

The prime example of a data-driven technique is metabolic flux analysis (MFA) where intracellular fluxes are calculated based on measured exchange rates. The mathematical technique used was initially developed for black box modeling [1, 2], but can easily be applied to stoichiometric models [3, 4]. Exchange rates are not always sufficient to solve a stoichiometric model. In such cases, the option exists to measure intracellular fluxes via ^{13}C enrichment analysis [5].

The counterpart of MFA is flux balance analysis (FBA) which uses linear optimisation to solve the stoichiometric model [6] and does not directly use measurements (although measurements can be included [7]). Whereas FBA typically yields one optimal solution, the calculation of elementary modes or extreme pathways gives a full view of the metabolic capabilities of an organism [8, 9]. An overview of the different extensions applied to MFA and FBA is given in [10].

Both MFA and FBA have the fundamental assumption that the cell is in a pseudo-steady state, thus that there is no accumulation or depletion of intracellular metabolites.

This implies that metabolic models can only be used for cells in pseudo-steady state. Models that fully describe the cellular dynamics exist, but are complex and cumbersome to implement and/or very limited in scope [11, 12].

However, some examples of FBA applied in nonstationary cases can be found. Varma and Palsson [13] used flux balance analysis to describe fed batches by iteratively solving the model for maximal biomass production. At each time point, the available substrate is calculated from the results of the FBA in the previous step. This way the time profiles of cell density, glucose, and by-products could be quantitatively predicted. This concept of dynamic FBA (DFBA) is further developed by Mahadevan et al. [14]. They formalise the methodology of Varma and Palsson [13] and name it static optimisation-based DFBA. They also introduce a new methodology, called dynamic optimisation-based DFBA, in which the optimisation is done over the entire time period of interest to obtain time profiles of fluxes. The methodology of dynamic optimisation-based DFBA is combined with the concept of minimisation of metabolic adjustment (MOMA) to model the myocardial energy metabolism [15]. Minimisation of metabolic adjustment (MOMA) is an alternative goal function to better predict the behaviour of mutated strains [16]. Using this technique, it was shown that the cellular metabolism does not always remain optimal during transient perturbations [16]. Lee et al. [17] introduced integrated dynamics FBA (idFBA) that dynamically simulates cellular phenotypes arising from integrated networks. These networks combine metabolic models with genetic, regulatory, and intracellular signalling information.

In this contribution, not FBA but MFA is extended so that it can be used for modelling cultures during transient conditions. This is reasonable as intracellular pseudo-steady state can also be assumed under certain dynamic conditions, because of the relatively small time constants of cellular processes, for example, mass action and metabolic adaption to novel conditions, in comparison with processes affecting the observed environmental conditions. Indeed, perturbation experiments performed in *E. coli* [11, 18–20] showed that intracellular metabolite pools reach a new pseudo-steady state after 20 seconds. Hence, an intracellular pseudo-steady state can be assumed during the transient.

Pseudo-steady state conditions are typically encountered in chemostat operated cultures [21, 22]. Chemostats are mostly run under the concept that only one substrate is limiting. However, conditions do exist where multiple substrates can be limiting at the same time. Practical applications of such dual limitations have been described by Zinn et al. [23]. Egli [24] has described concentration zones in which both glucose and nitrogen are limiting. In this contribution, however, the limiting compound in the feed medium of steady state cultures is switched from glucose to ammonia and *vice versa*. No abrupt changes are thus sensed by the cells (the reactor broth is only gradually replaced), an ideal test case for applying dynamic FBA. The exchange fluxes of metabolites are determined based on their concentration in the reactor broth and are subsequently used to solve an overdetermined metabolic model, resulting in the determination of the intracellular flux values during the transient period. As exchange fluxes are based on the derivative of the concentration data, the latter ones are first smoothed to avoid amplification of the noise in the exchange fluxes.

2. Materials and Methods

2.1. Bacterial Strain. *Escherichia coli* MG1655 [λ^-, F^-, *rph*–1, (del *fnr*)] was obtained from the Netherlands Culture Collection of Bacteria (NCCB, Utrecht, The Netherlands).

2.2. Culture Conditions

2.2.1. Media. Composition of the Luria Bertani Broth (LB) medium and shake flask medium (for the preculture) can be found in [25].

The minimal medium used in the reactor consisted of 2.5 g/L (N-limited medium) or 5 g/L (C-limited medium) $(NH_4)_2SO_4$, 2 g/L KH_2PO_4, 0.5 g/L NaCl, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 33 g/L (N-limited medium) or 16.5 g/L (C-limited medium) Glucose $\cdot H_2O$, 1 mL/L vitamin stock solution and 100 μL/L molybdate stock solution. Vitamin stock solution consisted of 3.6 g/L $FeCl_2 \cdot 4H_2O$, 5 g/L $CaCl_2 \cdot 2H_2O$, 1.3 g/L $MnCl_2 \cdot 2H_2O$, 0.38 g/L $CuCl_2 \cdot 2H_2O$, 0.5 g/L $CoCl_2 \cdot 6H_2O$, 0.94 g/L $ZnCl_2$, 0.0311 g/L H_3BO_4, 0.4 g/L $Na_2EDTA \cdot 2H_2O$, and 1.01 g/L thiamine $\cdot$ HCl. The molybdate stock solution contained 0.967 g/L $Na_2MoO_4 \cdot 2H_2O$. All components were dissolved and filter-sterilised (pore size 0.22 μm, Sartobran, Sartorius, Belgium). The pH was left at approximately 5.4.

2.2.2. Cultivations. Culture conditions and inoculation procedure are described in [25].

The experiment in which the limiting substrate was changed from glucose to ammonia was conducted at a dilution rate of 0.155 h^{-1}. The experiment in which the limitation was changed from nitrogen to carbon was conducted at a dilution rate of 0.142 h^{-1}.

For each experiment, steady state samples were taken: optical density (OD), cell dry weight (CDW), and HPLC analyses. These samples were taken after waiting at least five residence times after the batch phase. To ensure that the perturbations caused by sampling did not influence steady state, switching the limiting nutrient was only done after another five residence times. The transition between the two limitations was frequently sampled for OD and metabolite data (HPLC). Sampling of the second steady state was performed again after five residence times. Furthermore, during the culture, dissolved oxygen, pH, stirrer rate, temperature, airflow and oxygen and carbon dioxide content of the off-gas were continuously logged. Two balances, one under the influent and one under the effluent barrel, were coupled to the system to allow precise measurement of the dilution rate.

2.3. Measurements. The sampling and measurement methodologies used are described in [25].

2.4. Data Analysis. Due to the amount of data that had to be combined for each experiment, a custom-made program was written in Python using the SciPy scientific library [26, 27]

to process the data and apply the algorithm as explained in Section 3 (polynomial fitting, extracting the derivatives, and calculating the fluxes).

Each point in the time series of the transient experiments is only measured once. This is deemed sufficient, as nonsystematical variations are smoothed out when approximating the data with polynomials (as described in Section 3.2). In this polynomial approximation, error propagation calculations are not performed. For the MFA calculations, however, a covariance matrix is needed. This matrix is used to rescale the measurements so that fluxes that are known to be inaccurate do not weight as much during the data reconciliation as other, more trustworthy fluxes [1]. Therefore, a "mean" covariance matrix was calculated from data presented in [25] and used for the MFA part of this analysis. The experiments described therein are similar to those performed for this publication (also using the same equipment) and it is assumed that the utilised covariance matrix is a good approximation of the real one. The standard deviation flags in the figures are based on this covariance matrix.

The MFA modelling was performed as described in [4], although error analysis on the reconciliated fluxes was not done. The model used is based on the one in [4] and is the same metabolic model as described in [25]. It contains 136 reactions (Table 1) and 150 metabolites (Table 2), of which 12 are exchangeable. There are 142 independent equations and $136+12 = 148$ unknowns. Thus at least 6 measurements have to be performed in order to fully solve the model. Ten exchange metabolites were measured and used for solving the stoichiometric models (CO_2, O_2, NH_3, PiOH, acetate, lactate, pyruvate, succinate, glucose, and biomass) giving 4 redundant measurements that can be balanced. The methods for balancing measurements were initially developed for black box models [1, 2, 28], but can easily be applied on stoichiometric models [3, 4].

Although it is known that the biomass composition is different under C- and N-limitation [29, 30], the same model was used in both cases thus implying the same biomass composition for cells grown under N- and C-limitation. Limited testing showed that the influence of a different protein content of the biomass on the flux fluctuations was minimal.

3. Dynamic Metabolic Flux Analysis

3.1. DMFA: The Concept. In classical metabolic flux analysis, the measured fluxes are obtained by quantification of the metabolite concentrations in the reactor broth of a chemostat culture. But what should be done under transient conditions? Measurements of external metabolites cannot directly be used for fluxes. Provost and Bastin [31] applied MFA to the batch phase of the culture of Chinese Hamster Ovary (CHO) cells. The time series of the measured metabolite concentrations are approximated with a straight line, the slope of which is a direct measure of the fluxes entering and leaving the cells and is used to calculate internal reaction rates via MFA.

This approach of Provost and Bastin [31] can be formalised and extended. Instead of limiting oneself to cases where a straight line is sufficient to adequately capture the dynamics of the measured metabolites during the transients, one can take the derivative in each point, without approximating the whole interval with a straight line. This derivative is, by definition, a flux.

The differential equation that governs the change of a component in the reactor broth can be written as

$$\frac{dC}{dt} = D(C_{in} - C) + r_p, \tag{1}$$

where C is the concentration of the component, C_{in} the concentration of the component in the influent, r_p the production rate, and D the dilution rate.

From (1) the net reaction rate r_p can be calculated using the following approximation for the derivative in a point n of the time series:

$$\left.\frac{dC}{dt}\right|_n = \frac{C_n - C_{n-1}}{t_n - t_{n-1}}. \tag{2}$$

Eventually, more data around the working point might be used to perform the linear regression. However, because of the noisiness of the data, this approach does not give satisfying results. Therefore, approximation techniques have to be used. In this work, a polynomial approximation was chosen.

3.2. Polynomial Fitting. The polynomial fitting algorithms found in SciPy [26] were used to approximate the data. The whole time series were covered by different polynomials. To this end, a moving window approach was used (Figure 1). The points inside the outer window, W_1, are used for fitting the polynomial. The derivatives of the polynomial in the points of interest within in the inner window W_2 are used for dynamic MFA. Making the inner window smaller than the outer window ensures a smooth transition between the different polynomials. However, if the inner window is too small compared to the outer window, the noise on the data is not filtered sufficiently, yielding nonrealistic fluctuations in the derivatives.

Some of the available data series were impossible to smooth with polynomials. They behaved like a typical logistic curve. Hence, the logistic curve was added to the set of functions that could be used to approximate the data:

$$L(t) = p_0 \frac{1 + p_1 e^{p_3 t}}{1 + p_2 e^{p_3 t}}. \tag{3}$$

The parameters W_1 and W_2 and the maximal degree of the polynomials (or logistic curve) were manually selected by visually inspection of the generated curves. The actual degree of the polynomials could vary for each polynomial fitted (thus the curve for a time series of a certain metabolite consists of polynomials of different degrees chained together): for each W_1 window the polynomial giving the least amount of outliers was selected. Data points were identified as outliers when the difference between the measured value and

Table 1: List of reactions used in the metabolic model.

Name of reaction	The reaction
PGI:	G6P ↔ F6P
PFK:	ATP + F6P → ADP + FBP
ALD:	FBP ↔ G3P + DHAP
TPI:	DHAP ↔ G3P
G3PDH:	PiOH + NAD + G3P ↔ NADH + H + BPG
PGK:	ADP + BPG ↔ ATP+3PG
PGM:	3PG ↔2PG
ENO:	2PG ↔ H_2O + PEP
PyrK:	ADP + PEP → ATP + Pyr
PyrD:	NAD + Pyr + CoA → NADH + H + AcCoA + CO_2
CitSY:	H_2O + AcCoA + OAA → CoA + Cit
ACO:	Cit ↔ iCit
CitDH:	NAD + iCit ↔ NADH + H + CO_2 + aKGA
AKGDH:	NAD + CoA + aKGA → NADH + H + CO_2 + SucCoA
SucCoASY:	ADP + PiOH + SucCoA ↔ ATP + CoA + Suc
SucDH:	FAD + Suc → FADH2 + Fum
FumHY:	H_2O + Fum ↔ Mal
MalDH:	NAD + Mal ↔ NADH + H + OAA
PEPCB:	H_2O + PEP + CO_2 → PiOH + OAA
LacDH:	NADH + H + Pyr ↔ NAD + Lac
PFLY:	Pyr + CoA → AcCoA + FA
AcKNLR:	ADP + PiOH + AcCoA ↔ ATP + CoA + Ac
Resp:	1.33ADP + 1.33PiOH + NADH + H + 0.5O_2 → 1.33ATP + NAD + 2.33H_2O
H_2CO_3SY:	H_2O + CO_2 ↔ H_2CO_3
G6PDH:	NADP + G6P → NADPH + H + 6PGL
LAS:	H_2O + 6PGL → 6PG
PGDH:	NADP + 6PG → NADPH + H + CO_2 + Rl5P
PPI:	Rl5P ↔ R5P
PPE:	Rl5P ↔ Xu5P
TK1:	R5P + Xu5P ↔ G3P + S7P
TA:	G3P + S7P ↔ F6P + E4P
TK2:	Xu5P + E4P ↔ F6P + G3P
PTS:	GLC + PEP → G6P + Pyr
PPiOHHY:	PPiOH + H_2O → 2PiOH
GluDH:	NADPH + H + aKGA + NH_3 ↔ NADP + H_2O + Glu
GluLI:	ATP + NH_3 + Glu → ADP + PiOH + Gln
AspSY:	ATP + H_2O + Asp + Gln → AMP + PPiOH + Asn + Glu
AspTA:	OAA + Glu ↔ aKGA + Asp
AlaTA:	Pyr + Glu ↔ aKGA + Ala
ValAT:	aKIV + Glu ↔ aKGA + Val
LeuSYLR:	NAD + H_2O + AcCoA + aKIV + Glu → NADH + H + CoA + CO_2 + aKGA + Leu
aKIVSYLR:	NADPH + H + 2Pyr → NADP + H_2O + CO_2 + aKIV
IleSYLR:	NADPH + H + Pyr + Glu + Thr → NADP + H_2O + CO_2 + aKGA + NH_3 + Ile
ProSYLR:	ATP + 2NADPH + 2H + Glu → ADP + PiOH + 2NADP + H_2O + Pro
SerLR:	NAD + H_2O + 3PG + Glu → PiOH + NADH + H + aKGA + Ser
SerTHM:	Ser + THF → H_2O + Gly + MeTHF
H2SSYLR:	2ATP + 3NADPH + ThioredH_2 + 3H + H_2SO_4 → ADP + PPiOH + 3NADP + Thiored + 3H_2O +H_2S + PAP
PAPNAS:	H_2O + PAP → AMP + PiOH

Table 1: Continued.

Name of reaction	The reaction
CysSYLR:	$H_2S + AcCoA + Ser \rightarrow CoA + Cys + Ac$
PrppSY:	$ATP + R5P \rightarrow AMP + PRPP$
HisSYLR:	$ATP + 2NAD + 3H_2O + Gln + PRPP \rightarrow 2PPiOH + PiOH + 2NADH + 2H + aKGA + His + AICAR$
PheSYLR:	$Glu + Chor \rightarrow H_2O + CO_2 + aKGA + Phe$
TyrSYLR:	$NAD + Glu + Chor \rightarrow NADH + H + CO_2 + aKGA + Tyr$
TrpSYLR:	$Gln + Ser + Chor + PRPP \rightarrow PPiOH + 2H_2O + G3P + Pyr + CO_2 + Glu + Trp$
DhDoPHepAD:	$H_2O + PEP + E4P \rightarrow PiOH + Dahp$
DhqSY:	$Dahp \rightarrow PiOH + Dhq$
DhsSYLR:	$Dhq \rightarrow H_2O + Dhs$
ShiSY:	$NADPH + H + Dhs \leftrightarrow NADP + Shi$
ShiKN:	$ATP + Shi \rightarrow ADP + Shi3P$
ChorSYLR:	$PEP + Shi3P \rightarrow 2PiOH + Chor$
ThrSYLR:	$ATP + H_2O + HSer \rightarrow ADP + PiOH + Thr$
MDAPSYLR:	$NADPH + H + Pyr + SucCoA + Glu + AspSA \rightarrow NADP + CoA + aKGA + Suc + MDAP$
LysSY:	$MDAP \rightarrow CO_2 + Lys$
MetSYLR:	$H_2O + SucCoA + Cys + MTHF + HSer \rightarrow Pyr + CoA + Suc + NH_3 + Met + THF$
AspSASY:	$ATP + NADPH + H + Asp \rightarrow ADP + PiOH + NADP + AspSA$
HSerDH:	$NADPH + H + AspSA \leftrightarrow NADP + HSer$
CarPSY:	$2ATP + H_2O + H_2CO_3 + Gln \rightarrow 2ADP + PiOH + Glu + CarP$
OrnSYLR:	$ATP + NADPH + H + H_2O + AcCoA + 2Glu \rightarrow ADP + PiOH + NADP + CoA + aKGA + Orn + Ac$
ArgSYLR:	$ATP + Asp + Orn + CarP \rightarrow AMP + PPiOH + PiOH + Fum + Arg$
ThioredRD:	$NADPH + Thiored + H \leftrightarrow NADP + ThioredH2$
H_2O_{2ox}:	$2H_2O_2 \rightarrow 2H_2O + O_2$
FAD2NAD:	$NAD + FADH2 \leftrightarrow NADH + FAD + H$
AICARSYLR:	$6ATP + 3H_2O + CO_2 + Asp + 2Gln + Gly + FA + PRPP \rightarrow 6ADP + PPiOH + 6PiOH + Fum + 2Glu + AICAR$
IMPSYLR:	$FTHF + AICAR \rightarrow H_2O + THF + IMP$
AMPSYLR:	$Asp + GTP + IMP \rightarrow AMP + PiOH + Fum + GDP$
AdKN:	$ATP + AMP \leftrightarrow 2ADP$
ADPRD:	$ADP + ThioredH2 \rightarrow Thiored + H_2O + dADP$
dADPKN:	$ATP + dADP \rightarrow ADP + dATP$
IMPDH:	$NAD + H_2O + IMP \rightarrow NADH + H + XMP$
GMPSY:	$ATP + H_2O + Gln + XMP \rightarrow AMP + PPiOH + Glu + GMP$
GuKN:	$ATP + GMP \rightarrow ADP + GDP$
GDPKN:	$ATP + GDP \rightarrow ADP + GTP$
GDPRD:	$ThioredH2 + GDP \rightarrow Thiored + H_2O + dGDP$
dGDPKN:	$ATP + dGDP \rightarrow ADP + dGTP$
UMPSYLR:	$O_2 + Asp + PRPP + CarP \rightarrow PPiOH + PiOH + H_2O + CO_2 + UMP + H_2O_2$
UrKN:	$ATP + UMP \rightarrow ADP + UDP$
UDPKN:	$ATP + UDP \rightarrow ADP + UTP$
CTPSY:	$ATP + H_2O + Gln + UTP \rightarrow ADP + PiOH + Glu + CTP$
CDPKN:	$ATP + CDP \leftrightarrow ADP + CTP$
CMPKN:	$ATP + CMP \rightarrow ADP + CDP$
CDPRD:	$ThioredH2 + CDP \rightarrow Thiored + H_2O + dCDP$
dCDPKN:	$ATP + dCDP \rightarrow ADP + dCTP$
UDPRD:	$ThioredH2 + UDP \rightarrow Thiored + H_2O + dUDP$

Table 1: Continued.

Name of reaction	The reaction
dUDPKN:	ATP + dUDP → ADP + dUTP
dUTPPPAS:	H_2O + dUTP → PPiOH + dUMP
dTMPSY:	MeTHF + dUMP → DHF + dTMP
dTMPKN:	ATP + dTMP → ADP + dTDP
dTDPKN:	ATP + dTDP → ADP + dTTP
DHFRD:	NADPH + H + DHF → NADP + THF
FTHFSYLR:	NADP + H_2O + MeTHF → NADPH + H + FTHF
GlyCA:	NAD + Gly + THF ↔ NADH + H + CO_2 + NH_3 + MeTHF
MeTHFRD:	NADH + H + MeTHF → NAD + MTHF
AcCoACB:	ATP + H_2O + AcCoA + CO_2 ↔ ADP + PiOH + MalCoA
MalCoATA:	MalCoA + ACP ↔ CoA + MalACP
AcACPSY:	MalACP → CO_2 + AcACP
C120SY:	10NADPH + 10H + AcACP + 5MalACP → 10NADP + 5H_2O + 5CO_2 + C120ACP + 5ACP
C140SY:	12NADPH + 12H + AcACP + 6MalACP → 12NADP + 6H_2O + 6CO_2 + C140ACP + 6ACP
C160SY:	14NADPH + 14H + AcACP + 7MalACP → 14NADP + 7H_2O + 7CO_2 + C160ACP + 7ACP
C181SY:	15NADPH + 15H + AcACP + 8MalACP → 15NADP + 8H_2O + 8CO_2 + C181ACP + 8ACP
AcylTF:	C160ACP + C181ACP + Go3P → 2ACP + PA
Go3PDH:	NADPH + H + DHAP ↔ NADP + Go3P
DGoKN:	ATP + DGo → ADP + PA
CDPDGoSY:	CTP + PA ↔ PPiOH + CDPDGo
PSerSY:	Ser + CDPDGo → CMP + PSer
PSerDC:	PSer → CO_2 + PEthAn
GlnF6PTA:	F6P + Gln → Glu + GA6P
GlcAnMU:	GA6P ↔ GA1P
NAGUrTF:	AcCoA + UTP + GA1P → PPiOH + CoA + UDPNAG
LipaSYLR:	ATP + 2CMPKDO + 2UDPNAG + C120ACP + 5C140ACP → ADP + 2CMP + UMP + UDP +6ACP + Lipa + 2Ac
A5PIR:	Rl5P ↔ Ar5P
PGLCMT:	G6P ↔ G1P
CMPKDOSYLR:	2H_2O + PEP + Ar5P + CTP → PPiOH + 2PiOH + CMPKDO
ADPHEPSY:	ATP + S7P → PPiOH + ADPHEP
UDPGlcSY:	G1P + UTP → PPiOH + UDPGlc
EthANPT:	CMP + PEthAn ↔ CDPEthAn + DGo
LpsSYLR:	3ADPHEP + 3CMPKDO + 2UDPGlc + Lipa + 2CDPEthAn → 3ADP + 3CMP + 2CDP + 2UDP + Lps
PGSYLR:	Go3P + CDPDGo → PiOH + CMP + PG
CLSY:	PG + CDPDGo → CMP + CL
PeptidoSYLR:	5ATP + NADPH + H + PEP + 3Ala + MDAP + 2UDPNAG → 5ADP + 7PiOH + NADP + UMP+ UDP + Peptido
GlcgSY:	ATP + G1P → ADP + PPiOH + Glcg
ATPHY:	ATP + H_2O → ADP → PiOH
DNASYLR:	2H_2O + 0.246dATP + 0.254dGTP + 0.254dCTP + 0.246dTTP → 2PiOH + DNA
RNASYLR:	0.262ATP + 2H_2O + 0.322GTP + 0.2CTP + 0.216UTP → 2PiOH + RNA
ProtSYLR:	2ATP + 3H_2O + 0.0961Ala + 0.05506Arg + 0.04505Asn + 0.04505 Asp + 0.01702Cys + 0.04905Gln + 0.04905Glu + 0.1151 Gly + 0.01802 His + 0.05405 Ile + 0.08408 Leu + 0.06406 Lys + 0.02903 Met + 0.03504Phe + 0.04104 Pro + 0.04004 Ser + 0.04705 Thr + 0.01101Trp + 0.02603Tyr + 0.07908Val + 2GTP → 2ADP + 4 PiOH + 2GDP + Prot
LipidSYLR:	0.0266CL + 0.202PG + 0.7714PEthAn → Lipid
BiomSYLR:	0.004561Glcg + 0.0002663 Lps + 0.0008933Peptido + 0.002291DNA + 0.01446RNA + 0.1227 Prot + 0.003642 Lipid → Biom

Table 2: List of metabolites used in the metabolic model.

Metabolite	Molecular formula	Full name
2PG	$C_3H_7O_7P$	2-phophoglycerate
3PG	$C_3H_7O_7P$	3-phophoglycerate
6PG	$C_6H_{13}O_{10}P$	6-phosphogluconate
6PGL	$C_6H_{11}O_9P$	6-phosphogluconolactone
Ac	$C_2H_4O_2$	Acetate
AcACP	C_2H_3OPept	Acetyl ACP
AcCoA	$C_{23}H_{34}O_{17}N_7P_3S$	Acetyl CoA
ACP	HPept	Acyl carrier protein
ADP	$C_{10}H_{15}O_{10}N_5P_2$	Adenosine diphosphate
ADPHEP	$C_{17}H_{27}O_{16}N_5P_2$	ADP-Mannoheptose
AICAR	$C_9H_{15}O_8N_4P$	Amino imidazole carboxamide ribonucleotide
aKGA	$C_5H_6O_5$	Alpha ketoglutaric acid
aKIV	$C_5H_8O_3$	Alpha-keto-isovalerate
Ala	$C_3H_7O_2N$	Alanine
AMP	$C_{10}H_{14}O_7N_5P$	Adenosine monophosphate
Ar5P	$C_5H_{11}O_8P$	Arabinose-5-phosphate
Arg	$C_6H_{14}O_2N_4$	Arginine
Asn	$C_4H_8O_3N_2$	Aspartate
Asp	$C_4H_7O_4N$	Asparagine
AspSA	$C_4H_7O_3N$	Aspartate semialdehyde
ATP	$C_{10}H_{16}O_{13}N_5P_3$	Adenosine triphosphate
Biom	$CH_{1.63}O_{0.392}N_{0.244}P_{0.021}S_{0.00565}$	Biomass
BPG	$C_3H_8O_{10}P_2$	1-3-biphosphoglycerate
C120ACP	$C_{12}H_{23}OPept$	
C140ACP	$C_{14}H_{27}OPept$	
C160ACP	$C_{16}H_{31}OPept$	
C181ACP	$C_{18}H_{33}OPept$	
CarP	CH_4O_5NP	Carbamoyl phosphate
CDP	$C_9H_{15}O_{11}N_3P_2$	Citidine diphosphate
CDPDGo	$C_{46}H_{83}O_{15}N_3P_2$	CDP-diacylglycerol
CDPEthAn	$C_{11}H_{20}O_{11}N_4P_2$	CDP-ethanolamine
Chor	$C_{10}H_{10}O_6$	Chorismate
Cit	$C_6H_8O_7$	cisaconitate
CL	$C_{77}H_{144}O_{16}P_2$	Cardiolipin
CMP	$C_9H_{14}O_8N_3P$	Citidine monophosphate
CMPKDO	$C_{17}H_{26}O_{15}N_3P$	CMP-2-keto-3-deoxyoctanoate
CO_2	CO_2	Carbon dioxide
CoA	$C_{21}H_{32}O_{16}N_7P_3S$	Coenzyme A
CTP	$C_9H_{16}O_{14}N_3P_3$	Citidine triphosphate
Cys	$C_3H_7O_2NS$	Cysteine
dADP	$C_{10}H_{15}O_9N_5P_2$	deoxy ADP
Dahp	$C_7H_{13}O_{10}P$	Deoxy arabino heptulosonate
dATP	$C_{10}H_{16}O_{12}N_5P_3$	deoxy ATP
dCDP	$C_9H_{15}O_{10}N_3P_2$	deoxy CDP
dCTP	$C_9H_{16}O_{13}N_3P_3$	deoxy CTP
dGDP	$C_{10}H_{15}O_{10}N_5P_2$	deoxy GDP
DGo	$C_{37}H_{70}O_5$	Diacyl glycerol
dGTP	$C_{10}H_{16}O_{13}N_5P_3$	deoxy GTP
DHAP	$C_3H_7O_6P$	Dihydroxyaceton phosphate
DHF	$C_{19}H_{21}O_6N_7$	Dihydrofolate

Table 2: Continued.

Metabolite	Molecular formula	Full name
Dhq	$C_7H_{10}O_6$	Dehydroquinate
Dhs	$C_7H_8O_5$	Dehydroshikimate
DNA	$C_{9.75}H_{14.2}O_7N_{3.75}P$	DNA composition
dTDP	$C_{10}H_{16}O_{11}N_2P_2$	deoxy TDP
dTMP	$C_{10}H_{15}O_8N_2P$	deoxy TMP
dTTP	$C_{10}H_{17}O_{14}N_2P_3$	deoxy TTP
dUDP	$C_9H_{14}O_{11}N_2P_2$	deoxy UDP
dUMP	$C_9H_{13}O_8N_2P$	deoxy UMP
dUTP	$C_9H_{15}O_{14}N_2P_3$	deoxy UTP
E4P	$C_4H_9O_7P$	Erythrose-4-phosphate
F6P	$C_6H_{13}O_9P$	Fructose-6-phosphate
FA	CH_2O_2	Formic Acid
FAD	$C_{27}H_{33}O_{15}N_9P_2$	Flavine adeninen dinucleotide
FADH2	$C_{27}H_{35}O_{15}N_9P_2$	
FBP	$C_6H_{14}O_{12}P_2$	Fructose-1-6-biphosphate
FTHF	$C_{20}H_{23}O_7N_7$	Formyl tetrahydrofolate
Fum	$C_4H_4O_4$	Fumarate
G1P	$C_6H_{13}O_9P$	Glucose-1-phosphate
G3P	$C_3H_7O_6P$	Glyceraldehyde-3-phosphate
G6P	$C_6H_{13}O_9P$	Glucose-6-phosphate
GA1P	$C_6H_{14}O_8NP$	D-glucosamine-6-phosphate
GA6P	$C_6H_{14}O_8NP$	Glucose-6-phosphate
GDP	$C_{10}H_{15}O_{11}N_5P_2$	Guanosine diphosphate
GLC	$C_6H_{12}O_6$	Glucose
Glcg	$C_6H_{10}O_5$	Glycogen
Gln	$C_5H_{10}O_3N_2$	Glutamine
Glu	$C_5H_9O_4N$	Glutamate
Gly	$C_2H_5O_2N$	Glycine
GMP	$C_{10}H_{14}O_8N_5P$	Guanosine monophosphate
Go3P	$C_3H_9O_6P$	Glycerol-3-phosphate
GTP	$C_{10}H_{16}O_{14}N_5P_3$	Guanosine triphosphate
H	H^+	Hydrogene
H_2CO_3	CH_2O_3	Bicarbonate
H_2O	H_2O	Water
H_2O_2	H_2O_2	
H_2S	H_2S	Hydrogene sulde
H_2SO_4	H_2O_4S	Sulfuric acid
His	$C_6H_9O_2N_3$	Histidine
HSer	$C_4H_9O_3N$	Homoserine
iCit	$C_6H_8O_7$	Isocitrat
Ile	$C_6H_{13}O_2N$	Isoleucine
IMP	$C_{10}H_{13}O_8N_4P$	Inosine monophosphate
Lac	$C_3H_6O_3$	Lactate
Leu	$C_6H_{13}O_2N$	Leucine
Lipa	$C_{110}H_{196}O_{32}N_2P_2$	Lipid A
Lipid	$C_{40.2}H_{77.6}O_{8.41}N_{0.771}P_{1.03}$	Lipid composition
Lps	$C_{171}H_{298}O_{81}N_4P_2$	Lipo Poly saccharide
Lys	$C_6H_{14}O_2N_2$	Lysine
Mal	$C_4H_6O_5$	Malate
MalACP	$C_3H_3O_3Pept$	Malonyl ACP
MalCoA	$C_{24}H_{34}O_{19}N_7P_3S$	Malonyl CoA

Table 2: Continued.

Metabolite	Molecular formula	Full name
MDAP	$C_7H_{14}O_4N_2$	Meso-diaminopimelate
Met	$C_5H_{11}O_2NS$	Methionine
MeTHF	$C_{20}H_{23}O_6N_7$	Methylene tetrahydro folate
MTHF	$C_{20}H_{25}O_6N_7$	Methyl tetrahydrofolate
NAD	$C_{21}H_{28}O_{14}N_7P_2^+$	Nicotinamide adenine dinucleotide
NADH	$C_{21}H_{29}O_{14}N_7P_2$	
NADP	$C_{21}H_{28}O_{17}N_7P_3^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	$C_{21}H_{29}O_{17}N_7P_3$	
NH_3	H_3N	Ammonia
O_2	O_2	Oxygen
OAA	$C_4H_4O_5$	Oxaloacetate
Orn	$C_5H_{12}O_2N_2$	Ornithine
PA	$C_{37}H_{71}O_8P$	Phosphatidyl acid
PAP	$C_{10}H_{15}O_{10}N_5P_2$	Phospho adenosine phosphate
PEP	$C_3H_5O_6P$	Phosphoenolpyruvate
Peptido	$C_{35}H_{53}O_{16}N_7$	Peptidoglycan
PEthAn	$C_{39}H_{76}O_8NP$	Phosphatidyl ethanolamine
PG	$C_{40}H_{75}O_9P$	Phosphatidyl glycerol
Phe	$C_9H_{11}O_2N$	Phenylalanine
PiOH	H_3O_4P	Phosphate
PPiOH	$H_4O_7P_2$	Pyrophosphate
Pro	$C_5H_9O_2N$	Proline
Prot	$C_{4.8}H_{7.67}O_{1.4}N_{1.37}S_{0.046}$	Protein composition
PRPP	$C_5H_{13}O_{14}P_3$	5-phospho-alpha-D-ribosyl-1-pyrophosphate
PSer	$C_{40}H_{76}O_{10}NP$	Phosphatidyl Serine
Pyr	$C_3H_4O_3$	Pyruvate
R5P	$C_5H_{11}O_8P$	Ribose-5-phosphate
Rl5P	$C_5H_{11}O_8P$	Ribulose-5-phosphate
RNA	$C_{9.58}H_{13.8}O_{7.95}N_{3.95}P$	RNA composition
S7P	$C_7H_{15}O_{10}P$	Sedoheptulose-7-phosphate
Ser	$C_3H_7O_3N$	Serine
Shi	$C_7H_{10}O_5$	Shikimate
Shi3P	$C_7H_{11}O_8P$	Shikimate-3-phosphate
Suc	$C_4H_6O_4$	Succinate
SucCoA	$C_{25}H_{36}O_{19}N_7P_3S$	Succinyl CoA
THF	$C_{19}H_{23}O_6N_7$	Tetrahydrofolate
Thiored	Pept	Thioredoxin
ThioredH_2	H_2Pept	Reduced thioredoxin
Thr	$C_4H_9O_3N$	Threonine
Trp	$C_{11}H_{12}O_2N_2$	Tryptophan
Tyr	$C_9H_{11}O_3N$	Tyrosine
UDP	$C_9H_{14}O_{12}N_2P_2$	Uridine diphosphate
UDPGlc	$C_{15}H_{24}O_{17}N_2P_2$	UDP glucose
UDPNAG	$C_{17}H_{27}O_{17}N_3P_2$	UDP N-acetyl glucosamine
UMP	$C_9H_{13}O_9N_2P$	Uridine monophosphate
UTP	$C_9H_{15}O_{15}N_2P_3$	Uridine triphosphate
Val	$C_5H_{11}O_2N$	Valine
XMP	$C_{10}H_{13}O_9N_4P$	Xanthosine-5-phosphate
Xu5P	$C_5H_{11}O_8P$	Xylulose-5-phosphate

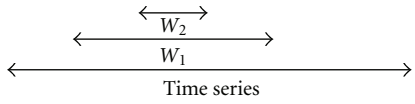


Figure 1: A moving window is run through a time series of data. W_1 is the polynomial fitting window, W_2 is the interpolation window.

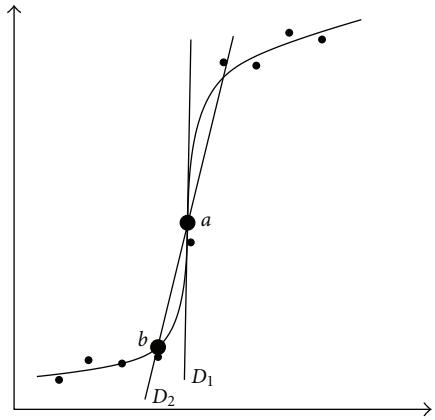


Figure 2: Two ways of estimating the derivative in point a: D_1 is calculated by symbolically deriving the function in point a; D_2 is the slope between point a and point b (where a and b are calculated from the curve). Dots represent sample points through which a curve is fitted.

Table 3: Parameters for the polynomial fitting of the different time series. W_1 is expressed in fraction of the time between the first measurement and the last measurement. W_2 is expressed as a fraction of W_1. PD is the highest degree a polynomial may have; an L is displayed if the data were fitted to a logistic curve.

	C → N limitation			N → C limitation		
	W_1	W_2	PD	W_1	W_2	PD
GLC	0.40	0.87	5	0.35	0.45	3
NH_3	0.50	0.80	2	1	1	L
PiOH	0.70	0.75	5	0.45	0.80	4
O_2	0.18	0.90	7	0.11	0.90	7
CO_2	0.12	0.90	4	0.12	0.90	7
Ac	0.70	0.88	5	0.15	0.70	3
Lac	1	1	7	0.41	0.80	7
Pyr	0.40	0.80	6	1	1	L
Suc	0.40	0.80	6	0.82	0.99	5
Biomass	0.50	0.70	8	0.37	0.80	4

the estimated value was larger than 5% of the range of the response variable. Table 3 summarized the fitting parameters used for each metabolite.

The polynomial or logistic functions yield a smoothed time profile of the concentrations and enable to calculate the derivative of the concentrations to time. However, when calculating this derivative directly by symbolically deriving the function and filling in the time for the point of interest, one sometimes obtains unrealistically high values (D_1 in Figure 2). Therefore, the derivative was approximated by taking the slope of the point of interest and the previous point on the polynomial (D_2 in Figure 2).

The derivative can subsequently be used in (1) to calculate the cellular production (or consumption) fluxes:

$$r_p = \frac{dC}{dt} - D(C_{in} - C). \quad (4)$$

Once r_p has been calculated, classical MFA techniques (as described in [4]) can be applied at each time point. The results of those models, evaluated at the different time points, can then be combined to obtain a time profile of internal fluxes.

3.3. Growth Rate Calculation. During the transient period, the assumption that the growth rate is equal to the dilution rate does not hold. However, (4) can explicitly be rewritten for the following biomass:

$$r_{p,Biom} = \frac{dC_{Biom}}{dt} + DC_{Biom} \quad (5)$$

from which the growth rate can be calculated as

$$\mu = \frac{r_{p,Biom}}{C_{Biom}} = \frac{dC_{Biom}}{dt}\frac{1}{C_{Biom}} + D. \quad (6)$$

4. Results

4.1. Polynomial Fitting. The polynomials fitted through the data (Figures 3 and 4) are used to obtain the concentration, C, and the derivative of the concentration to the time, dC/dt, in each point. From those values, the conversion rate can be calculated (4). Biomass concentrations were calculated from OD values, after generating a calibration curve between cell dry weight (CDW) and OD values. This calibration curve was generated at the end of each experiment. As the curves for both experiments were very close to each other, no distinct calibration was used for the C-limited and N-limited conditions.

The steady state concentration of NH_3 during N-limitation in the C- to N-limitation experiment is around 0.2 g/L (data are not shown on Figure 3, as this value falls out the time frame shown), while it is 0.1 g/L in the N to C-limitation case. These values are high compared to what is found in the literature for the residual nitrogen concentration in similar strains [29] and the nitrogen measurement kit used is probably not suitable for accurately measuring very low NH_3 concentrations.

4.2. Growth Rate and Biomass Production. At time zero, the influent medium was simply switched. No cells were extracted and resuspended in the other medium, as done in [32]. Thus at the switch of the limiting compound, there is an abundance of both, as the medium limited in carbon is saturated with nitrogen while the medium limited in nitrogen is saturated with glucose. In the C to N-limitation experiment, it seems as if N-limitation is reached not much later than when carbon abundance starts (Figure 3). In the

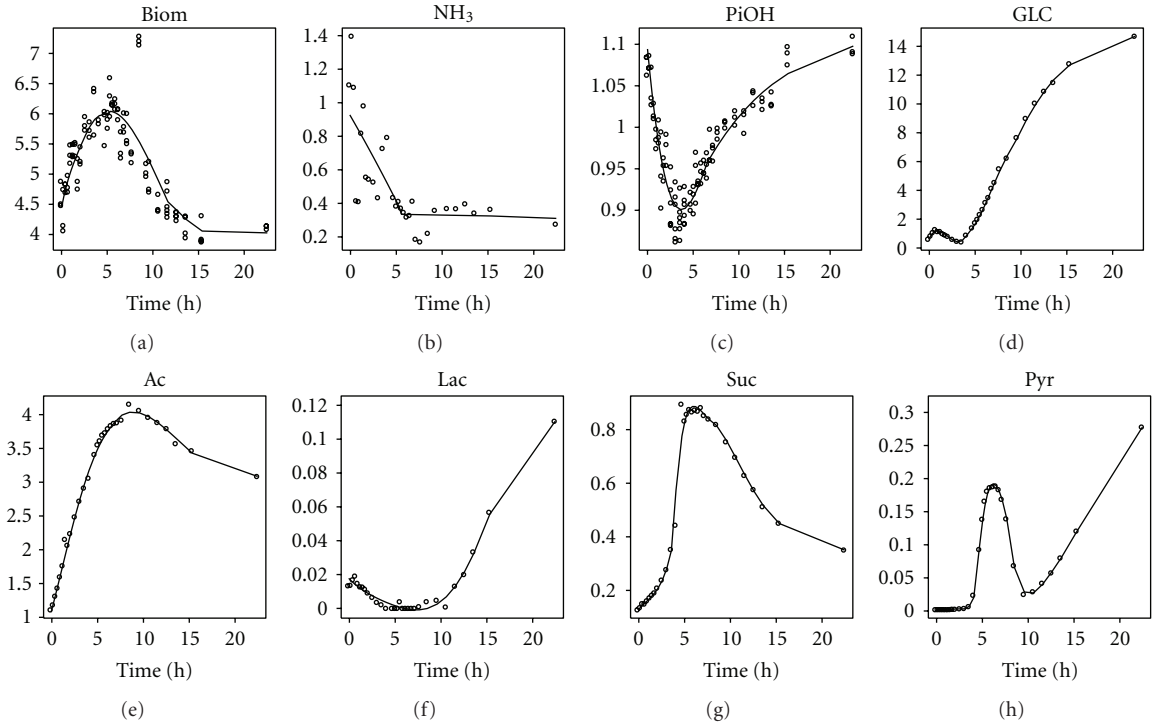


Figure 3: Polynomial fit of some metabolites (expressed in g/L) for the experiment where carbon-limiting medium was replaced with nitrogen-limiting medium. The switch is made at time zero, when the cells are in carbon-limited steady state. Steady state values of five residence times before the switch and five residence times after the switch are not shown.

N to C-limitation case, nitrogen abundance is reached before all the glucose is depleted: after 5 hours, both glucose and nitrogen are not limiting. This can be seen in Figure 4, in the glucose profile. The glucose concentration becomes only zero after 12 hours while the nitrogen concentration already reaches steady state after 5 hours. At the medium switch, as there is excess nitrogen, the growth rate could increase, but apparently *E. coli* needs some time to adapt to the new conditions. The biomass flux even decreases as nitrogen is added and becomes almost zero (top of Figure 5). Only seven hours after the medium switch, the growth rate increases sharply to attain almost $0.6\,h^{-1}$ (right part of Figure 5). This increased growth rate coincides with the uptake of all glucose left in the broth (Figure 4).

Figure 5 depicts the different growth rates during the transient conditions occurring when switching the limiting substrate. The same trends as in the biomass flux (top of Figure 5) can be observed. For the C- to N-limitation case, a sharp increase in growth rate is observed in the beginning, while for the N- to C-limitation the growth rate decreases to almost zero before increasing to $0.6\,h^{-1}$, not far from the maximal growth rate of $0.7\,h^{-1}$. In the carbon to nitrogen-limitation case, the maximal possible growth rate is never achieved and after a sharp increase in the beginning to $0.35\,h^{-1}$ the growth rate decreases.

The sharp increase in growth rate of the cells at the beginning of the medium switch in the C- to N-limitation experiment (left part of Figure 5) is not accompanied by an increased flux through the Krebs cycle (Figure 8, the first hour after the steady state point, labeled with a triangle). An increase in biomass flux does not imply that intracellular fluxes relative to the biomass flux should increase as well. On the contrary, in the beginning the cells are more energy-efficient than during the carbon-limited steady state (left subfigure of Figure 7 shows that the ATP hydrolysis flux relative to the biomass flux is lower in the beginning implying that the maintenance energy requirement is lower). However, very quickly the flux through the glycolysis increases.

During the switch of limiting substrate from carbon to nitrogen, less oxygen is consumed and less CO_2 is produced than during steady state (Figure 6). In the beginning of the medium switch, a sharp decline in O_2 consumption and CO_2 production is observed. This correlates with the increased growth rate and biomass production rate (Figure 5). For the experiment where nitrogen limitation is switched to carbon limitation, two parts can be observed: in the first period (until five hours), the O_2 consumption rate and CO_2 production rate increase. Then they suddenly decrease (around 7 hours), to steadily increase again until they reach the steady state level. In both experiments, the steady state value of oxygen consumption and carbon dioxide production is around 16 g/mol Biomass/h. This is in accordance with values obtained with a similar strain [4].

The ATP hydrolysis reaction (Figure 7) lumps all the maintenance requirements and futile cycles (as including all the futile cycles in the metabolic model would yield parallel pathways and the system of equations would not be solvable without measuring intracellular fluxes). It can be

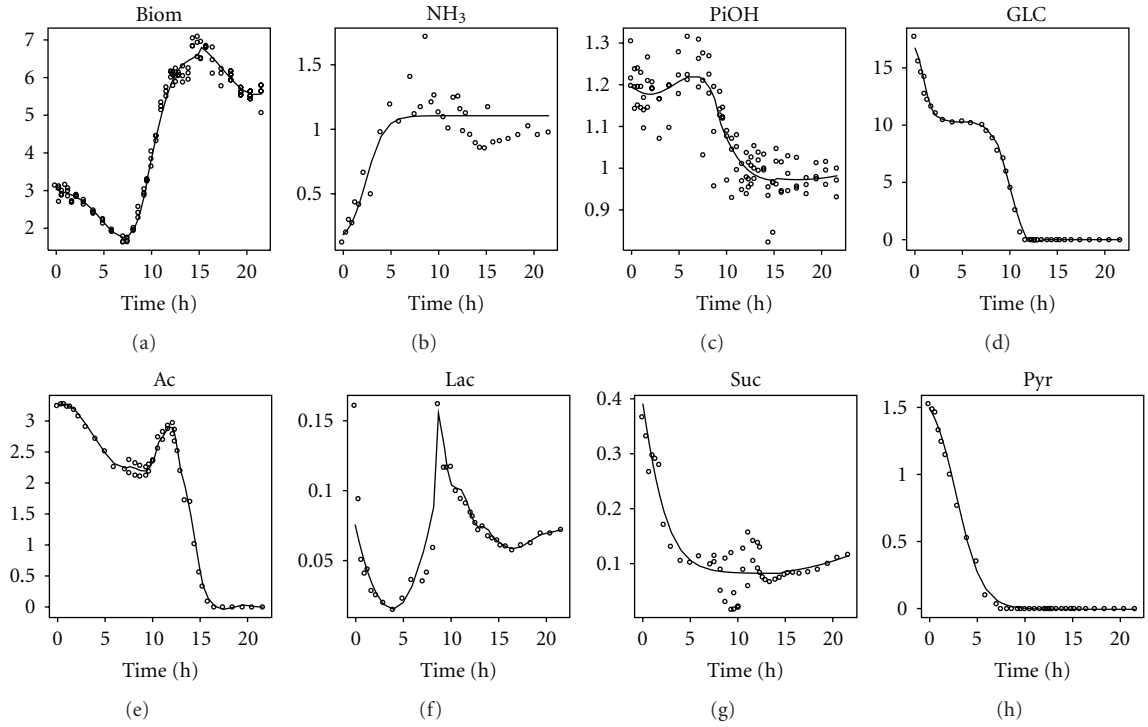


Figure 4: Polynomial fit of some metabolites (expressed in g/L) for the experiment where nitrogen-limiting medium was replaced with carbon-limiting medium. The switch is made at time zero, when the cells are in nitrogen-limited steady state. Steady state values of five residence times before the switch and five residence times after the switch are not shown.

seen that during the lag phase of the experiment switching nitrogen to carbon-limitation (right part of Figure 7), the ATP maintenance requirement is around 0.6 mol ATP/mol Biom/h, which is high compared to the values found in a similar *E. coli* [4]. It seems even more unusually high because there is minimal growth during this period. This high ATP requirement in the beginning of the lag phase correlates with a high uptake of glucose (PTS reaction in Figure 9) and an increased flux in the TCA cycle (Figure 9). The somewhat higher steady state ATP flux at the end of the C- to N-limitation (left part of Figure 7) may be due to multiple physiological steady states [33].

4.3. Central Carbon Metabolism. During the transient in the C- to N-limitation experiment, the fluxes through the glycolysis increase while those through the Krebs cycle decrease. Also an increase in acetate excretion is observed (Figure 8). The opposite is observed in the N- to C-limitation experiment: fluxes through the glycolysis and acetate production decrease while the fluxes through the Krebs cycle increase (Figure 9). The decrease is not linear: at 15 hours the fluxes through the glycolysis reach a minimum to increase again until the steady state values are obtained after the subsequent 5 hours. In the Krebs cycle, the fluxes gradually increase the first five hours (coinciding with the lag phase observed in the biomass production, Figure 5) after which a sharp decline is observed, immediately followed by a gradual increase, peaking at around 15 hours, and then decrease to steady state level. The succinate flux sharply decreases in the beginning of the switch to gradually increase during the lag phase (5 hours) at which point it reaches a value that does not change anymore. The PEPCB flux, which is responsible for replenishing to the Krebs cycle, follows (or predates) this succinate excretion rate. The excretion of pyruvate decreases during the lag phase and then sharply increases again to reach a lower steady state than under N limitation.

The pentose phosphate pathway is entirely driven by the demands in biomass components, and as each flux is expressed relatively to the biomass, to remove the influence of biomass on the flux values, the fluxes of the pentose phosphate pathway do not change, which results in seemingly huge error bars.

5. Discussion

5.1. Fluxes. When going from an nitrogen-limited medium to a carbon-limited one, a lag phase was observed (Figure 5). A hypothesis could be that when more nitrogen becomes available, the cell starts to break down the accumulated glycogen which is redirected to energy metabolism for protein production. As the biomass composition is assumed to be constant in the model, this increased respiration is measured in the off-gas (O_2 and CO_2 fluxes, Figure 6) and propagated in the model to the ATPHY flux (Figure 7). However, the question remains why the cell would degrade glycogen while there is still plenty of carbon available in the environment. Furthermore, the carbon balance data (not shown) do not reflect an increase in carbon output compared to the carbon input.

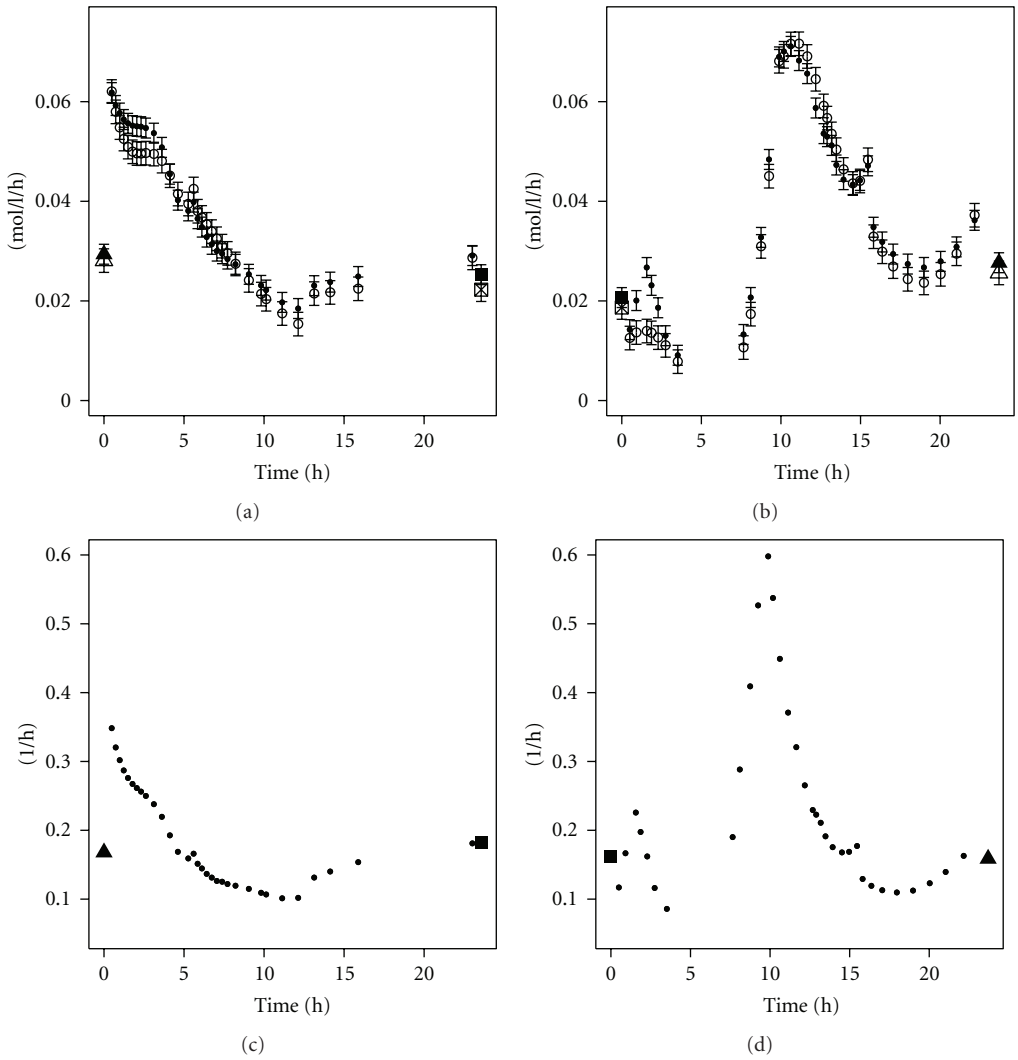


Figure 5: Fluxes in mol/L/h of the biomass production. Open symbols are the values as derived from the polynomials with formula 4; closed symbols are values obtained after flux balancing (top). The growth rate in 1/h of the cells during the transients (bottom). Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

Another explanation is that ammonia abundance for cells previously grown under ammonia limiting conditions causes a futile cycle of ammonia import and export. Such cycles have been described previously in micro-organisms, although not under the experimental conditions described in this work [34]. Normally bacterial cells are highly resistant to ammonium, and the negative effects of high NH_4^+ concentrations are due to an enhanced osmolarity or increased ionic strength of the medium and not caused by ammonium itself [35]. But when changing the medium from nitrogen-limitation to carbon-limitation, the environment of the bacteria switches from low NH_4^+ concentrations to high ones, and it could be that *E. coli* needs a certain adaptation period. The scavenging active during N-limitation could still be active in the beginning of the transition, resulting in a high influx of ammonia. To counter this, active efflux is needed, draining ATP and explaining the high energetic cost in the beginning of the addition of ammonium (right part of Figure 7).

Such a mechanism has been observed in plants, where NH_4^+ toxicity is due to the inability of root cells to limit the influx of ammonium. The high cytosolic NH_4^+ concentration activates the high-capacity and energy demanding ammonium efflux systems. This ammonium efflux can constitute as much as 80% of primary influx, resulting in a futile cycle of nitrogen across the plasma membrane of root cells. This futile cycle imposes a high energetic cost on the cell that is independent of N metabolism and is accompanied by a decline in growth. Plants that are resistant to high ammonium concentrations (e.g., *Oriza sativa*) limit the influx of ammonium by lowering the polarisation of the cellular membrane. This lowering of membrane polarisation is not observed in *Hordeum vulgare*, sensitive to ammonium [36].

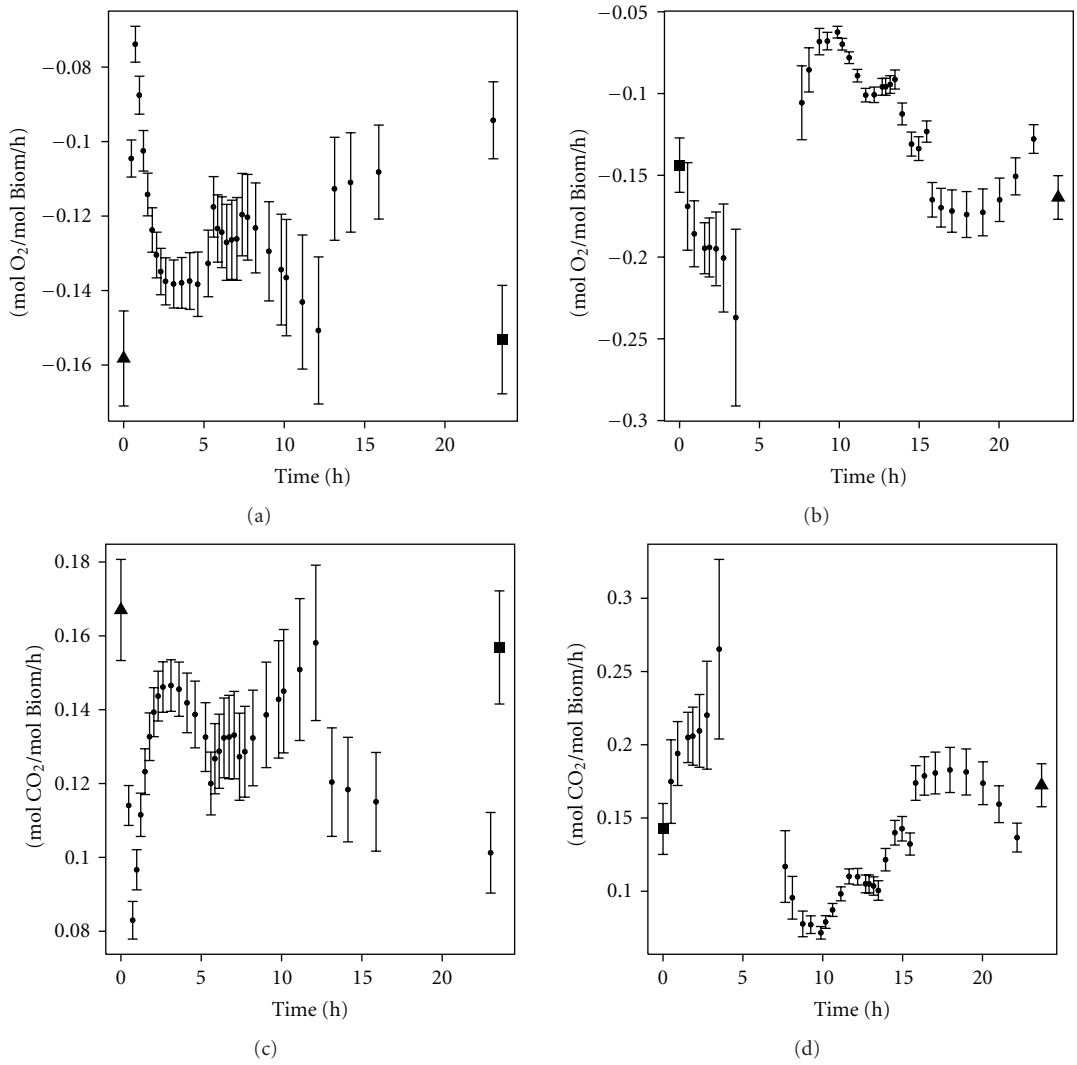


Figure 6: Amount of oxygen consumed (top) and carbon dioxide produced (bottom). Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

The lack of growth resulting from adding ammonium to cells that are adapted to ammonium limitation has previously only been observed in cells lacking *glnE* (giving the gene product ATase, responsible for activating/inactivating glutamine synthetase) while the wild-type *E. coli* did not show an impaired growth when shifted from ammonium limitation to excess [37, 38]. Furthermore, the lack of growth in *glnE* mutants could be alleviated in cells constitutively expressing only low levels of glutamine synthetase [37], suggesting that the toxicity of ammonium in this case is not due to NH_4^+ itself, but maybe to accumulation of glutamine and/or glutamate. Normally, when shifting from ammonium limitation to carbon-limitation, the amount of active glutamine synthetase decreases instantly [39]. Possibly due to the lack of ATase, the enzymatic glutamine synthetase regulation is lost, and only genetic regulation is present. If this would be the case for the strain used in this study, it could explain the long lag phase after the medium switch. The cells survive until enough glutamine synthetase is degraded before they are able to grow again.

Thus it seems as *E. coli* MG1655, the strain used in this study, has some anomalies in its nitrogen metabolism. It is known that multiple strains are denominated under this name [40]. The strain used in this work originated from the Coli Genetic Stock Center (CGSC) (personal communications from the Netherlands Culture Collection of Bacteria). The NCCB acquired this strain from CGSC prior to 1991. At that time, the strain was sent out as being fnr^-. Later it was found that the strain was a mixture of fnr^+ and fnr^- strains (personal communication from Mary Berlyn). For the experiments in this work, it turned out that the fnr^- strain was used. More research is needed to find out why this strain does not grow well when NH_4^+ is added to a nitrogen-limiting environment.

The lag phase found in the N-limitation to C-limitation experiment stands in sharp contrast with the experiment

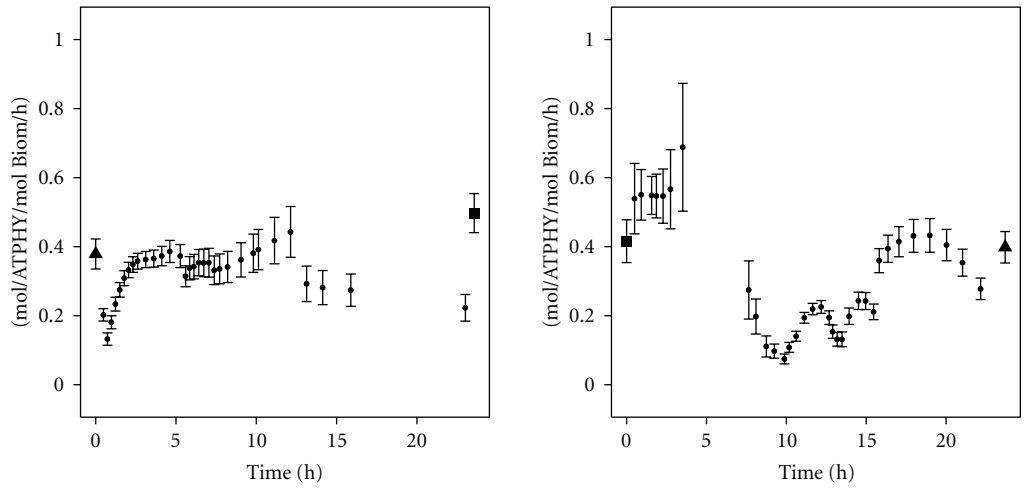


Figure 7: Fluxes through the ATP hydrolysis reaction (ATPHY reaction): it can be considered as a measure for the maintenance requirement of the cells, as it combines all the futile cycles and ATP hydrolysed in nonspecific reactions. Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

in which the carbon-limiting medium is replaced with the nitrogen-limiting one and thus carbon is added in the beginning of the switch while nitrogen is not yet limiting. The growth rate increases instantly (top of Figure 5). Death and Ferenci [41] demonstrated that during carbon-limitation, sugar regulons are upregulated by internally synthesised sugars. Lequeux et al. [4] observed that under glucose limitation *ptsG* is upregulated. The results of the experiments described here show that this up-regulation permits the cells to almost instantly increase their growth rate when carbon is added to a carbon-limited culture. The cells grow as fast as possible, until all nitrogen is consumed and thus the glucose concentration in the reactor broth increases around the same time as the nitrogen decreases (Figure 3).

Hence the peak of succinate (and the smaller peak of pyruvate) occurring around 5 hours after the medium switch (Figure 8) could be due to the depletion of ammonium (Figure 3). The high flux through the glycolysis is not stopped quickly enough and the carbon is excreted as pyruvate or diverted to PEPCB to be excreted as succinate.

5.2. Dynamic MFA. The extension of metabolic flux analysis to allow to predict also intracellular fluxes for cells growing under nonsteady state conditions was successfully applied to transient cultures. As raw measurement data are too noisy, smoothing had to be performed prior to the calculation of the exchange rates, needed for solving the metabolic model. An improvement for this smoothing process would be to incorporate more than two points in the calculation of the derivative, to avoid steep derivatives (and thus unrealistic fluxes) caused by the quirks of the higher-order polynomials.

A possible improvement to the model is to allow fluctuations of the biomass composition. However, this should be accompanied with the appropriate measurements such as protein, RNA, DNA, lipid... content of the biomass. The biomass reaction can then be removed from the model and instead the different biomass compounds can be considered as excreted by the cell.

6. Conclusion

This work introduced the concept of dynamic metabolic flux analysis. We demonstrated how to transform extracellular measurement data from dynamic experiments to flux values. It is not always clear in transient experiments whether the decrease in concentration of the extracellular metabolites is due to the cells stopping production and the remaining product diluting out or whether the cells are actively taking up the metabolites. Transforming the extracellular time series of concentrations to flux values can give the answer to that question. Furthermore, those flux values can help to get insight into the intracellular reactions.

The transformation of time series of concentration measurements to flux values is based on differentiation (in the mathematical sense: finding the derivative) of those time series. Differentiation typically amplifies the noise on the data; therefore a noise-reducing step is needed prior to the differentiation. In this work, this was done by polynomial filtering. Selecting the right parameters for this filtering was done manually and produced satisfying results for the case presented. However, a statistically more sound approach may be desirable.

We subsequently applied dynamic metabolic flux analysis on *E. coli* cultures in which the limiting compound in the medium was changed from nitrogen to glucose or from glucose to nitrogen.

A lag phase occurred when cells adapted to a nitrogen-limiting environment were supplied with nitrogen. During this lag phase of several hours, the ATP demand was high. No clear physiological reason could be given. The only similar

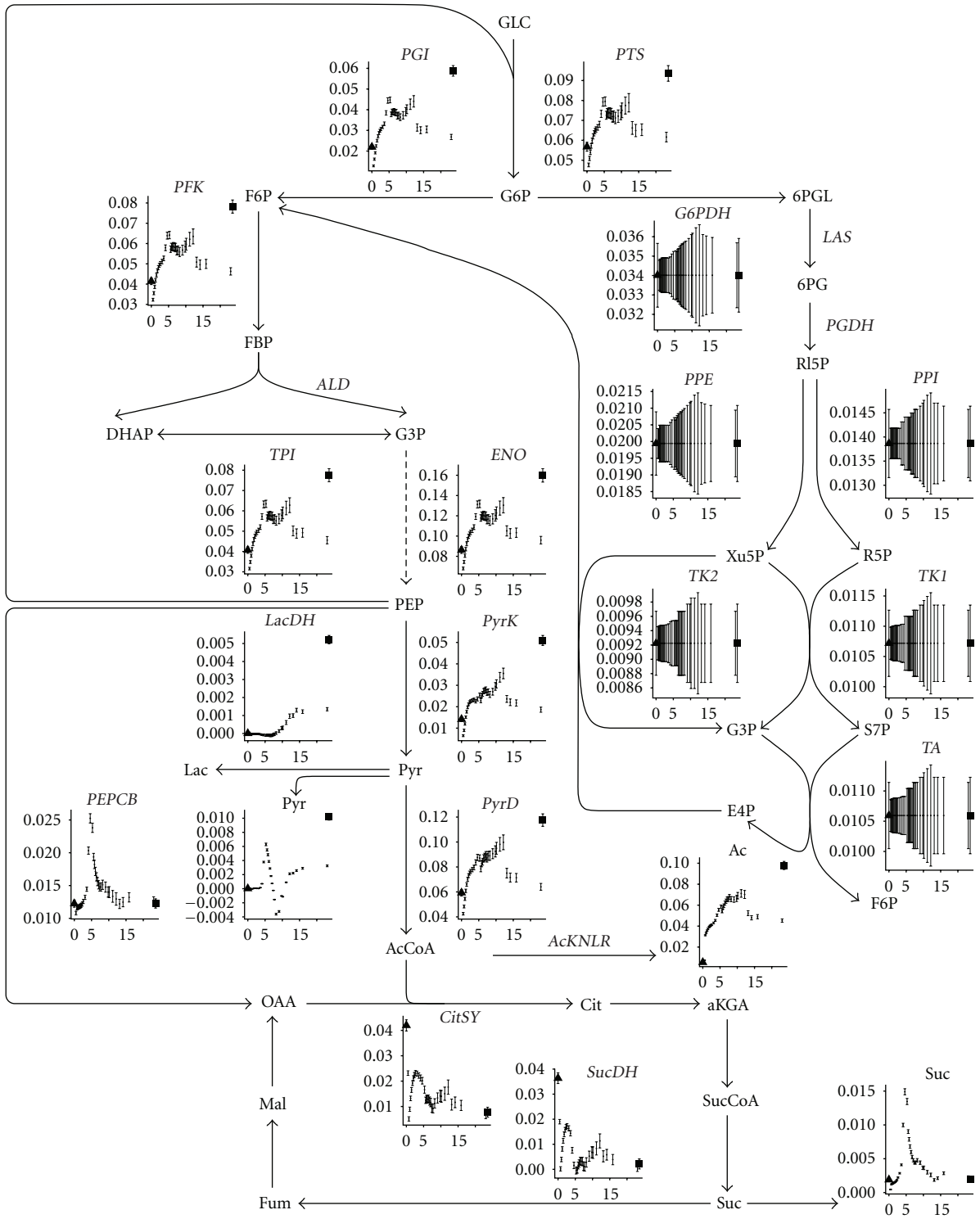


Figure 8: Fluxmap of the glycolysis, penthose phosphate pathway, and citric acid cycle for the experiment where carbon-limiting medium is changed to nitrogen-limiting medium at time zero. The ordinate on each graph represents the flux expressed in mol/mol Biomass/h while the abscissa represents time. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

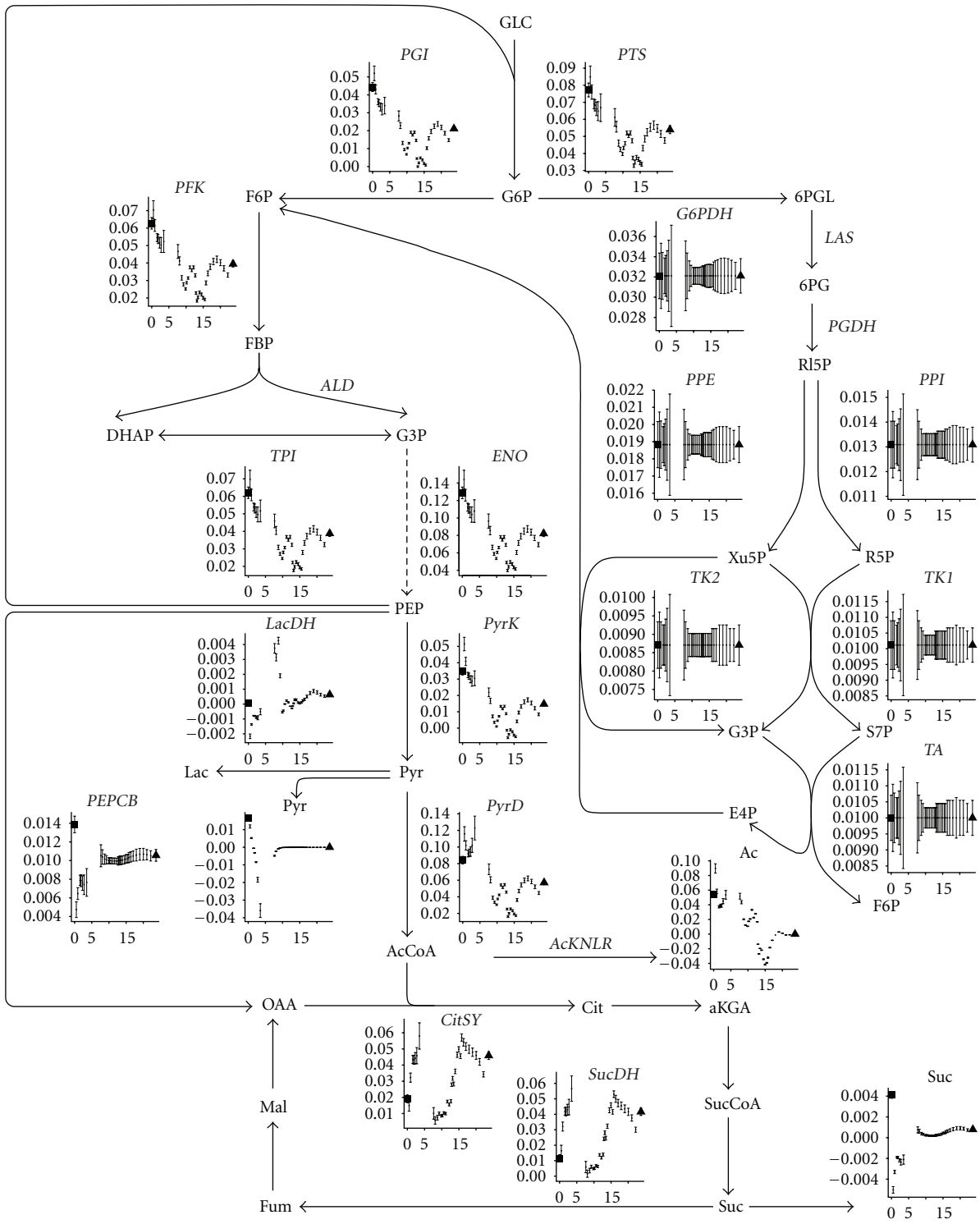


Figure 9: Fluxmap of the glycolysis, penthose phosphate pathway, and citric acid cycle for the experiment where nitrogen-limiting medium is changed to carbon-limiting medium at time zero. The ordinate on each graph represents the flux expressed in mol/mol Biomass/h while the abscissa represents time. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

case described in literature for *E. coli* is a *glnE* knock out, in which case the toxicity was not due to ammonium itself, but to the accumulation of glutamine/glutamate because the down-regulation of glutamine synthetase is not operative (glutamine synthetase is down-regulated by the gene product of *glnE*). A possible explanation for the long lag phase observed in the N-to-C-limitation experiment could be that sufficient glutamine synthetase has to be degraded so that the accumulation of glutamine/glutamate does not impair growth anymore.

No lag phase occurred when supplying cells adapted to carbon limitation, with carbon. Instead the growth rate increased immediately. However, the highest growth rates were attained in the N to C-limitation experiment and not in the C- to N-limitation one.

Acknowledgments

The authors thank Professor. Raymond Cunin for fruitful discussions about nitrogen metabolism. This project was funded by the FWO-Vlaanderen (Fund for Scientific Research, Flanders, Belgium) Grant no. G.0184.04 and by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) via the MEMORE project (040125). J. Maertens was research assistant of the FWO-Vlaanderen (2003–2007). They also thank the reviewers to substantially help improve the paper by their throughout reading and constructive comments.

References

[1] R. T. J. M. van der Heijden, J. J. Heijnen, C. Hellinga, B. Romein, and K. C. A. M. Luyben, "Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates," *Biotechnology and Bioengineering*, vol. 43, no. 1, pp. 3–10, 1994.

[2] R. T. J. M. van der Heijden, B. Romein, J. J. Heijnen, C. Hellinga, and K. C. A. M. Luyben, "Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors," *Biotechnology and Bioengineering*, vol. 43, no. 1, pp. 11–20, 1994.

[3] B. Romein, *Mathematical modelling of mammalian cells in suspension culture*, Ph.D. thesis, Technische Universiteit Delft, Delft, The Netherlands, 2001.

[4] G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem, and G. Lidén, "MFA for overdetermined systems reviewed and compared with RNA expression data to elucidate the difference in shikimate yield between carbon- and phosphate-limited continuous cultures of *E. coli* W3110.shik1," *Biotechnology Progress*, vol. 22, no. 4, pp. 1056–1070, 2006.

[5] W. Wiechert, M. Möllney, S. Petersen, and A. A. De Graaf, "A universal framework for ^{13}C metabolic flux analysis," *Metabolic Engineering*, vol. 3, no. 3, pp. 265–283, 2001.

[6] A. Varma and B. O. Palsson, "Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors," *Journal of Theoretical Biology*, vol. 165, no. 4, pp. 477–502, 1993.

[7] G. J. E. Baart, B. Zomer, A. de Haan et al., "Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes," *Genome Biology*, vol. 8, no. 7, article R136, 2007.

[8] C. H. Schilling, D. Letscher, and B. O. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 229–248, 2000.

[9] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson, "Metabolic pathways in the post-genome era," *Trends in Biochemical Sciences*, vol. 28, no. 5, pp. 250–258, 2003.

[10] J. Maertens and P. A. Vanrolleghem, "Modeling with a view to target identification in metabolic engineering: a critical evaluation of the available tools," *Biotechnology Progress*, vol. 26, no. 2, pp. 313–331, 2010.

[11] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss, "Dynamic modeling of the central carbon metabolism of *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 79, no. 1, pp. 53–73, 2002.

[12] B. E. Peercy, S. J. Cox, S. Shalel-Levanon, K.-Y. San, and G. Bennett, "A kinetic model of oxygen regulation of cytochrome production in *Escherichia coli*," *Journal of Theoretical Biology*, vol. 242, no. 3, pp. 547–563, 2006.

[13] A. Varma and B. O. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110," *Applied and Environmental Microbiology*, vol. 60, no. 10, pp. 3724–3731, 1994.

[14] R. Mahadevan, J. S. Edwards, and F. J. Doyle III, "Dynamic flux balance analysis of diauxic growth in Escherichia coli," *Biophysical Journal*, vol. 83, no. 3, pp. 1331–1340, 2002.

[15] R.-Y. Luo, S. Liao, G.-Y. Tao et al., "Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions," *Molecular Systems Biology*, vol. 2, article 71, 2006.

[16] D. Segrè, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, pp. 15112–15117, 2002.

[17] J. M. Lee, E. P. Gianchandani, J. A. Eddy, and J. A. Papin, "Dynamic analysis of integrated signaling, metabolic, and regulatory networks," *PLoS Computational Biology*, vol. 4, no. 5, Article ID e1000086, 2008.

[18] A. Buchholz, J. Hurlebaus, C. Wandrey, and R. Takors, "Metabolomics: quantification of intracellular metabolite dynamics," *Biomolecular Engineering*, vol. 19, no. 1, pp. 5–15, 2002.

[19] A. Hoque, H. Ushiyama, M. Tomita, and K. Shimizu, "Dynamic responses of the intracellular metabolite concentrations of the wild type and *pykA* mutant *Escherichia coli* against pulse addition of glucose or NH_3 under those limiting continuous cultures," *Biochemical Engineering Journal*, vol. 26, no. 1, pp. 38–49, 2005.

[20] J. Schaub, K. Mauch, and M. Reuss, "Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ^{13}C labeling data," *Biotechnology and Bioengineering*, vol. 99, no. 5, pp. 1170–1185, 2008.

[21] A. Novick and L. Szilard, "Description of the chemostat," *Science*, vol. 112, no. 2920, pp. 715–716, 1950.

[22] G. Lidén, "Understanding the bioreactor," *Bioprocess and Biosystems Engineering*, vol. 24, no. 5, pp. 273–279, 2001.

[23] M. Zinn, B. Witholt, and T. Egli, "Dual nutrient limited growth: models, experimental observations, and applications," *Journal of Biotechnology*, vol. 113, no. 1-3, pp. 263–279, 2004.

[24] T. Egli, "On multiple-nutrient-limited growth of microorganisms, with special reference to dual limitation by carbon and

nitrogen substrates," *Antonie van Leeuwenhoek*, vol. 60, no. 3-4, pp. 225–234, 1991.

[25] M. De Mey, G. J. Lequeux, J. J. Beauprez et al., "Comparison of different strategies to reduce acetate formation in *Escherichia coli*," *Biotechnology Progress*, vol. 23, no. 5, pp. 1053–1063, 2007.

[26] E. Jones, T. Oliphant, P. Peterson, et al., "SciPy: Open source scientific tools for Python," (2001–2007), http://www.scipy.org/.

[27] F. Pérez and B. E. Granger, "IPython: a system for interactive scientific computing," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21–29, 2007.

[28] N. S. Wang and G. Stephanopoulos, "Application of macroscopic balances to the identification of gross measurement errors," *Biotechnology and Bioengineering*, vol. 25, no. 9, pp. 2177–2208, 1983.

[29] Q. Hua, C. Yang, T. Baba, H. Mori, and K. Shimizu, "Response of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts," *Journal of Bacteriology*, vol. 185, no. 24, pp. 7053–7067, 2003.

[30] I. Nookaew, M. C. Jewett, A. Meechai et al., "The genome-scale metabolic model *iIN800* of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism," *BMC Systems Biology*, vol. 2, article 71, 2008.

[31] A. Provost and G. Bastin, "Dynamic metabolic modelling under the balanced growth condition," *Journal of Process Control*, vol. 14, no. 7, pp. 717–728, 2004.

[32] M. R. Atkinson, T. A. Blauwkamp, V. Bondarenko, V. Studitsky, and A. J. Ninfa, "Activation of the *glnA*, *glnK*, and *nac* promoters as *Escherichia coli* undergoes the transition from nitrogen excess growth to nitrogen starvation," *Journal of Bacteriology*, vol. 184, no. 19, pp. 5358–5363, 2002.

[33] A. Narang, A. Konopka, and D. Ramkrishna, "New patterns of mixed-substrate utilization during batch growth of *Escherichia coli* K12," *Biotechnology and Bioengineering*, vol. 55, no. 5, pp. 747–757, 1997.

[34] J. B. Russell and G. M. Cook, "Energetics of bacterial growth: balance of anabolic and catabolic reactions," *Microbiological Reviews*, vol. 59, no. 1, pp. 48–62, 1995.

[35] T. Müller, B. Walter, A. Wirtz, and A. Burkovski, "Ammonium toxicity in bacteria," *Current Microbiology*, vol. 52, no. 5, pp. 400–406, 2006.

[36] D. T. Britto, M. Y. Siddiqi, A. D. M. Glass, and H. J. Kronzucker, "Futile transmembrane NH_4^+ cycling: a cellular hypothesis to explain ammonium toxicity in plants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 7, pp. 4255–4258, 2001.

[37] S. Kustu, J. Hirschman, D. Burton, J. Jelesko, and J. C. Meeks, "Covalent modification of bacterial glutamine synthetase: physiological significance," *Molecular & General Genetics*, vol. 197, no. 2, pp. 309–317, 1984.

[38] M. R. Atkinson and A. J. Ninfa, "Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*," *Molecular Microbiology*, vol. 29, no. 2, pp. 431–447, 1998.

[39] B. Friedrich and B. Magasanik, "Urease of *Klebsiella aerogenes*: control of its synthesis by glutamine synthetase," *Journal of Bacteriology*, vol. 131, no. 2, pp. 446–452, 1977.

[40] E. Soupene, W. C. van Heeswijk, J. Plumbridge et al., "Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression," *Journal of Bacteriology*, vol. 185, no. 18, pp. 5611–5626, 2003.

[41] A. Death and T. Ferenci, "Between feast and famine: endogenous inducer synthesis in the adaptation of *Escherichia coli* to growth with limiting carbohydrates," *Journal of Bacteriology*, vol. 176, no. 16, pp. 5101–5107, 1994.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 753904, 13 pages
doi:10.1155/2010/753904

Review Article

Which Metabolic Pathways Generate and Characterize the Flux Space? A Comparison among Elementary Modes, Extreme Pathways and Minimal Generators

Francisco Llaneras and Jesús Picó

Instituto de Automática e Informática Industrial, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Correspondence should be addressed to Francisco Llaneras, kikollan@gmail.com

Received 18 September 2009; Revised 29 December 2009; Accepted 11 February 2010

Academic Editor: Ryan Senger

Important efforts are being done to systematically identify the relevant pathways in a metabolic network. Unsurprisingly, there is not a unique set of network-based pathways to be tagged as relevant, and at least four related concepts have been proposed: extreme currents, elementary modes, extreme pathways, and minimal generators. Basically, there are two properties that these sets of pathways can hold: they can generate the flux space—if every feasible flux distribution can be represented as a nonnegative combination of flux through them—or they can comprise all the nondecomposable pathways in the network. The four concepts fulfill the first property, but only the elementary modes fulfill the second one. This subtle difference has been a source of errors and misunderstandings. This paper attempts to clarify the intricate relationship between the network-based pathways performing a comparison among them.

1. Introduction

A metabolic network can be represented with a stoichiometric matrix **N**, where rows correspond to the m metabolites and columns to the n reactions. Assuming that intracellular metabolites are at steady state, material balances can be formulated as follows [1]:

$$\mathbf{N} \cdot \mathbf{v} = 0, \tag{1}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_n)^{\mathrm{T}}$ is the n-dimensional vector of flux through each reaction. Each feasible steady state is represented by a flux vector **v**. Taking into account these mass balances and the irreversibility of certain reactions, the space of feasible steady state *flux distributions,* or *flux space*, can be defined as follows (see glossary for words in italics):

$$\mathbf{P} = \{\mathbf{v} \in R^n : \mathbf{N} \cdot \mathbf{v} = 0,\ \mathbf{D} \cdot \mathbf{v} \geq 0\}, \tag{2}$$

where **D** is a diagonal $n \times n$-matrix with $\mathbf{D_{ii}} = 1$ if the flux i is irreversible (otherwise 0).

The concept of the flux space is the cornerstone of constraint-based modeling, an approach supported by the fact that cells are subject to governing constraints that limit their behavior [2, 3]. In this context, network-based pathways are used to investigate the modeled metabolism by the analysis of a finite set of relevant pathways, which ideally represents all of the metabolic states that a cell can show. Some outstanding applications of this approach are enumerated in Table 1.

However, there is not a unique set of network-based pathways to be tagged as "relevant" and different proposals have been applied with success: extreme currents, elementary modes, extreme pathways, and minimal generators. These concepts are not equivalent, but closely related. There are three major properties that a set of network-based pathways can hold: (P1) they can generate the flux space **P**, (P2) they can be the minimal set of vectors fulfilling the first property, and (P3) they can be all the non-decomposable pathways in the network. The fact that all of the network-based pathways—elementary modes, extreme pathways, and

Table 1: Applications of network-based pathways analysis. Partially extracted from [4–6].

Applications	References
Identification of pathways	[7, 8]
Determination of minimal medium requirements	[9]
Analysis of pathway redundancy and robustness	[10–12]
Linkage between structure and regulation...	
Correlated reactions (enzyme subsets)	[11, 13]
Detect excluding reaction pairs	[4]
Prediction of transcription ratios	[10, 14]
Include regulatory rules	[15]
Support for metabolic engineering...	
Identification of pathways with optimal yields	[8]
Evaluation of effect of addition/deletion of genes	[16]
Inference of viability of mutants	[10, 17]
Detection of minimal cut sets	[18]
Suggest operations to increase product yield	[19]
Translation of a flux distribution into pathways activities...	
Particular solution methods	[20, 21]
Alpha-spectrum	[22, 23]
Aid in the reconstruction of metabolic reaction networks...	
Assignment of function to orphan genes	[24]
Detection of infeasible circles	[12, 25]
Detection of network dead ends	[9, 26, 27]
Support in the reconstruction of metabolic maps	[28]
Development of reduced, kinetic models	[29–31]

so forth—fulfill the first property but not the others has been a source of errors, imprecisión, and misunderstandings.

This paper discusses the relationship between the different network-based pathways from a theoretical point of view. We will start defining four pathway concepts and then we will perform a comparison among them. Finally, we will present some examples and outline the major conclusions.

2. Results and Discussion

The first attempts to systemically extract a set of pathways from a given metabolic network were based on the assumption that all of the fluxes were irreversible, or more precisely, that its dominant direction could be presumed. Convex algebra shows that in this case the flux space **P** is a *pointed convex polyhedral cone* in the positive orthant R^n_+, which can be generated by non-negative combination of certain vectors, its edges, or *extreme rays* [32]. See Figure 1 for a geometric illustration of the concept.

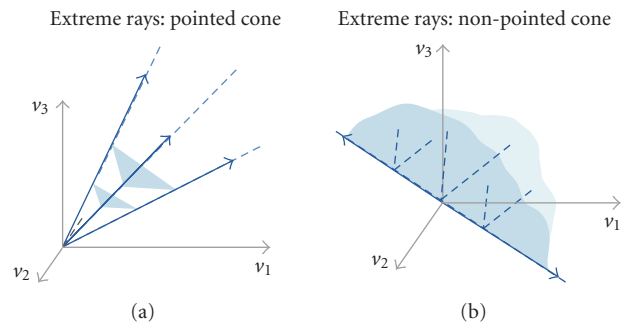


Figure 1: Extreme rays of two flux spaces.

These extreme rays were flux vectors, or pathways, with a remarkable property (P1): the extreme rays generate the flux space **P**; that is, every flux distribution **v** in **P** can be represented as a non-negative combination of fluxes through these pathways ($\mathbf{e_k}$ denotes the extreme rays):

$$\mathbf{P} = \left(\mathbf{v} : \mathbf{v} = \sum_{k}^{e} w_k \cdot \mathbf{e_k},\ w_k \geq 0 \right). \tag{3}$$

Notice that, in general, a given **v** cannot be uniquely decomposed into an activity pattern **w**, but a space of valid solutions exists [22, 23]. This is also true for the rest of generating sets that will be introduced in subsequent sections.

Moreover, the set of extreme rays had two additional properties: (P2) it was the smallest (minimal) generating set of **P**, and (P3) the extreme rays were all the *non-decomposable* vectors in **P**, those that cannot be decomposed in simpler vectors [6]. A *non-decomposable* vector is a minimal set of reactions that form a "functional unit"; if any of its participant reactions is not carrying flux, the others cannot operate alone. These functional units are the simplest steady-state flux distributions that cells can show, and the rest of feasible states can be seen as the aggregated action of these units. This property makes it possible to investigate the infinite behaviors that cells can show by inspection of the finite set of non-decomposable vectors.

But what happens if not all fluxes can be assumed to be irreversible? If so, the extreme rays may lose these properties. Moreover, a set of vectors holding the three properties simultaneously (P1, P2, and P3) will not exist; there will be sets fulfilling P1 and P2, or P1 and P3, but not P2 and P3 in a general case.

3. Different Network-Based Pathways

3.1. Extreme Currents. Extreme currents are probably the first attempt to define a set of network-based pathways [33]. Their computation is based on splitting up each reversible reaction into two irreversible ones. Thus, if fluxes are reordered to separate the irreversible fluxes $\mathbf{v_I}$ and the

reversible ones $\mathbf{v_R}$, the flux space (2) is augmented ($\mathbf{N} = [\mathbf{N_I N_R}]$):

$$\mathbf{P_{rc}} = \left\{ \mathbf{v} \in \mathbf{R}^{n+r} : \begin{pmatrix} \mathbf{N_I} & \mathbf{N_R} & -\mathbf{N_R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v_I} \\ \mathbf{v_R} \\ \mathbf{v'_R} \end{pmatrix} = \mathbf{0}, \begin{pmatrix} \mathbf{v_I} \\ \mathbf{v_R} \\ \mathbf{v'_R} \end{pmatrix} \geq \mathbf{0} \right\}. \tag{4}$$

The extreme rays of the cone $\mathbf{P_{rc}}$ are defined as the extreme currents of $\mathbf{P}$. Notice that $\mathbf{P_{rc}}$ is a pointed cone in the positive orthant $\mathbf{R}^{n+r}$, so its extreme rays have all of the properties mentioned above (P1–P3). However, $\mathbf{P_{rc}}$ lives in a higher-dimensional vector space (augmented in one dimension for each split reversible reaction) and the extreme currents lose their properties when they are translated to the original vector-space.

In fact, it has been recently shown that the set of extreme currents (ECS) coincides with the set of elementary modes, which will be introduced below, when it is translated to the original vector-space [34]—when computing the ECS a set of r spurious cycles appear (pathways formed by the forward and backward reaction of each reversible flux); however, these pathways are not considered meaningful in most applications [35] and they disappear when the ECs are expressed in the original vector-space $\mathbf{R}^n$.

3.2. Elementary Modes. The concept of elementary modes was introduced to extend the property of non-decomposability of the extreme rays (P3) to networks with reversible fluxes [7, 8]. Hence, a flux vector $\mathbf{e}$ is an elementary mode (EM) if and only if [36]

(C1) $\mathbf{e} \in \mathbf{P}$,

(C2) there is no nonzero vector $\mathbf{v} \in \mathbf{P}$ such that the support of $\mathbf{v}$ supp($\mathbf{v}$) is a proper subset of the support of $\mathbf{e}$ supp($\mathbf{e}$). In other words, $\mathbf{e}$ cannot be decomposed as a positive combination of two "simpler" vectors $\mathbf{v'}$ and $\mathbf{v''}$ in $\mathbf{P}$ that contain zero elements wherever $\mathbf{e}$ does and include at least one additional zero component each. This condition is the so-called nondecomposability, simplicity, or genetic independence.

Thereby, the set of elementary modes (EMS) is defined as the set of all the nondecomposable vectors in the flux space (P3). This definition implies that the EMS fulfills property P1, as in (3), but also a more restrictive condition due to C2: each flux distribution can always be represented as a non-negative combination of elementary modes without cancelations [36]:

$$\mathbf{P} = \left(\mathbf{v} : \mathbf{v} = \sum_{k}^{e} w_k \cdot \mathbf{e_k}, w_k \geq 0 \right) \text{ without cancelations } (^*). \tag{5}$$

(*) *If the sum runs over two or more indices* $\mathbf{k}$, *all of the* $\mathbf{e_k}$ *have zero components wherever* $\mathbf{v}$ *has zero components and include at least one additional zero each.*

That means that the elementary modes are all the simple states, or functional units, that a cell can show (the non-decomposable vectors) and the rest of feasible states can be seen as its strictly aggregated action, that is, its aggregated action without cancelations. The "no cancelation rule" is relevant for several applications of network-based pathways. The no cancelation rule is what makes it possible to investigate the infinite behaviors that cells can show by simply inspection of the finite set of elementary modes, because there is no possibility of cancelations of reversible fluxes. This allows to answer many interesting questions in an easy manner; consider, for example the following:

(i) Which reactions are essential to produce the compound Y? Those that participate in all of the elementary modes producing Y.

(ii) Is there a route connecting the educt A with the product Y? Only if there is an elementary mode connecting them.

(iii) Which are the capabilities of the network if a reaction r is not carrying flux or has been knocked-out? The feasible states in these circumstances are only those that result from aggregating, with no cancelations, the elementary modes not involving r (i.e., the consequences of r not carrying flux can be directly predicted ignoring the elementary modes participated by r).

(iv) Which is the optimal yield to produce Y from A? The (stoichiometrically) optimal pathway is the elementary mode consuming A and producing Y with the best yield.

As we will see in subsequent sections, the main difference among network-based pathways is that all of them satisfy (3), but only the elementary modes satisfy (5), and this fact determines their practical applications.

3.3. Minimal Generators. We have seen that the elementary modes generate the flux space, as in (3), but usually they are not the smallest set satisfying this condition because they have to fulfill the most exigent condition of (5). Which is then the minimal set of vectors that generates $\mathbf{P}$ by non-negative combination? The term minimal generating set (MGS) has been recently coined to refer to this set [37]. It was also shown how to obtain an MGS that is subset of EMS. However, in general there is not a unique minimal generating set: different MGSs may exist within the EMS, and even vectors that are not EMs can be part of an MGS. Both cases will be discussed in following sections. Yet, the concept of the minimal generating set also arises from a different point of view. It is well known that the elementary modes are not *systemically independent* because it is possible to represent some modes as non-negative combination of others [5]. Clearly, dependent modes that are not necessary to fulfill (3) can be removed. Thus, any resultant *irreducible subset* of the elementary modes is a minimal generating set.

In summary, a set of minimal generators fulfill properties P1 and P2, whereas the elementary modes fulfill P1 and P3. The elementary modes include additional non-decomposable vectors to fulfill P3, which are redundant in (3) but necessary in (5).

The fact that an MGS does not fulfill (5) reduces its usability for analysis of the underlying metabolism. Remarkably, the questions mentioned in the previous section cannot be easily addressed using the MGS because the cancelation of reversible fluxes hides simple pathways. For example, the MGS has to be recalculated after a gene deletion, and similar difficulties arise in other applications. The advantage of the MGSs against the EMS is its reduced size: considering the central carbon metabolism of *E. coli*, the computation of the EMS returns more than 500 000 EMs, whereas an MGS contains around 3000 MGs [34]. This also implies that obtaining the MGS is computationally more efficient. Thereby, the MGS will be preferred in those applications that just require a set of vectors generating the fluxspace. For instance, the MGS has been used to perform phenotype phase-plane analysis [37] and it can be used to extract the minimal connections between extracellular compounds, information that can then be used to develop unstructured, kinetic models [29–31, 38].

3.4. Extreme Pathways. As it happens with the extreme currents, extreme pathways are obtained in an augmented vector-space [35]; however, only the internal fluxes are decomposed in both forward and backward directions (the exchange fluxes, those that connect internal and external metabolites with one-to-one correspondence [4], are kept as reversible). Hence, if fluxes are reordered to separate the irreversible internal fluxes $\mathbf{v_I}$, the reversible ones $\mathbf{v_R}$, and the exchange fluxes $\mathbf{v_B}$, as $\mathbf{v} = [\mathbf{v_I}\ \mathbf{v_B}\ \mathbf{v_R}]^{\mathbf{T}}$, the flux space (2) can be reformulated as follows (where $\mathbf{N} = [\mathbf{N_I}\ \mathbf{N_B}\ \mathbf{N_R}]$):

$$\mathbf{P_{rc}} = \left\{ \mathbf{v} \in \mathbf{R}^{n+r} : \begin{pmatrix} \mathbf{N_I} & \mathbf{N_B} & \mathbf{N_R} & -\mathbf{N_R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v_I} \\ \mathbf{v_B} \\ \mathbf{v_R} \\ \mathbf{v'_R} \end{pmatrix} = \mathbf{0}, \begin{pmatrix} \mathbf{v_I} \\ \mathbf{v_B} \\ \mathbf{v_R} \\ \mathbf{v'_R} \end{pmatrix} \geq \mathbf{0} \right\}. \tag{6}$$

In this augmented vector-space, and only there, the set of extreme pathways (EPs) is a subset of the elementary modes that is systemically independent [5]. The EPs are still capable of generating **P**, as in (3), because only dependent elementary modes are discarded. However, the extreme pathways are not systemically independent in the original one (and even the ECs, which are equivalent to the EMs, are systemically independent in the augmented space where they are obtained). Therefore, they are not the irreducible subset of the elementary modes in the original vector-space; that is, they are not the minimal generating set [37]. Unfortunately, this notion was unclear in the literature until recently.

The extreme pathways fulfill property P1, but not P2 nor P3 in the original vector-space. As it happens with the MGS, the fact that the EPS does not fulfill (5) reduces its usability in certain applications. Their advantage with respect to the EMS is its smaller size, but it must be kept in mind that, in general, the MGS will be smaller than the EPS (and never larger).

Example: Two Different Vector-Spaces. Consider the small network depicted in Figure 2, Case 2A. The three EPs of this network represented in the augmented vector-space $\{v_1, v_2, v_3, -v_3\}$ are E1 = (1 0 1 0), E2 = (0 1 0 1), and E3 = (1 1 0 0). These three vectors are systemically independent. However, when translated to the original vector-space $\{v_1, v_2, v_3\}$, these vectors are E1 = (1 0 1), E2 = (0 1 −1) and E3 = (1 1 0), which are not longer systemically independent, since E1 = E2 + E3. Figure 2 also illustrates the systemic dependancy of the EPs.

4. Comparison among Network-Based Pathways

This section is devoted to the comparison of the network-based pathways described above: extreme currents, minimal generators, elementary modes and extreme pathways. The case where all of the fluxes are irreversible will be introduced first to contextualize the problem; then, the presence of reversible fluxes will be considered and the differences will become apparent (see Figure 2).

Reference Vector-Space. Hereinafter we consider the original vector-space as the reference one: all of the generating sets will be expressed as elements of the vector-space $\mathbf{R^n}$ where each flux corresponds to an axis. We choose $\mathbf{R^n}$ because it is the original space of the fluxes that connect the metabolites of the network, and thus it is the meaningful one. For instance, in the previous example the EPs expressed in the augmented vector-space were unable to capture the fact that pathway E1 can be seen as a combination of E2 and E3 (E1 = E2 + E3). Notice also that the relevant difference between equations (3) and (5), which depends on the cancelation of reversible fluxes, cannot be easily observed in the augmented vector-spaces. Since ECs and EPs are computed in augmented vector-spaces, once obtained, they have to be translated to $\mathbf{R^n}$, simply merging again the decomposed reversible fluxes. This process also removes the spurious cycles, those pathways formed only by the forward and backward reaction of each reversible flux and appearing as EPs and ECs in the augmented vector-spaces.

4.1. Case 1: All Fluxes Are Irreversible. As explained in a previous section, when all of the reactions are irreversible, the flux space **P** is a convex cone that satisfies two conditions: (a) it is in the positive orthant $\mathbf{R}^+$ and (b) it is a *pointed cone*.

Condition (b) implies that **P** can be generated by non-negative combination of its extreme rays (3) (more details below). In fact, the extreme rays always belong to every generating set because by definition they cannot be generated by non-negative combination of other vectors within the cone (see glossary). Thus, if the extreme rays are able to generate the cone, as it happens in this case, they are necessarily the minimal generating set. On the

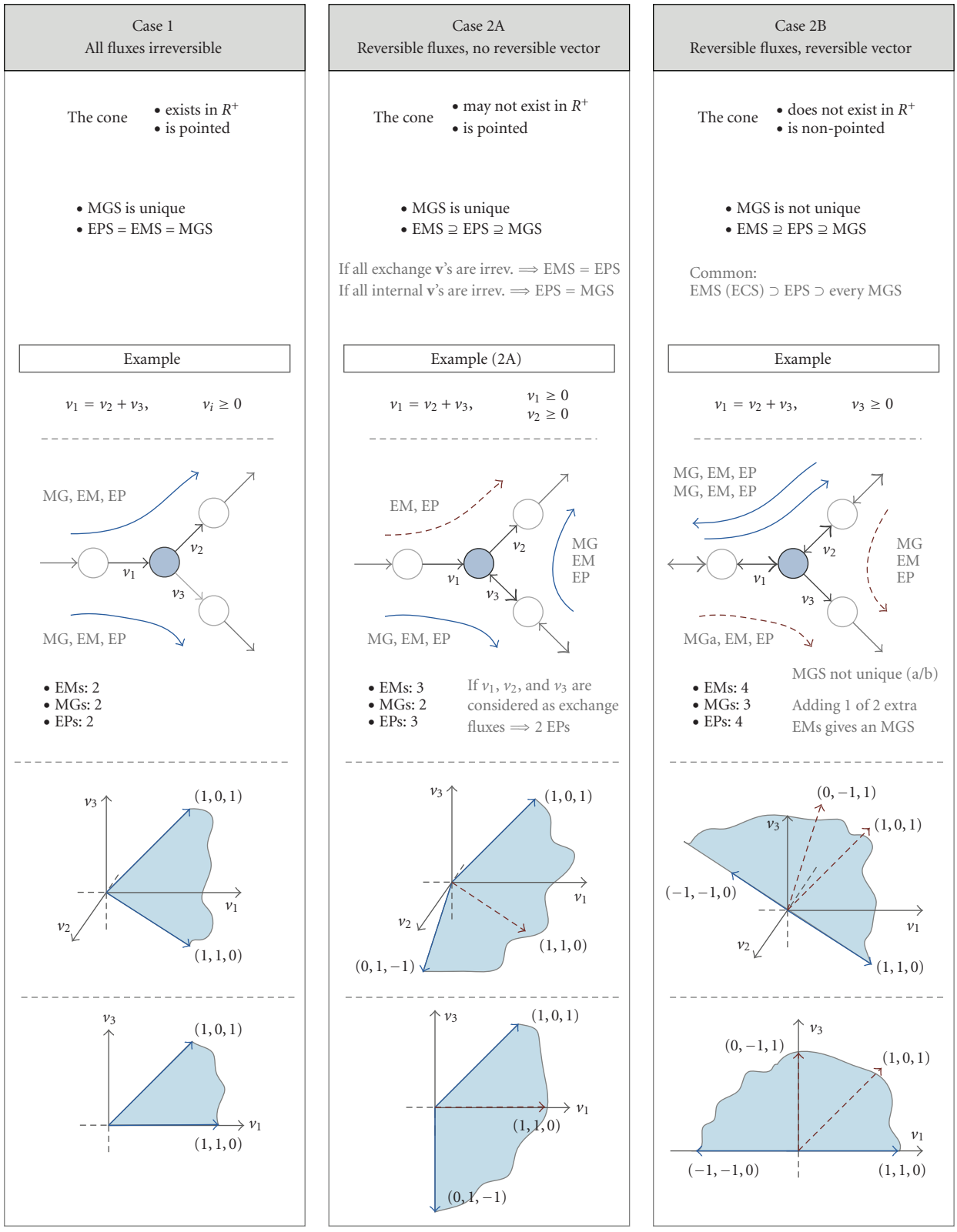


Figure 2: Case-based scheme of the different network-based pathways. In each example metabolites are represented with circles connected with thin arrows that represent the fluxes. The reversible fluxes are double arrowed (solid arrowhead defines the sign criteria). The blue thick arrows denote generating vectors that correspond to extreme rays of the cone and the red ones to the rest of generating vectors. The axis at the bottom depicts the flux-space over $\{v_1, v_2, v_3\}$, blue area, and its generating vectors.

other hand, the extreme rays are always non-decomposable vectors of **P** (since they cannot be generated by non-negative combination). Moreover, condition (a) implies that the intersections of the cone with the (positive or negative) axis of the vector-space, which are potential non-decomposable vectors, cannot be interior points of **P**. Thus, the extreme rays will be all the non-decomposable vectors in **P**.

These two conditions imply that in this case the extreme rays are not only the minimal generating set of the flux space (P1 and P2), but also the set of all non-decomposable vectors (P3). Since the ECs and the EPs are the extreme rays of two cones defined in augmented vector-spaces where the reversible fluxes are decomposed, it is obvious that, if there are no reactions to be decomposed, the ECs and EPs are the extreme rays of the original cone **P**. Therefore the following is maintained.

Rule 1. If all fluxes are irreversible, all the generating sets are equivalent, EMS = ECS = EPS = MGS, and coincide with the extreme rays of the flux space **P**.

4.2. Case 2: There Are Reversible Fluxes. Now we consider the situation where certain fluxes are reversible. The flux space **P** is still a convex cone, but it is not necessarily in the positive orthant $\mathbf{R}^+$ and it can be non-pointed. If one or more reversible reactions are effectively reversible—that is, both forward and backward directions can be realized by flux distributions—the cone will not be in the positive orthant (otherwise **P** would remain a pointed one in the positive orthant as in Case 1). Then, two situations are possible: Case 2A, the cone is pointed, and Case 2B, it is not.

Consider the *lineality space* of **P**, defined as $\text{lin.space}(\mathbf{P}) := \{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{A} \cdot \mathbf{x} = \mathbf{0}\}$. It represents the linear subspace contained in the cone and allows to characterize the cone as follows: **P** *is pointed if lin.space* $(\mathbf{P}) = \{0\}$, *otherwise non-pointed*. Hence, **P** will be a non-pointed cone if a vector **x** and its opposite **–x** exist in **P**. These vectors would involve only reversible fluxes and represent *reversible vectors* that can operate in both directions. Thus **P** is non-pointed cone if and only if it contains a reversible vector. It is also possible to check whether a cone is pointed inspecting **K**, the kernel of **N**, arranged in a suitable way (see [37] for details).

The more important consequence of this classification is the following: a pointed cone **P** can be generated by non-negative combination of its extreme rays, but this no longer true for a non-pointed one. A non-pointed cone still can be generated by non-negative combination, but a unique MGS will not exist.

4.3. Case 2A: Reversible Fluxes but Not Reversible Vectors. If there are reversible fluxes but not a reversible vector, the flux-space **P** is still a pointed cone and it can be generated by its extreme rays [39]. As explained above, if the extreme rays generate the cone, they are necessarily the minimal generating set because they belong to every generating set by definition.

Rule 2. If the flux space **P** does not contain a reversible vector, a unique MGS exists and it coincides with the extreme rays of **P**.

However, if there are reversible fluxes, and they are effectively used in both directions, the cone is not restricted to the positive orthant $\mathbf{R}^+$. This implies that the intersections of vector-space axis with the cone will be non-decomposable vectors of **P**. That is, there are non-decomposable vectors in **P** that are not extreme rays. The EMS sill contains the extreme rays, which are always non-decomposable, but could also contain other non-decomposable vectors. Notice that these extra EMs are necessary to generate the flux space **P** without cancelations (5), but can be redundant to fulfill (3).

Rule 3. The EMS (ECS) is always a superset of the extreme rays of the flux space **P**. If there are reversible fluxes, more EMs than extreme rays may exist.

By Rules 2 and 3 it follows that, if the flux space **P** does not contain a reversible vector, the MGS is a subset of the EMS. Moreover, those EMs not belonging to the MGS will be systemically dependent and the MGS will be the unique irreducible subset of the EMS.

Rule 4. If the flux space **P** does not contain a reversible vector, the unique MGS is the irreducible subset of the EMS. It can be extracted from the EMS selecting the systemically independent vectors (see the appendix).

This property was incorrectly assigned to the extreme pathways in the past, but these are systemically independent only in an augmented vector-space and not in the original one (see example below). The EPs are the extreme rays of the cone obtained when the internal, reversible reactions are split, whereas the EMs (ECs) are the extreme rays of the cone obtained when *all* of the reversible reactions are split. This difference determines the relationship among the concepts (Figure 3).

Rule 5. If the flux space **P** does not contain a reversible vector, the EPS can be a subset of the EMS, but in general it is not the MGS. That is, EMS (ECS) $\supseteq$ EPS $\supseteq$ MGS, and the following two particular cases exist.

(a) If all **exchange** fluxes are irreversible, EMS (ECS) = EPS.

(b) If all **internal** fluxes are irreversible, EPS = MGS.

The two rules can be rephrased as follows

(a) EPS can be a proper subset of the EMS $\Longleftrightarrow$ there are reversible exchange fluxes.

(b) MGS can be a proper subset of the EPS $\Longleftrightarrow$ there are reversible internal fluxes.

Proof Outline. (a) If all of the reversible fluxes are internal, the EPs and the ECs (EMs) are the extreme rays of the same cone. (b) If all of the internal fluxes are irreversible, the EPs are the extreme rays of the original cone, which coincide with the extreme rays due to Rule 2.

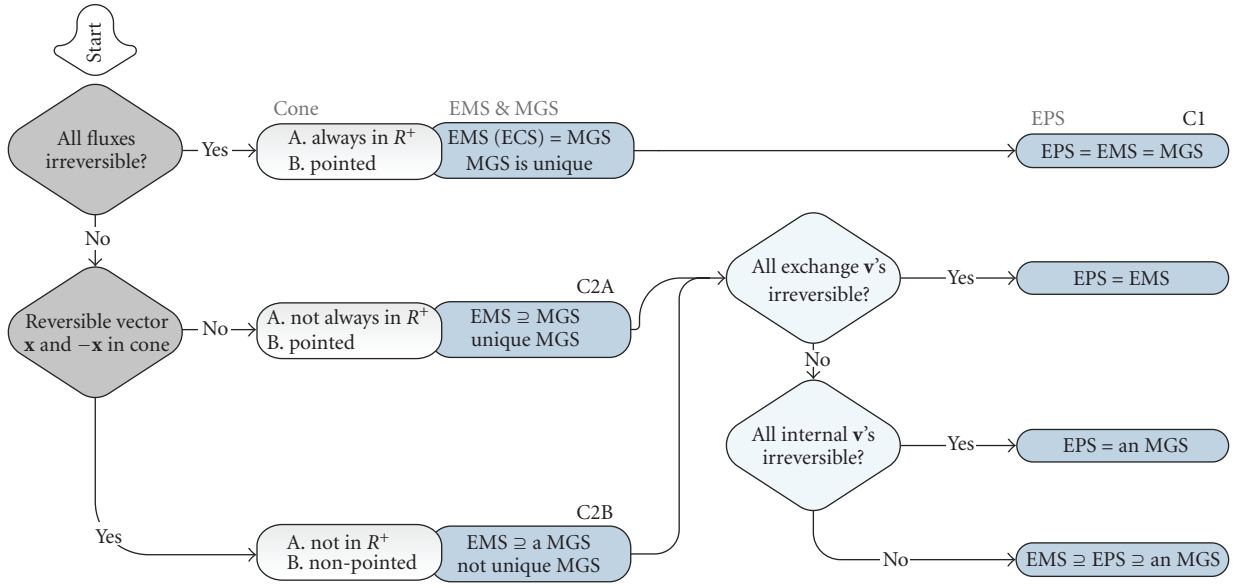


Figure 3: Relationship between different network-based pathways.

4.4. Case 2B: Reversible Fluxes and a Reversible Vector. If the reversible fluxes form a reversible vector, the convex cone $\mathbf{P_r}$ is non-pointed. A non-pointed cone can be represented as $\mathbf{P_r} = \mathbf{H} + \mathbf{Q}$, where $\mathbf{H}$ is the linear space lin.space($\mathbf{P_r}$), and $\mathbf{Q}$ is a pointed subcone, with $\mathbf{Q} \subseteq \mathbf{H}^{\perp}$ ($\mathbf{H}^{\perp}$ denotes the orthogonal complement of $\mathbf{H}$). In fact, this is the general representation of a convex polyhedral cone and Cases 1 and 2A are particular cases where $\mathbf{H} = \{0\}$. Thus, a non-pointed cone can be generated as follows [39]:

$$\mathbf{P_r} = \left\{ \mathbf{v} : \mathbf{v} = \sum_{k}^{nf} \lambda_k \cdot \mathbf{f_k} + \sum_{j}^{nb} \beta_j \cdot \mathbf{x_j},\ \lambda_k \geq 0 \right\}, \tag{7}$$

where $\mathbf{f_k}$ are the "irreversible" generating vectors, for which its opposites are not contained in $\mathbf{P_r}$, and $\mathbf{x_j}$ are the "reversible" ones, for which its opposite $-\mathbf{x_j}$ is also contained in $\mathbf{P_r}$. Vectors $\mathbf{x_j}$ must form a base of $\mathbf{H}$, whereas vectors $\mathbf{f_k}$ generate the sub-cone $\mathbf{Q}$. Notice that $\mathbf{P_r}$ can still be generated by non-negative combination, as in (3), using $\mathbf{f_k}$, $\mathbf{x_j}$, and $-\mathbf{x_j}$ as generating vectors. Unfortunately, there is a price to pay for the cone being non-pointed: the set of minimal generating vectors is not unique anymore.

In fact, a minimal generating set of $\mathbf{P_r}$ is obtained choosing an arbitrary base $\{\mathbf{x_j}\}$ of $\mathbf{H}$, and taking one arbitrary ray $\mathbf{f_k}$ from each minimal proper face of the cone [39]. When the cone is pointed, there are no vectors $\{\mathbf{x_j}\}$ and the minimal proper faces are the extreme rays, so they are uniquely defined.

The extreme rays of $\mathbf{P_r}$ will be present in any generating set—EMS, EPS, or an MGS—because they cannot be represented as non-negative combination of other vectors in $\mathbf{P_r}$. However, they are insufficient to generate a non-pointed cone, they could even not exist (e.g., if all fluxes are reversible, the cone is an n-dimensional vector-space generated only by vectors $\mathbf{x_j}$ and $-\mathbf{x_j}$). Additional vectors $\{\mathbf{x_j}\}$ and $\{\mathbf{f_k}\}$ must be combined with the extreme rays to form an MGS, but the choice is not unique.

Rule 6. If the flux space $\mathbf{P_r}$ contains a reversible vector, its extreme rays are not a complete generating set and there is not a unique MGS.

However, it is still possible to define a MGS containing only non-decomposable vectors, and thus being a subset of the EMS. This kind of MGS can be obtained with a lexico-smallest representation [40] or extracted from the set of EMs, as explained in the appendix.

Rule 7. If the flux space $\mathbf{P_r}$ contains a reversible vector, an irreducible subset of the EMS constitutes an MGS formed only with non-decomposable vectors.

Notice that other MGSs exist. Indeed, even more than one MGS formed with different non-decomposable vectors may exist, since there is not necessarily a unique irreducible subset of EMS. Both situations will be illustrated in subsequent examples.

With respect to the EPS, Rule 5 should be rephrased recalling that the MGS is no longer unique. Moreover, since a reversible vector will be (a) always participated by at least one internal flux—a reversible vector only with external fluxes has little sense and (b) in most cases also participated by external ones (except that if all of the reversible vectors are futile cycles), a common situation arises where EMS (ECS) ⊃EPS ⊃ an MGS.

Rule 8. If the flux-space $\mathbf{P_r}$ contains a reversible vector, the EPS can be a subset of the EMS, but in general the EPS is not an MGS. The most common case will be EMS (ECS) ⊃EPS ⊃ an MGS.

4.5. Examples. Some examples will be used to illustrate the different cases described above. The first examples (1 to 5) use a simple network taken from Papin et al. [5]. The network has 6 reactions—3 internal and 3 exchange—and three metabolites, so it has 3 degrees of freedom. If all of the reactions were reversible, the kernel of **N** would provide a basis of the flux space formed by three reversible vectors. Herein we consider five cases where different reactions are irreversible (results are depicted in Figure 4).

Example 1. In the first example all fluxes are assumed to be irreversible (Case 1). In this case, the flux space is a pointed cone in $\mathbf{R}^+$ and all of the network-based pathways—ECS, EMS, EPS, and MGS—are equivalent.

Example 2. Now the exchange flux v_4 is assumed to be reversible. This example corresponds to Case 2A (the flux space is a pointed cone not in $\mathbf{R}^+$). In this case the EMS can be a superset of the MGS, as indeed happens in this example: EM4 is systemically dependent (EM4 = MG1 + MG2), so it is an EM but not a MG. On the other hand, the EPS is equal to the MGS because the internal fluxes are all irreversible. EM4 is not an EP because the reversible flux being cancelled in MG1 + MG2 is an exchange, so EM4 is systemically dependent in the vector-space where EPs are computed.

Example 3. In this third example the exchange flux v_4 and the internal flux v_2 are reversible. This is a general case and therefore, EMS $\supseteq$ EPS $\supseteq$ MGS. EM5 is neither an EP nor an MG (EM5 = MG1 + MG2). EM4 is not a MG (EM4 = MG3 + MG2), but it is an EP; one of the fluxes cancelled in MG3 + MG2 is an internal flux, so this cancelation cannot be done in the augmented vector-space where the EPs are computed.

Example 4. In this example only two internal fluxes, v_1 and v_3, are reversible. Again, the EMS is a superset of the MGS: EM4 is not an MG because it is systemically dependent (EM4 = MG3 + MG2). On the other hand, as all of the reversible fluxes are internal, the EPs and the EMs are necessarily equivalent.

Example 5. Now there are four reversible fluxes—v_1, v_2, v_5, and v_6— that define a reversible vector. This corresponds to Case 2B, where the flux space is a non-pointed cone. There are 7 EMs and 5 of them are also EPs. The two vectors that form the reversible vector are extreme rays in this example. To form an MGS they need to be combined with 2 other vectors, but the choice is not unique. For instance, 2 subsets of EMs are minimal generating sets, MGS1 and MGS2.

Example 6. Klamt and Stelling use a simple example, referred to as N2 in their article, to investigate the relationship between the EMS and the EPS [4]. This network has 9 reactions (3 exchanges) and 6 metabolites. After computing the EMS, the EPS, and the MGS, it turns out that there are 8 EMs and 5 EPs (the extra EM9/EP6 in [4] disappears in the original vector-space because it is a spurious cycle caused by decomposing the reversible fluxes). Yet, the MGS contains only 4 vectors, indicating that there is an EP that is not systemically independent: it can be checked by simple inspection that EP1 = EP2 + EP4 (when they are represented in the original vector-space).

Example 7. Another example to be analyzed is the small network used by Schilling et al. [35]. We obtained 7 EMs and the 5 relevant EPs given in the paper. Again, the EPs are not systemically independent when translated to the original vector-space (EP2 = EP3 + EP5) and 4 vectors are sufficient to form an MGS. It turns out that the MGS is not unique because there is a reversible vector in the flux-space (in fact, the reversible vector defines two EPs: EP3 and EP4 use the same reactions but in opposite directions).

Example 8. We have also analyzed the metabolic network of CHO cells given in [31]. The network has 24 reactions (9 reversible) and 18 internal metabolites, so it has 6 degrees of freedom. There are 18 EMs and 8 EPs, but only 6 vectors form the unique MGS (see supplementary file, Figure 2 in supplementry material available online at doi:10.1155/2010/753904). The metabolic pathways that correspond to the MGs are given in the supplementary file, Figure 1.

5. Conclusions

The purpose of network-based pathways analysis is to identify a finite set of systemic pathways in a metabolic network, and then use these pathways to study the cell metabolism. In this paper four similar definitions of network-based pathways have been compared.

We have seen that all of the flux states of a given metabolic network can be represented as an aggregation of flux through its elementary modes, which are all the simple, or non-decomposable, pathways in the network. Nevertheless, the set of elementary modes is not the smallest set of pathways fulfilling this property; this role corresponds to the so-called minimal generating sets. In certain cases there is a unique minimal generating set, but in general there are several alternatives. Interestingly, the set of elementary modes can be reduced by eliminating modes that are systemically dependent, resulting in a minimal generating set formed only with elementary modes. It has been also highlighted that, contrarily to what has sometimes been stated, the extreme pathways are not the minimal generating set, because they are usually systemically dependent in the original vector-space.

The minimal generating sets can be of use in applications where a set of generating vectors are required. In these cases they will be preferred due to its reduced size and because their computation is more efficient. For instance, minimal generators are suitable for extracting the fundamental connections between extracellular compounds, information that can be used to develop unstructured, kinetic models [29–31, 38]. However, the analysis of the elementary modes is more powerful. The fact that the set of elementary modes comprises all of the simple pathways in the network—its functional states—makes it possible to investigate the infinite

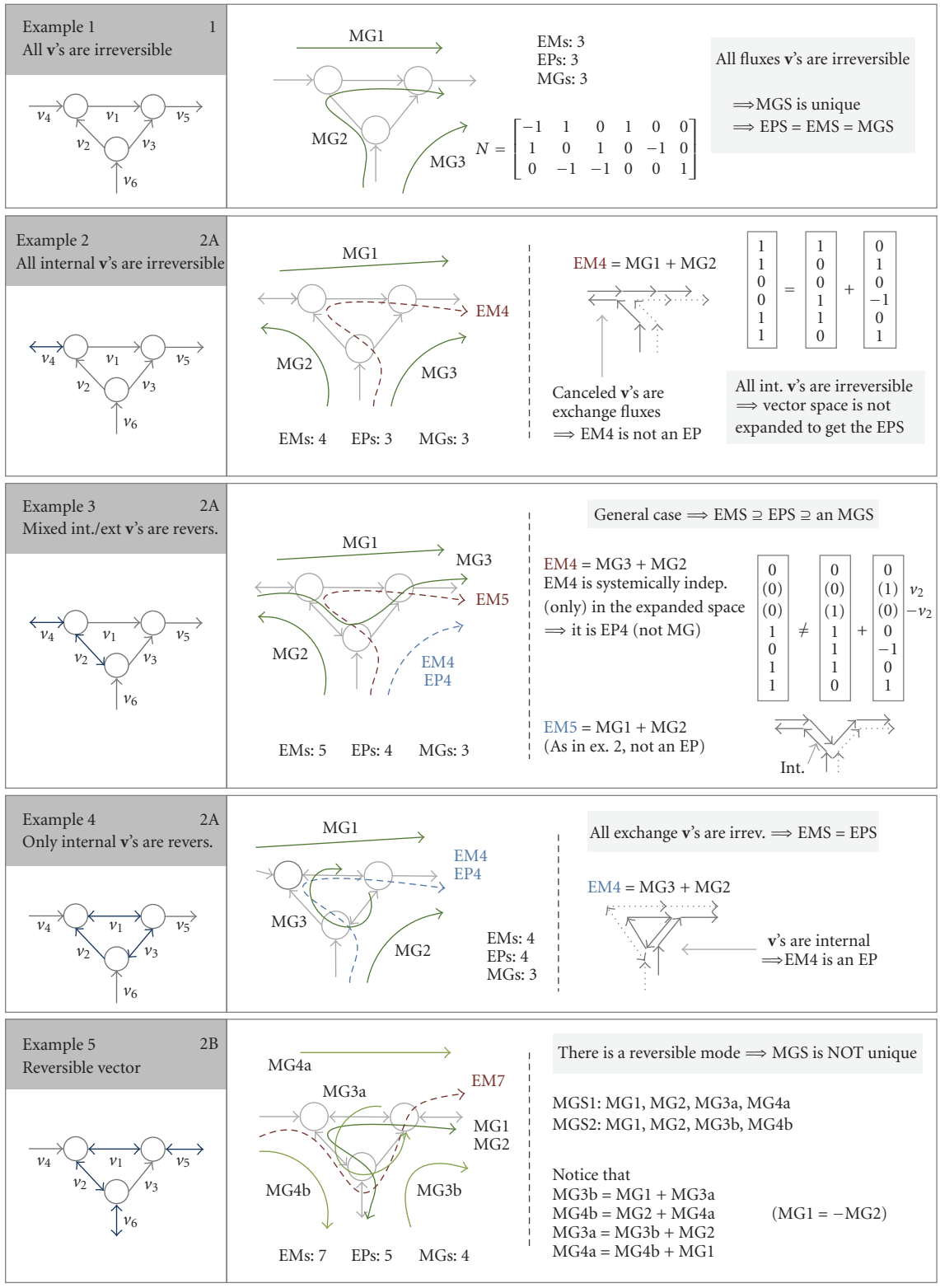


Figure 4: Examples illustrating the differences among network-based pathways.

behaviors that cells can show by simply inspecting them. This makes it easy to answer several questions: which reactions are essential to produce a certain compound, which will be the capabilities of the network if a reaction is knock-out, and so forth. Answering these questions using the minimal generators or the extreme pathways may be difficult because one has to take into account the possible cancelations of reversible fluxes.

Significant efforts are being done to improve network-based pathways analysis, particularly in the context of genome-scale metabolic network, where their more critical limitation appears: when the number of reactions in the network grows, the number of pathways dramatically increases, reducing understandability and even becoming not computable [5, 6]. Recent works have improved the computation algorithms [41, 42], and proposed methods to get particular subsets of pathways [43] and to decompose large networks in modules [44, 45]. New concepts of pathways have been also recently introduced. Kaleta et al. have introduced "elementary flux patterns", which explicitly takes into account possible steady-states fluxes through a genome-scale network when analyzing pathways through a subsystem, thus allowing the application of many (not all) elementary-mode-based tools to genome-scale networks [46]. Barrett et al. have used Monte Carlo sampling in conjunction with principal component analysis to obtain a low-dimensional set of pathways generating the flux space of genome-scale networks [47].

Most applications of network-based pathway analysis are found not only in the context of microbial production [9, 11, 12, 17, 20], but also in botany [48, 49] or in biomedicine [50, 51]. The number of applications increases steadily, and we believe that this will continue in the foreseeable future.

Glossary Box

Flux Distribution. The values of every metabolic flux of a given network at a particular (steady) state form a flux distribution.

Flux Space. The space $\mathbf{P}$ that contains all of the feasible flux distributions of a given metabolic network is the flux space. The flux space is often a convex polyhedral cone (2).

Convex Polyhedral Cone. A nonempty set of points $\mathbf{P} \subseteq \mathbf{R}^n$ is a convex cone if and only if any non-negative combination of elements from $\mathbf{P}$ remains in $\mathbf{P}$. A convex cone $\mathbf{P}$ is polyhedral if, for some matrix $\mathbf{A}$, $\mathbf{P} = \{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{A} \cdot \mathbf{x} \leq \mathbf{0}\}$. A convex polyhedral cone $\mathbf{P}$ is the set of solutions of a homogeneous system of inequalities (or the intersection of finitely many affine half-spaces).

Cone. For brevity, we use the term "cone" to refer to a convex polyhedral cone.

Nonnegative Generation. The Farkas theorem asserts that a convex cone is polyhedral if and only if it is finitely generated. A cone is finitely generated if there exist a set of vectors $\mathbf{G} = \{g_i\}$ that generate it by non-negative combination.

Generating Set. Any set of vectors $\mathbf{G}$ in $\mathbf{P}$ that generates $\mathbf{P}$ by non-negative combination is a generating set of $\mathbf{P}$. The EMS, the ECS, the EPS, and the MGSs are generating sets of $\mathbf{P}$.

Lineality Space. Let $\mathbf{P}$ be a convex polyhedral cone, $\mathbf{P} = \{\mathbf{x} \in \mathbf{R^n} \mid \mathbf{A} \cdot \mathbf{x} \leq \mathbf{0}\}$. Then, lin.space($\mathbf{P}$) $= \{\mathbf{x} \in \mathbf{R^n} \mid \mathbf{A} \cdot \mathbf{x} = \mathbf{0}\}$ is called the lineality space of $\mathbf{P}$. It is the t-dimensional linear subspace contained in the cone.

Reversible Vectors. The lineality space, lin.space($\mathbf{P}$), contains the nonzero vectors $\mathbf{r}$ in $\mathbf{P}$ whose opposite $-\mathbf{r}$ is also in $\mathbf{P}$. These involve only reversible fluxes and represent flux distributions that can operate in both directions. They can be called reversible vectors.

Pointed Cone. A convex polyhedral cone $\mathbf{P}$ is said to be pointed if lin.space($\mathbf{P}$) $= \{0\}$. In other words, a cone is non-pointed if it contains a reversible vector, and pointed otherwise.

Generation of a Pointed Cone. A pointed cone $\mathbf{P}$ can be generated by non-negative combination of its extreme rays, which is the unique, minimal generating set of $\mathbf{P}$ (MGS).

Generation of a Nonpointed Cone. A non-pointed cone $\mathbf{P}$ can still be generated by convex combination of a set of vectors, but there is no longer a unique MGS.

Extreme Rays or Edges. A vector $\mathbf{d}$ is a ray of the convex polyhedral cone $\mathbf{P}$ if for all $\mathbf{x} \in \mathbf{P}$, $\mathbf{x} + \lambda \cdot \mathbf{d} \in \mathbf{P}$ for each $\lambda \geq 0$. If a ray $\mathbf{d}$ cannot be expressed as non-negative combination of other rays in $\mathbf{P}$, it is an extreme ray. In metabolic pathway analysis, extreme rays are important because (i) if the cone is pointed, the extreme rays are the unique MGS. In the general case, the extreme rays (ii) belong to every generating set of $\mathbf{P}$, and (iii) they are always non-decomposable.

Nondecomposability. Given a cone $\mathbf{P}$, a vector $\mathbf{n} \in \mathbf{P}$ is non-decomposable if it cannot be represented as a positive combination of two vectors $\mathbf{v'}$ and $\mathbf{v''}$ in $\mathbf{P}$ that contain zero elements wherever $\mathbf{n}$ does and include at least one additional zero each. These vectors represent the simple states or functional units that a cell can show; the rest of feasible states can be seen as its aggregated action without cancelations. This property is relevant in several applications. All the generating set may contain non-decomposable vectors, but only the EMS is the set of all of the non-decomposable vectors in $\mathbf{P}$.

Systemic Independence. A set of vectors $\mathbf{I}$ are systemically independent if no vector in $\mathbf{I}$ can be represented as a non-negative combination of others. The extreme rays are always a systemically independent set of vectors.

Irreducible Subset. Given a generating set **G** of **P** that is not systemically independent, every smallest subset of **G** that generates **G**, and therefore **P**, is an irreducible subset of **G**. If the cone is pointed, there is a unique irreducible subset for every generating set and it coincides with the extreme rays of the cone, but if the cone is non-pointed, several irreducible subsets may exist.

(Flux) Vector-Space. The term (flux) vector-space refers to the space with the metabolic fluxes as axis. The original flux vector-space has dimensions n (n is number of reaction in the network), but some network-based pathways are computed in auxiliary vector-spaces of higher dimension.

Appendix

Computation of Network-Based Pathways

The elementary modes can be computed with *Metatool* [13] and *cellNetAnalyzer* [52], both running under MATLAB. The extreme pathways can be computed using *expa* [53]. Minimal generating sets can be obtained using SNA [54], a software package running under Mathematica, or using ccd [55] as reported in [40]. In addition, we describe a simple method to get an MGS from the EMS extracting an irreducible subset.

Extracting an MGS from the EMS. The procedure can be outlined with the following pseudocode:

for each elementary mode $\mathbf{e_i}$ in $\mathbf{E}$

define $\mathbf{A} = [\mathbf{M}\ \mathbf{E_r}]$

if (there is no $\mathbf{w} \geq \mathbf{0} \mid \mathbf{A} \cdot \mathbf{w} = \mathbf{e}$) then: add $\mathbf{e_i}$ to $\mathbf{M}$

end

where $\mathbf{E}$ is the matrix formed with EMs as columns, $\mathbf{E_r}$ is the submatrix of $\mathbf{E}$ with only columns after i, and $\mathbf{M}$ is the matrix collecting the MGs (thus empty on first iteration).

If the cone is pointed, the resultant set is the unique MGS (and coincides with the extreme rays of the cone). Otherwise, it is a nonunique MGS formed with non-decomposable vectors.

Acknowledgments

This research has been partially supported by the Spanish Government (DPI2008-06880-C03-01 and A/016560/08). The first author is recipient of a fellowship from the Spanish Ministry of Education and Science (FPU AP2005-1442).

References

[1] G. N. Stephanopoulos and A. A. Aristidou, *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego, Calif, USA, 1998.

[2] B. O. Palsson, *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press, Cambridge, Mass, USA, 2006.

[3] F. Llaneras and J. Picó, "Stoichiometric modelling of cell metabolism," *Journal of Bioscience and Bioengineering*, vol. 105, no. 1, pp. 1–11, 2008.

[4] S. Klamt and J. O. Stelling, "Two approaches for metabolic pathway analysis?" *Trends in Biotechnology*, vol. 21, no. 2, pp. 64–69, 2003.

[5] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, "Comparison of network-based pathway analysis methods," *Trends in Biotechnology*, vol. 22, no. 8, pp. 400–405, 2004.

[6] J. Gagneur and S. Klamt, "Computation of elementary modes: a unifying framework and the new binary approach," *BMC Bioinformatics*, vol. 5, article 175, 2004.

[7] S. Schuster, D. A. Fell, and T. Dandekar, "A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks," *Nature Biotechnology*, vol. 18, no. 3, pp. 326–332, 2000.

[8] S. Schuster, T. Dandekar, and D. A. Fell, "Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering," *Trends in Biotechnology*, vol. 17, no. 2, pp. 53–60, 1999.

[9] C. H. Schilling and B. ∅. Palsson, "Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 249–283, 2000.

[10] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles, "Metabolic network structure determines key aspects of functionality and regulation," *Nature*, vol. 420, no. 6912, pp. 190–193, 2002.

[11] J. A. Papin, N. D. Price, and B. ∅. Palsson, "Extreme pathway lengths and reaction participation in genome-scale metabolic networks," *Genome Research*, vol. 12, no. 12, pp. 1889–1900, 2002.

[12] N. D. Price, J. A. Papin, and B. ∅. Palsson, "Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis," *Genome Research*, vol. 12, no. 5, pp. 760–769, 2002.

[13] T. Pfeiffer, I. Sánchez-Valdenebro, J. C. Nuño, F. Montero, and S. Schuster, "METATOOL: for studying metabolic networks," *Bioinformatics*, vol. 15, no. 3, pp. 251–257, 1999.

[14] T. Çakir, B. Kirdar, and K. O. Ülgen, "Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks," *Biotechnology and Bioengineering*, vol. 86, no. 3, pp. 251–260, 2004.

[15] M. W. Covert and B. ∅. Palsson, "Constraints-based models: regulation of gene expression reduces the steady-state solution space," *Journal of Theoretical Biology*, vol. 221, no. 3, pp. 309–325, 2003.

[16] R. Carlson, D. Fell, and F. Srienc, "Metabolic pathway analysis of a recombinant yeast for rational strain development," *Biotechnology and Bioengineering*, vol. 79, no. 2, pp. 121–134, 2002.

[17] S. J. van Dien and M. E. Lidstrom, "Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph Methylobacterium extorquens AM1, with application to reconstruction of C3 and C4 metabolism," *Biotechnology and Bioengineering*, vol. 78, no. 3, pp. 296–312, 2002.

[18] S. Klamt and E. D. Gilles, "Minimal cut sets in biochemical reaction networks," *Bioinformatics*, vol. 20, no. 2, pp. 226–234, 2004.

[19] J. C. Liao, S.-Y. Hou, and Y.-P. Chao, "Pathway analysis, engineering, and physiological considerations for redirecting central metabolism," *Biotechnology and Bioengineering*, vol. 52, no. 1, pp. 129–140, 1996.

[20] J.-M. Schwartz and M. Kanehisa, "Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis," *BMC Bioinformatics*, vol. 7, article 186, 2006.

[21] R. Schwarz, P. Musch, A. von Kamp, et al., "YANA—a software tool for analyzing flux modes, gene-expression and enzyme activites," *BMC Bioinformatics*, vol. 6, article 135, 2005.

[22] S. J. Wiback, R. Mahadevan, and B. ∅. Palsson, "Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum," *Journal of Theoretical Biology*, vol. 224, no. 3, pp. 313–324, 2003.

[23] F. Llaneras and J. Picó, "An interval approach for dealing with flux distributions and elementary modes activity patterns," *Journal of Theoretical Biology*, vol. 246, no. 2, pp. 290–308, 2007.

[24] J. Forster, A. K. Gombert, and J. Nielsen, "A functional genomics approach using metabolomics and in silico pathway analysis," *Biotechnology and Bioengineering*, vol. 79, no. 7, pp. 703–712, 2002.

[25] D. A. Beard, S.-D. Liang, and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophysical Journal*, vol. 83, no. 1, pp. 79–86, 2002.

[26] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. ∅. Palsson, "Genome-scale metabolic model of Helicobacter pylori 26695," *Journal of Bacteriology*, vol. 184, no. 16, pp. 4582–4593, 2002.

[27] T. Dandekar, S. Schuster, B. Snel, M. Huynen, and P. Bork, "Pathway alignment: application to the comparative analysis of glycolytic enzymes," *Biochemical Journal*, vol. 343, no. 1, pp. 115–124, 1999.

[28] A. Cornish-Bowden and M. L. Cárdenas, "From genome to cellular phenotype—a role for metabolic flux analysis?" *Nature Biotechnology*, vol. 18, no. 3, pp. 267–268, 2000.

[29] J. Gao, V. M. Gorenflo, J. M. Scharer, and H. M. Budman, "Dynamic metabolic modeling for a MAb bioprocess," *Biotechnology Progress*, vol. 23, no. 1, pp. 168–181, 2007.

[30] A. P. Teixeira, C. Alves, P. M. Alves, M. J. T. Carrondo, and R. Oliveira, "Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control," *BMC Bioinformatics*, vol. 8, article 30, 2007.

[31] A. Provost, G. Bastin, S. N. Agathos, and Y.-J. Schneider, "Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells," *Bioprocess and Biosystems Engineering*, vol. 29, no. 5-6, pp. 349–366, 2006.

[32] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1970.

[33] B. L. Clarke, "Stoichiometric network analysis," *Cell Biophysics*, vol. 12, no. 1, pp. 237–253, 1988.

[34] C. Wagner and R. Urbanczik, "The geometry of the flux cone of a metabolic network," *Biophysical Journal*, vol. 89, no. 6, pp. 3837–3845, 2005.

[35] C. H. Schilling, D. Letscher, and B. ∅. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 229–248, 2000.

[36] S. Schuster, C. Hilgetag, J. H. Woods, and D. A. Fell, "Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism," *Journal of Mathematical Biology*, vol. 45, no. 2, pp. 153–181, 2002.

[37] R. Urbanczik and C. Wagner, "An improved algorithm for stoichiometric network analysis: theory and applications," *Bioinformatics*, vol. 21, no. 7, pp. 1203–1210, 2005.

[38] A. Provost and G. Bastin, "Dynamic metabolic modelling under the balanced growth condition," *Journal of Process Control*, vol. 14, no. 7, pp. 717–728, 2004.

[39] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley & Sons, Amsterdam, The Netherlands, 1998.

[40] A. Larhlimi and A. Bockmayr, "A new constraint-based description of the steady-state flux cone of metabolic networks," *Discrete Applied Mathematics*, vol. 157, no. 10, pp. 2257–2266, 2009.

[41] S. Klamt, J. Gagneur, and A. von Kamp, "Algorithmic approaches for computing elementary modes in large biochemical reaction networks," *Systems Biology*, vol. 152, no. 4, pp. 249–255, 2005.

[42] M. Terzer and J. Stelling, "Large-scale computation of elementary flux modes with bit pattern trees," *Bioinformatics*, vol. 24, no. 19, pp. 2229–2235, 2008.

[43] L. F. de Figueiredo, A. Podhorski, A. Rubio, et al., "Computing the shortest elementary flux modes in genome-scale metabolic networks," *Bioinformatics*, vol. 25, no. 23, pp. 3158–3165, 2009.

[44] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar, "Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae," *Bioinformatics*, vol. 18, no. 2, pp. 351–361, 2002.

[45] J.-M. Schwartz, C. Gaugain, J. C. Nacher, A. de Daruvar, and M. Kanehisa, "Observing metabolic functions at the genome scale," *Genome Biology*, vol. 8, no. 6, article R123, 2007.

[46] C. Kaleta, L. F. De Figueiredo, and S. Schuster, "Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns," *Genome Research*, vol. 19, no. 10, pp. 1872–1883, 2009.

[47] C. L. Barrett, M. J. Herrgard, and B. ∅. Palsson, "Decomposing complex reaction networks using random sampling, principal component analysis and basis rotation," *BMC Systems Biology*, vol. 3, article 30, 2009.

[48] M. G. Poolman, D. A. Fell, and C. A. Raines, "Elementary modes analysis of photosynthate metabolism in the chloroplast stroma," *European Journal of Biochemistry*, vol. 270, no. 3, pp. 430–439, 2003.

[49] R. Steuer, A. N. Nesi, A. R. Fernie, T. Gross, B. Blasius, and J. Selbig, "From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle," *Bioinformatics*, vol. 23, no. 11, pp. 1378–1385, 2007.

[50] J.-J. Zhong, "Plant cell culture for production of paclitaxel and other taxanes," *Journal of Bioscience and Bioengineering*, vol. 94, no. 6, pp. 591–599, 2002.

[51] R. P. Nolan, A. P. Fenley, and K. Lee, "Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis," *Metabolic Engineering*, vol. 8, no. 1, pp. 30–45, 2006.

[52] S. Klamt, J. Stelling, M. Ginkel, and E. D. Gilles, "FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps," *Bioinformatics*, vol. 19, no. 2, pp. 261–269, 2003.

[53] S. L. Bell and B. ∅. Palsson, "Expa: a program for calculating extreme pathways in biochemical reaction networks," *Bioinformatics*, vol. 21, no. 8, pp. 1739–1740, 2005.

[54] R. Urbanczik, "SNA—a toolbox for the stoichiometric analysis of metabolic networks," *BMC Bioinformatics*, vol. 7, article 129, 2006.

[55] K. Fukuda and A. Prodon, "Double description method revisited," in *Combinatorics and Computer Science*, vol. 1120, pp. 91–111, Springer, London, UK, 1996.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 518743, 14 pages
doi:10.1155/2010/518743

Research Article

Elementary Mode Analysis for the Rational Design of Efficient Succinate Conversion from Glycerol by *Escherichia coli*

Zhen Chen,[1,2] Hongjuan Liu,[1] Jianan Zhang,[1] and Dehua Liu[2]

[1] *Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China*
[2] *Department of Chemical Engineering, Institute of Applied Chemistry, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Hongjuan Liu, liuhongjuan@tsinghua.edu.cn and Jianan Zhang, zhangja@tsinghua.edu.cn

Received 2 December 2009; Revised 20 May 2010; Accepted 7 July 2010

Academic Editor: Jennifer Reed

By integrating the restriction of oxygen and redox sensing/regulatory system, elementary mode analysis was used to predict the metabolic potential of glycerol for succinate production by *E. coli* under either anaerobic or aerobic conditions. It was found that although the theoretical maximum succinate yields under both anaerobic and aerobic conditions are 1.0 mol/mol glycerol, the aerobic condition was considered to be more favorable for succinate production. Although increase of the oxygen concentration would reduce the succinate yield, the calculation suggests that controlling the molar fraction of oxygen to be under 0.65 mol/mol would be beneficial for increasing the succinate productivity. Based on the elementary mode analysis, the rational genetic modification strategies for efficient succinate production under aerobic and anaerobic conditions were obtained, respectively. Overexpressing the phosphoenolpyruvate carboxylase or heterogonous pyruvate carboxylase is considered to be the most efficient strategy to increase the succinate yield.

1. Introduction

Glycerol has become an abundant and inexpensive carbon source due to its generation as an inevitable byproduct of biodiesel production. Over the past few years, the price of crude glycerol has decreased 10-fold due to the tremendous growth of the biodiesel industry [1]. Much effort has been paid for the development of processes to convert crude glycerol into higher-value products to maximize the full economic potential of biodiesel process. For example, the transformation of glycerol into 1,3-propanediol has been extensively studied in the past few years [2–4].

Several recent studies also tried to utilize glycerol as a carbon source for the transformation of other valued products such as ethanol [5] and amino acids [6]. Succinate is traditionally produced from sugars and suffers the limitation due to the availability of reducing equivalents. Compared with glucose, glycerol has a higher reduced state and also several microorganisms such as *E. coli* can transform glycerol into succinate [7]. So, the byproduct glycerol is a potential substrate for the succinate production. Despite few attempts in the past, no industrially competitive organisms can effectively produce succinate from glycerol so far. In the light of the new powerful tools of metabolic engineering, the quest for targeted development of strains that can effectively utilize glycerol for succinate production is strongly revived. *E. coli* is one of the most promising organisms since it can directly utilize glycerol and it has been traditionally developed for succinate production [8–10].

The dissimilation of glycerol in *E. coli* is catalyzed by proteins encoded by *glp* regulon under aerobic conditions. Glycerol is first phosphorylated into glycerol 3-phophate (G3P) by ATP-dependent glycerol kinase encoded by *glpK* gene, and then glycerol 3-phophate is converted into dihydroxyacetone phosphate (DHAP) by aerobic G3P dehydrogenase encoded by *glpD* gene (Figure 1) [11, 12].

Although the aerobic utilization of glycerol by *E. coli* has been known for a long time, the fermentative pathway of glycerol has just been clarified recently [13, 14]. It has been suggested that the feasibility of fermenting glycerol into fuels and other reduced chemicals is through the inducing of its native 1,2-propanediol fermentative pathway without

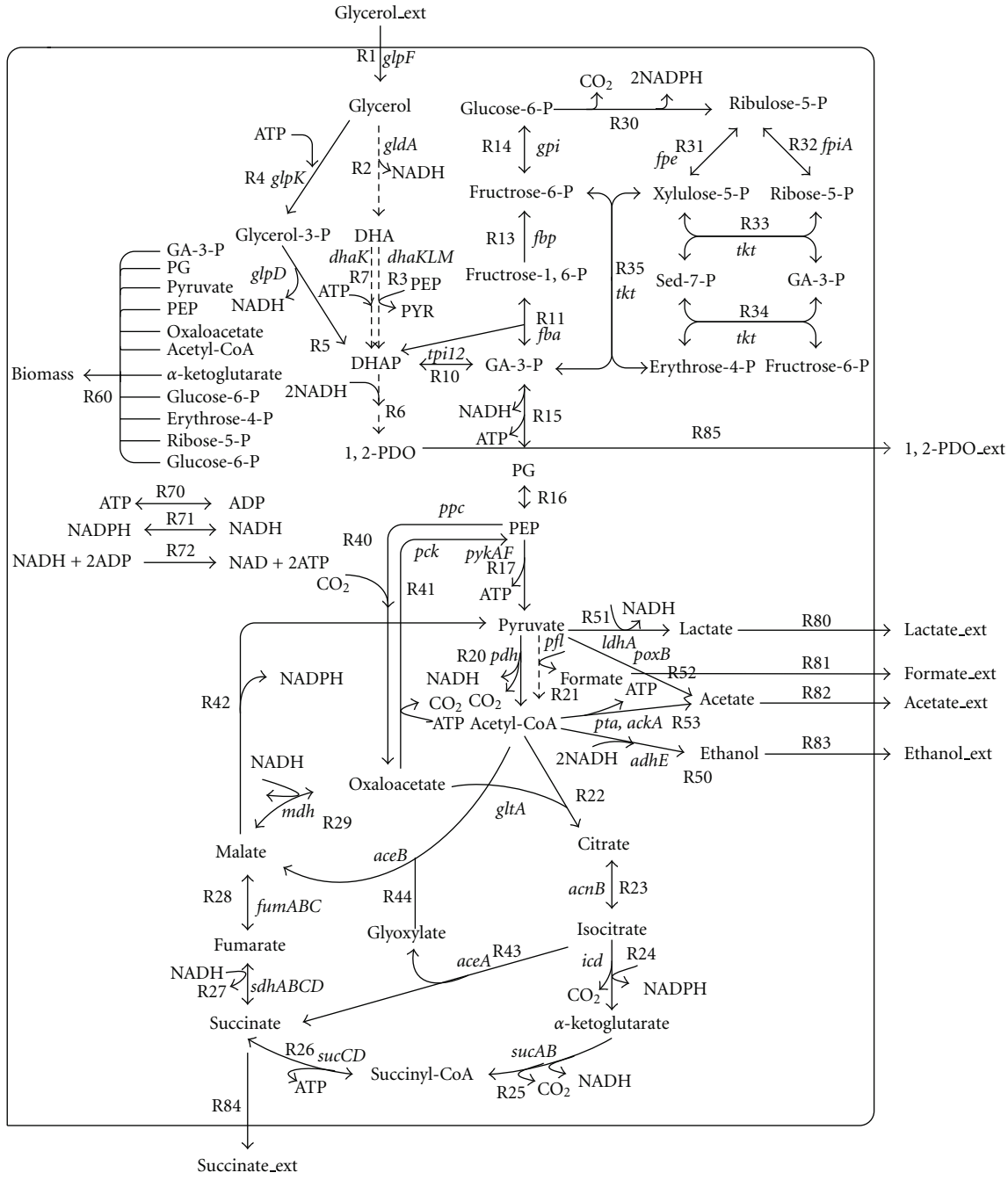


Figure 1: Central metabolic network of glycerol in wildtype *E. coli*. The dashed arrows represent the particular pathways in anaerobic conditions. Reversible reactions are represented by a double-headed arrow. Key genes associated with the pathway are included.

using external electron acceptors. In this pathway, glycerol is converted to dihydroxyacetone (DHA) by NAD^+-linked glycerol dehydrogenase (GDH), and the DHA is phosphorylated to DHAP via the ATP-dependent or phosphoenolpyruvate (PEP)-dependent DHA kinase (DHAK). DHAP is then reduced into 1,2-propanediol or enter glycolysis [15].

Since *E. coli* can utilize glycerol in both aerobic and anaerobic conditions, it is necessary to analyze the potential and the feasibility of engineering *E. coli* for the succinate production in either condition. A careful metabolic pathway analysis is very helpful in such kind of estimation and rational strain development. Elementary mode analysis is one of the most powerful tools for metabolic pathway analysis using for the metabolic properties study of cellular systems [16–18]. Elementary mode analysis allows the calculation of a solution space that contains all possible steady-state flux distributions of a network. The stoichiometry of the metabolic network, including carbon as well as cofactor requirements, is fully considered in elementary mode analysis. On the other hand, it also allows determining the overall capacity, that is,

theoretical maximum yield, of a cellular system and studying the effects of any genetic modification. Based on such studies, rational design can be obtained for the efficient production and genetic modification. Moreover, knowledge of the theoretical maximum yield allows estimating the potential economic efficiency of a process. Recently, elementary mode analysis has been used for genome scale metabolic studies dealing with, for example, the rational design of methionine production in *E. coli* and *C. glutamicum* [19], the production of polyhydroxybutanoate in yeast [20], and growth-related aspects in *Saccharomyces cerevisiae* [21, 22] and *E. coli* [23, 24].

In this work, the elementary mode analysis was carried out for succinate production by comparing the metabolic networks of *E. coli* in anaerobic and aerobic conditions. The pathways involved in the flux scenario representing optimal succinate production were investigated and the effect of oxygen level on succinate production and biomass was also discussed. Furthermore, the rational design for the genetic modification of the *E. coli* to enhance the succinate production was developed. This work is considered useful for the further strain improvement and metabolic regulation in the succinate production by *E. coli* with glycerol as the substrate.

2. Materials and Methods

2.1. Metabolic Reaction Network. The glycerol metabolic network of *E. coli* was constructed (Figure 1) based on KEGG database (http://www.genome.jp/kegg/metabolism.html) as well as biochemical and physiological literatures [13, 14, 25]. It includes glycerol dissimilation pathways, glycolysis pathway (EMP), pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, biosynthesis pathway, anaplerosis, and respiratory chain. The metabolic network was depicted in Figure 1. For the interconversion of NADH and NADPH, a cytosolic transhydrogenase transferring protons from NADPH to NAD^+ and a membrane-bound transhydrogenase reducing $NADP^+$ by oxidation of NADH were implemented [26]. For ATP production in the respiratory chain, a P/O ratio of 2 for NADH was assumed [27].The precursor demand for biomass formation was calculated according to the literature [28]. The biomass term is represents as Cmol basis ($CN_{0.24}S_{0.008}$). The cell physiology of *E. coli* is strongly affected by oxygen from different levels such as the transcriptional regulatory which cannot be represented simply in the metabolic network. For example, one component fumarate and nitrate reduction (FNR) protein is aerobic/anaerobic response regulator [25]. FNR appears to sense oxygen directly through a redox-sensitive iron-sulphur cluster in the protein and is active only during anaerobic growth. The two iron-sulphur ($[4Fe–4S]^{2+}$) clusters in the dimeric FNR protein are converted to two $[2Fe–2S]^{2+}$ clusters upon exposure to stoichiometric levels of oxygen. Active FNR protein activates and represses target genes in response to anaerobiosis. It acts as a positive regulator of genes expressed under anaerobic fermentative conditions such as aspartase, formate dehydroganases, fumarate reductase, and pyruvate formate lyase. To account for the effects of FNR regulator and other experimental discovery, the following constraints are used to discriminate the metabolic networks under aerobic and anaerobic network.

For the anaerobic model, glycerol is assumed to be dissimilated into DHA by NAD^+-dependent glycerol dehydrogenase and then DHA is phosphorylated into DHAP by ATP-dependent or PEP-dependent DHA kinase [13, 14]. The pathway from DHAP to 1,2-propanediol is considered to be active under anaerobic condition [13]. The pyruvate dehydrogenase complex is inactive under anaerobic condition and thus pyruvate-formate lyase was the only active enzyme that catalyzes the transformation of pyruvate into acetyl-CoA [29]. The TCA cycle is broken at the alpha-ketoglutarate dehydrogenase step and the respiratory chain is assumed to be inactive [30]. The detailed description of the model is listed in Appendices A.1 and A.2.

For the aerobic model, glycerol is firstly phosphorylated into G3P by ATP-dependent glycerol kinase and then G3P is transferred into DHAP by NAD^+-dependent G3P dehydrogenase [11, 12]. The 1, 2-propanediol pathway is assumed to be inactive. Pyruvate oxidase is active under aerobic condition which will transfer pyruvate into acetate. The detailed description of the model is listed in Appendices A.1 and A.3.

2.2. Computational Methods. In the present work, the elementary mode analysis was carried out for studying the aerobic and anaerobic metabolism of glycerol in *E. coli* by using METATOOL 5.1 [31]. The script files and compiled shared library of METATOOL 5.1 can be downloaded from the METATOOL website (http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool/metatool5.1/metatool5.1.html). The mathematical details of the algorithm were described elsewhere [32]. Metabolic pathway analysis resulted in tens to hundreds of elementary flux modes for each situation investigated. For each of these flux modes, the fluxes were calculated as relative molar values normalized to the glycerol uptake rate and were expressed as mol/mol (glycerol).

3. Results and Discussion

3.1. Elementary Mode Analysis of Glycerol Metabolism under Anaerobic Condition. Under anaerobic condition, the metabolic network model got 55 elementary flux modes. The relationship between the yields of products and biomass was shown in Figure 2. The maximum molar yield of biomass under anaerobic condition is 0.187 mol/mol in which the respective yields of 1,2-propanediol, ethanol, and formate were 0.248 mol/mol, 0.495 mol/mol, and 0.57 mol/mol and no succinate, acetate, and lactate were produced. It was found that the cell growth was always associated with the production of 1,2-propandediol, ethanol, and formate. Therefore, the production of 1,2-propanediol, ethanol, and formate was necessary for the biomass synthesis during the glycerol metabolism. The biomass synthesis process consumes ATP and produces reducing equivalents (NADH). Both ATP

Table 1: Reactions and enzymes involved in Figure 1.

Reactions	Genes	Enzymes	Aerobic/anaerobic specificity	References
R1	*glpF*	glycerol facilitator		[1]
R2	*gldA*	Glycerol dehydrogenase II	anaerobic	[2]
R3	*dhaKLM*	PTS-dependent Dihydroxyacetone kinase	anaerobic	[2]
R4	*glpK*	glycerol kinase	aerobic	[1]
R5	*glpD*	Glycerol-3-phosphate dehydrogenase	aerobic	[1]
R6		sets of reactions	anaerobic	[2]
R7	*dhaK*	ATP-dependent Dihydroxyacetone kinase		[2]
R10	*tpi12*	Triose-phosphate isomerase		KEGG
R11	*fba*	Fructose-bisphosphate aldolase		KEGG
R13	*fbp*	Fructose-bisphosphatase		KEGG
R14	*gpi*	Glucose-6-phosphate isomerase		KEGG
R15		sets of reactions		KEGG
R16		sets of reactions		KEGG
R17	*pykAF*	Pyruvate kinase		KEGG
R20	*pdh*	Pyruvate dehydrogenase	aerobic	[3]
R21	*pfl*	Pyruvate formate-lyase	anaerobic	[4]
R22	*gltA*	Citrate synthase		KEGG
R23	*acnB*	Aconitate hydratase		KEGG
R24	*icd*	Isocitrate dehydrogenase		KEGG
R25	*sucAB*	Oxoglutarate dehydrogenase	aerobic	[5]
R26	*sucCD*	Succinate-CoA ligase	aerobic	[5]
R27	*sdhABCD*	Succinate dehydrogenase		KEGG
R28	*fumABC*	Fumarate hydratase		KEGG
R29	*mdh*	Malate dehydrogenase		KEGG
R30		sets of reactions		KEGG
R31	*rpe*	Ribulose-phosphate 3-epimerase		KEGG
R32	*rpiA*	Ribose-5-phosphate isomerase		KEGG
R33	*tkt*	Transketolase		KEGG
R34	*tkt*	Transaldolase		KEGG
R35	*tkt*	Transketolase		KEGG
R40	*ppc*	PEP carboxylase		KEGG
R41	*pck*	PEP carboxykinase		KEGG
R42	*malE*	malic enzyme		KEGG
R43	*aceA*	Isocitrate lyase		KEGG
R44	*aceB*	Malate synthase		KEGG
R50	*adhE*	Aldehyde dehydrogenase		KEGG
R51	*ldhA*	Lactate dehydrogenase		KEGG
R52	*poxB*	pyruvate oxidase	aerobic	[6]
R53	*pta,ackA*	Phosphate acetyltransferase, Acetate kinase		KEGG
R60		Biomass formation		[7]
R70		ATP-hydrolysis		[7]
R71		Transhydrogenase		[7]
R72		respiratory chain 1	aerobic	[7]
R80		Membrane transport reaction		
R81		Membrane transport reaction		
R82		Membrane transport reaction		
R83		Membrane transport reaction		
R84		Membrane transport reaction		
R85		Membrane transport reaction		

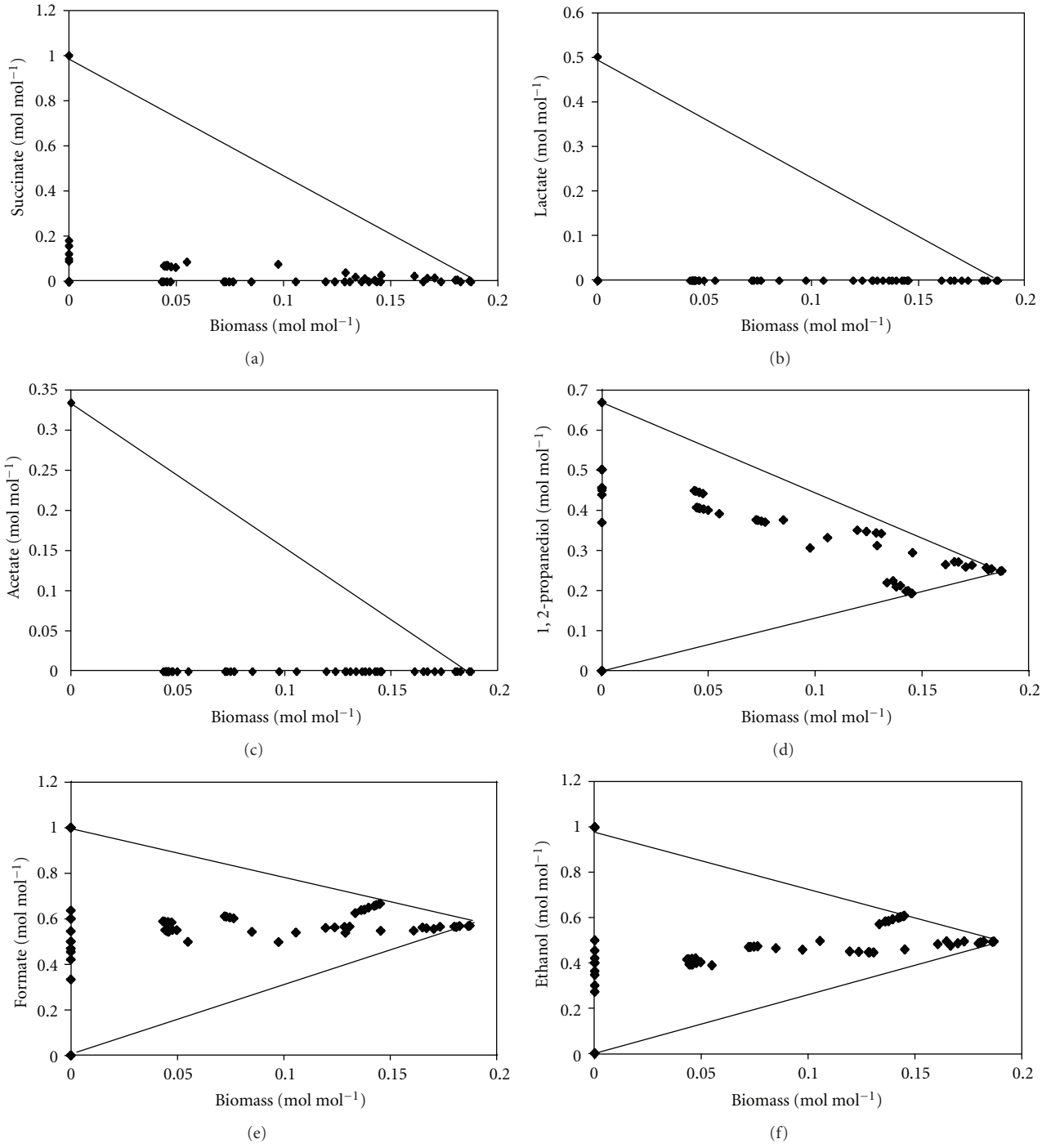


Figure 2: Relationship between the yields of biomass and byproducts for the obtained elementary modes of *E. coli* under anaerobic conditions. (a) Succinate, (b) Lactate, (c) Acetate, (d) 1,2-Propanediol, (e) Formate, and (f) Ethanol. The enclosed regions represent the possible solution space. The fluxes were normalized by glycerol uptake rate and expressed as mol/mol (glycerol).

and NAD need to be regenerated through the production of other byproducts (the biomass synthesis equation in Appendix A.2). For the glycerol metabolism in anaerobic condition, only the glycerol to 1,2-propanediol pathway can consume extra NADH (see (1)) and thus provide the mean to consume the reducing equivalents generated during the synthesis of biomass. The conversion of glycerol to ethanol and formate (see (2)), a redox-balanced pathway, fulfills energy requirements by generating ATP via substrate-level phosphorylation (Appendix A.4). The calculation results were consistent with the experiment observation that 1,2-propanediol and ethanol were growth-associated products

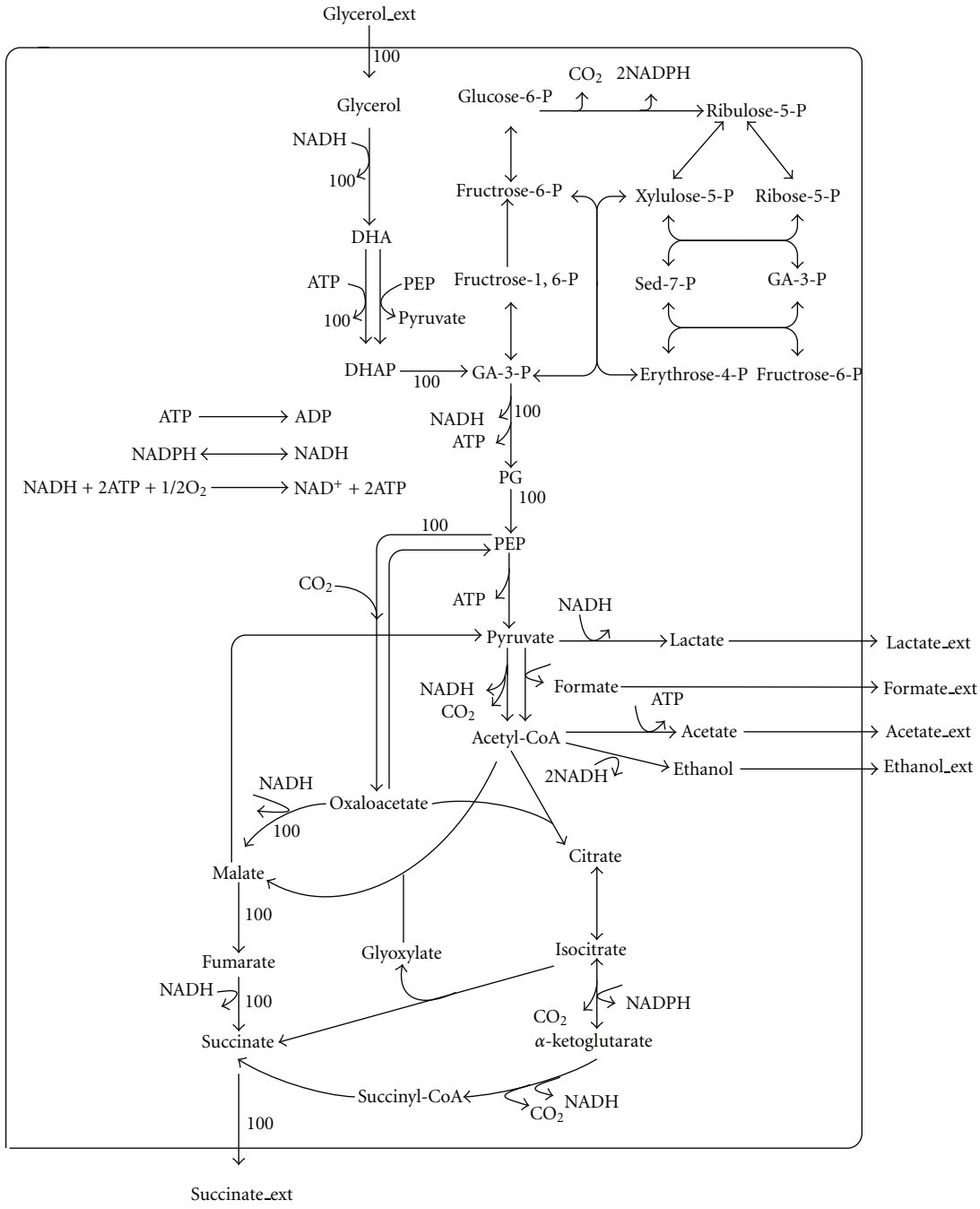


Figure 3: The optimum flux distribution of glycerol metabolism for succinate production in *E. coli* under anaerobic conditions when only ATP-dependent DHA kinase plays a function.

[13, 14]

$$\text{glycerol} + \text{NADH} \longrightarrow 1,2\text{-propanediol} + \text{NAD}, \quad (1)$$

$$\text{glycerol} + \text{ADP} \longrightarrow \text{ethanol} + \text{formate} + \text{ATP}. \quad (2)$$

The maximum succinate yield under anaerobic condition is 1.0 mol/mol when CO_2 or carbonate salts are added as cosubstrates and the optimal flux distribution for succinate production was shown in Figure 3. In this case, there was no production of biomass and other byproducts. The key points for this mode were that the phosphorylation of DHA was only catalyzed by ATP-dependent DHA kinase and PEP was totally carboxylated into oxaloacetate by PEP carboxylase and the latter was further transferred into succinate. This required a very high activity of ATP-dependent DHA kinase and PEP carboxylase. However, it was reported that the PEP-dependent DHA kinase plays the main role in *E. coli* which dramatically reduced the yield of succinate. With single PEP-dependent DHA kinase function, there will be no succinate production [14]. Thus the PEP-dependent DHA kinase is

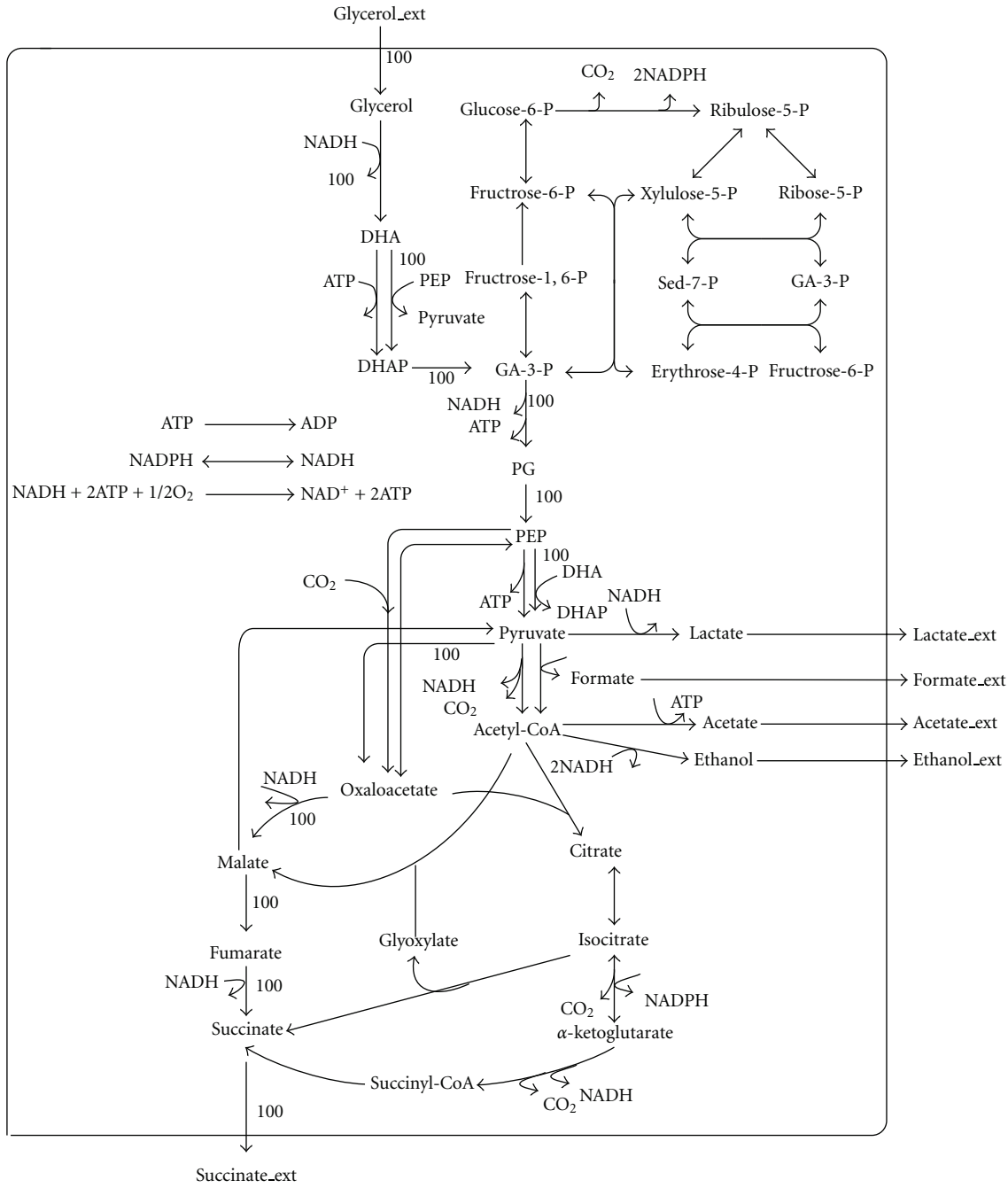


Figure 4: The alternative optimum flux distribution of glycerol metabolism for succinate production in *E. coli* under anaerobic conditions when only PEP-dependent DHA kinase plays a function and heterogeneous PEP carboxylase (pyc) are introduced and overexpressed in *E. coli*.

the bottleneck of succinate production under anaerobic condition.

3.2. Element Ary Mode Analysis of Glycerol Metabolism under Aerobic Condition. The aerobic metabolic network model got 259 elementary flux modes. The relationship between the yields of products and biomass was shown in Figure 5. It was indicated that the flux distribution modes under aerobic condition were completely different from that of anaerobic condition. The maximum molar yield of biomass under aerobic condition was 0.725 mol/mol in which only CO_2 was produced. The results suggested that aerobic condition was more favorable for biomass formation and the most effective way for biomass formation was through the TCA cycle. The cell growth process was no longer associated with the production of 1,2-propanedio, formate, and ethanol. Comparing with the anaerobic condition, the succinate production modes were more diversely distributed throughout

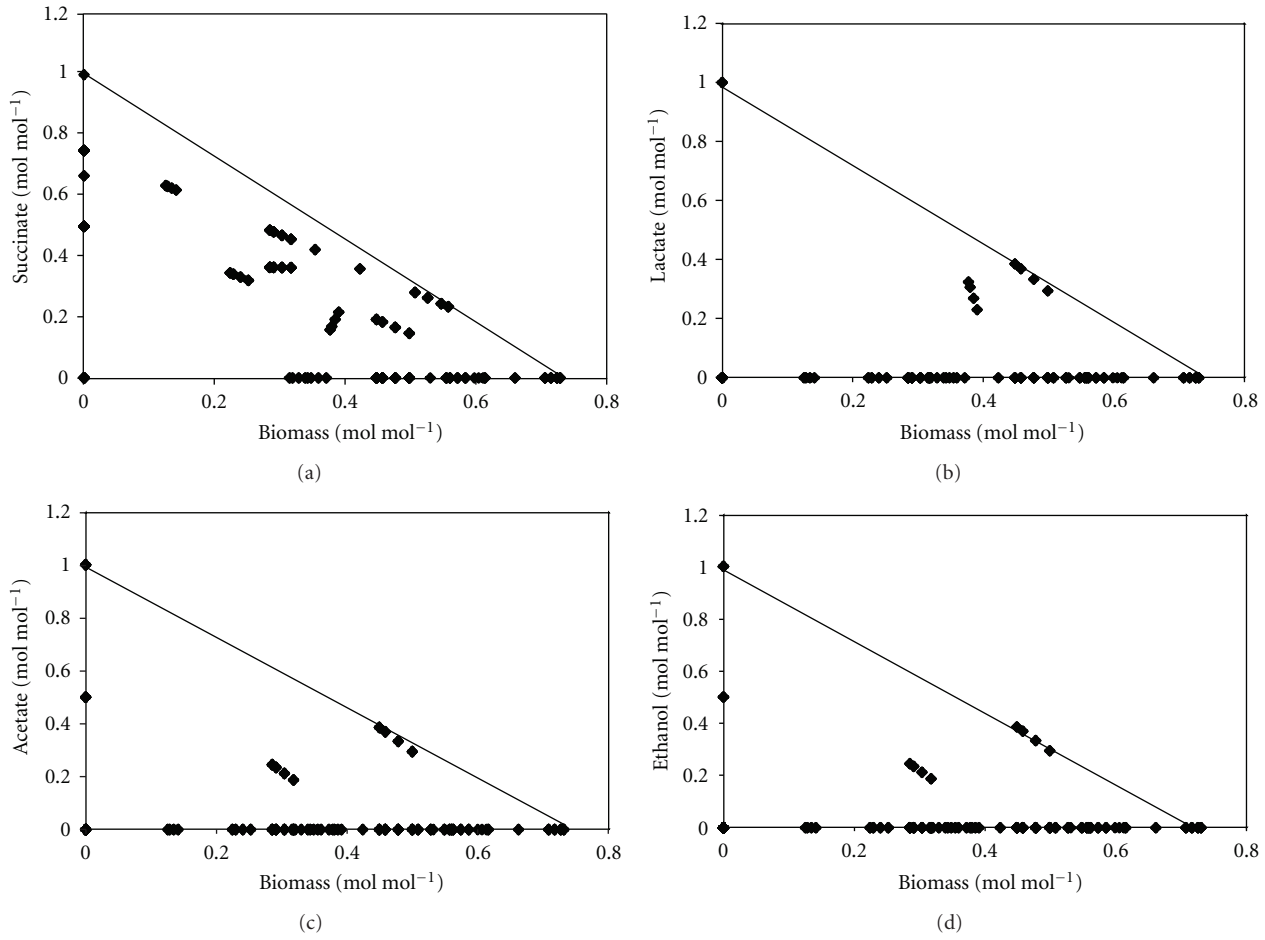


Figure 5: Relationship between the yields of biomass and byproducts for the obtained elementary modes of *E. coli* under aerobic conditions. (a) Succinate, (b) Lactate, (c) Acetate, and (d) Ethanol. The enclosed regions represent the possible solution space. The fluxes were normalized by glycerol uptake rate and expressed as mol/mol (glycerol).

Table 2: Degree of reduction of considered substrates and products.

Glycerol	14
O_2	−4
CO_2	0
Succinate	14
Ethanol	12
Lactate	12
Formate	2
1,2-propanediol	16
Acetate	8

the feasible solution space, and the corresponding yield of succinate at the same biomass yield was higher which indicated that the high potential of succinate production is associated with the cell growth by metabolic modification under aerobic condition.

The maximum succinate yield under aerobic condition (an aerobic mode suggests oxygen consumption) is also 1.0 mol/mol which requires the CO_2 or carbonate salts to be added as cosubstrates, and the optimal flux distribution for succinate production was shown in Figure 6. The optimal flux distribution modes were quite similar for the aerobic and anaerobic conditions. The key point for obtaining high succinate yield was considered as that PEP was totally carboxylated into oxaloacetate by PEP carboxylase and the latter was further transferred into succinate. This required a very high activity of PEP carboxylase.

The network robustness and its sensitivity to perturbation were critical to the optimal metabolic pathway. The sensitivity of succinate yield to the flux ratios at the key branch nodes PEP and acetyl-CoA was considered in this work. PEP was consumed in reactions R40 and R17 which are catalyzed by PEP carboxylase and Pyruvate kinase, respectively; the flux ratio was denoted as

$$\mathrm{R}(40, 17) = \frac{\mathrm{R40}}{(\mathrm{R40} + \mathrm{R17})}. \quad (3)$$

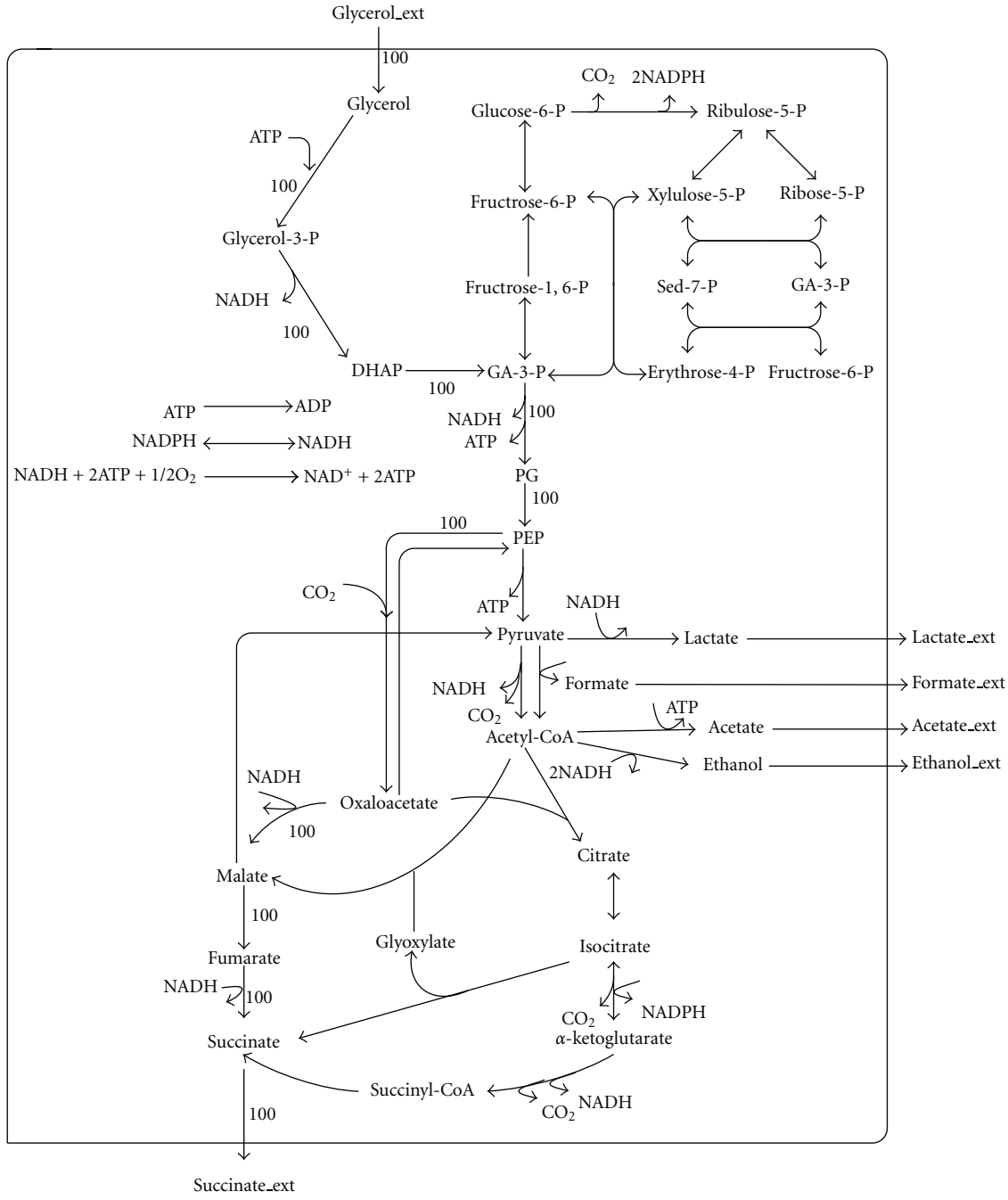


Figure 6: The optimum flux distribution of glycerol metabolism for succinate production in *E. coli* under aerobic conditions.

The influence of flux distribution at PEP node was shown in Figure 7(a). The succinate yield increased from 0.5 mol/mol to 1.0 mol/mol when R(40, 17) increased from 0 to 1.0. The results showed that increasing the flux distribution from PEP to oxaloacetate was beneficial for succinate production. For acetyl-CoA node, it was consumed in reactions R22, R44, R50, and R53 which are catalyzed by citrate synthase, aldehyde dehydrogenase, malate synthase, and phosphate acetyltransferase, respectively. The flux ratio was denoted as

$$\mathrm{R}(22, 44, 50, 53) = \frac{\mathrm{R22}}{(\mathrm{R22} + \mathrm{R44} + \mathrm{R50} + \mathrm{R53})}. \quad (4)$$

The influence of flux distribution at acetyl-CoA node was shown in Figure 7(b). The succinate yield decreased from 1.0 mol/mol to 0.75 mol/mol when R(22, 44, 50, 53) increased from 0 to 1.0. The results indicated that decreasing the flux distribution from acetyl to TCA cycle was beneficial for succinate production since the carbon would be lost during the TCA cycle.

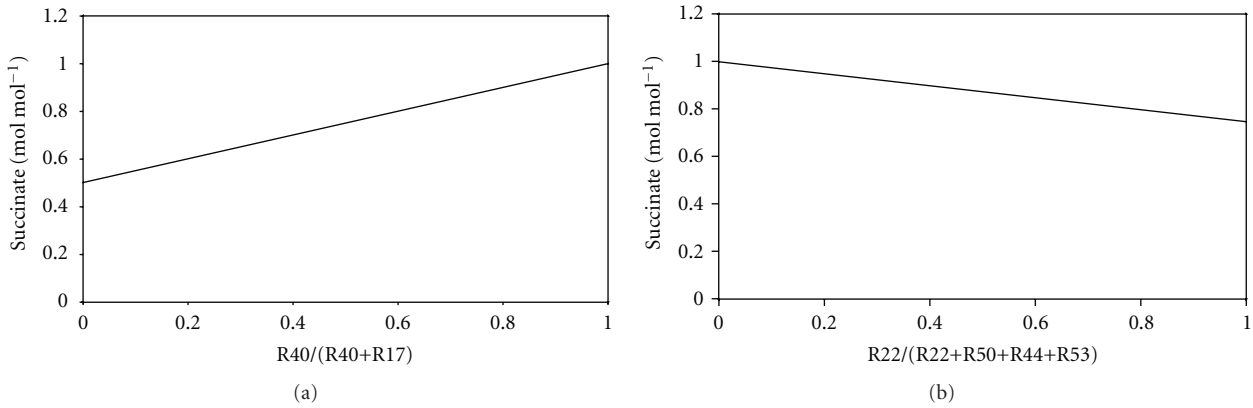


Figure 7: Sensitivity of succinate yield to the relative fluxes at the (a) PEP node and (b) AcCoA node under aerobic conditions. The PEP node involves the catabolic reactions of R40 and R17 which are catalyzed by PEP carboxylase and pyruvate kinase, respectively. The AcCoA node involves the catabolic reactions of R22, R50, R44, and R53 which are catalyzed by citrate synthase, aldehyde dehydrogenase, malate synthase, and phosphate acetyltransferase, respectively.

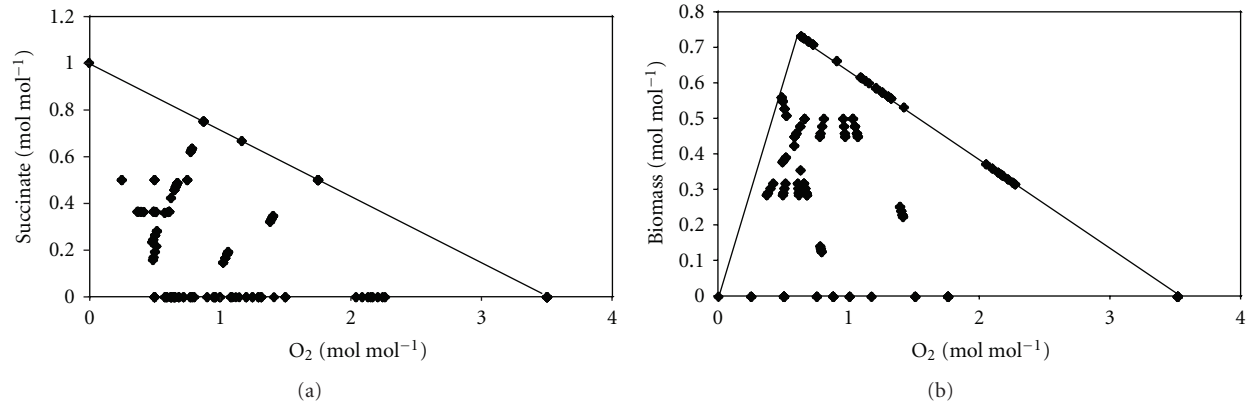


Figure 8: Effect of oxygen consumption on the succinate production and biomass formation under aerobic conditions. The enclosed regions represent the possible solution space. The fluxes were normalized by glycerol uptake rate and expressed as mol/mol (glycerol).

3.3. Effect of Oxygen Consumption on the Production of Succinate and Biomass. The effect of oxygen consumption on succinate and biomass production was further investigated according to the elementary mode analysis; the relationship between the molar fractions of oxygen consumption and succinate or biomass under aerobic conditions was calculated and shown in Figure 8. The theoretical succinate yield decreased as the oxygen consumption ratio increased (Figure 8(a)). The elementary modes of succinate yield distributed only on the left part of the solution space, which indicated that the higher consumption of oxygen was unfavorable for the succinate production. This is reasonable from the redox consideration. The higher consumption of oxygen results in more NAD which is critical for succinate production, to be utilized for the ATP production through oxidative phosphorylation. Especially as the molar oxygen consumption fraction of oxygen was more than 1.75 mol/mol, the elementary mode that produced succinate did not exist.

The theoretical biomass yield increased firstly and then decreased when the oxygen consumption ratio increased (Figure 8(b)). This is reasonable and consistent with the results of electron conservation because the rational increase of oxygen flux would be favorable for ATP synthesis through oxidative phosphorylation which is essential for biomass synthesis. However, high oxygen flux also results in less carbon source and reducing equivalents available for biomass because more carbon source would be oxidized to CO_2. The maximum biomass was achieved when the molar fraction of oxygen was 0.65 mol/mol. Since higher cell concentration is beneficial for increase of the productivity, the optimal regulation strategy is controlling the molar fraction of oxygen consumption less than 0.65 mol/mol.

3.4. Rational Design to Improve the Succinate Production by Genetic Modifications. Comparing the results of elementary flux mode analysis above, although the maximum succinate yields were 1.0 mol/mol under both anaerobic and aerobic

conditions, the aerobic condition seemed to be more favorable for succinate production in fact. The slow cell growth under anaerobic condition hindered the practical application of glycerol fermentation for succinate production [33]. What is more, the association of cell growth with the production of 1,2-propanediol, ethanol, and formate reduced the total yield of succinate.

As discussed above, to improve the yield of succinate yield under anaerobic condition, the substitution of PEP-dependent DHA kinase into ATP-dependent DHA kinase and overexpressing the PEP carboxylase would be a prior consideration. An alternative choice is to express the heterogeneous pyruvate carboxylase in *E. coli*. The overexpression of pyruvate carboxylase could redistribute the flux of pyruvate into oxaloacetate for succinate production. The optimal flux distribution for succinate production in such case could also reach 1.0 mol/mol which was shown in Figure 4. It has been reported that expressing the ATP-dependent DHA kinase and pyruvate carboxylase could both increase the yield of succinate [14]. Another consideration to increase the succinate yield is reducing the byproducts production. Since the production of 1,2-propanediol, ethanol, and formate is necessary for biomass synthesis, the proper strategy is reducing the acetate and lactate production. Knockout of the phosphate acetyltransferase gene (*pta*) and lactate dehydrogenase gene (*ldh*) would be expected to increase the succinate production.

For the case of aerobic condition, overexpressing the PEP carboxylase or expressing pyruvate carboxylase would be a prior consideration as discussed above. Knockdown of the isocitrate dehydrogenase gene (*icd*) would enhance the succinate production since the flux flowed from isocitrate to alpha-ketoglutarate and succinyl-CoA would result in the carbon lost. Since acetate is the main byproduct under aerobic condition [34], knocking out the pyruvate oxidase gene (*poxB*) and phosphate acetyltransferase gene (*pta*) is also expected to increase the succinate yield.

4. Conclusions

Nowadays, the conversion of byproduct glycerol has attracted more and more attention with the development of biodesiel industrial. The potentials of using glycerol for succinate production in *E. coli* under the anaerobic and aerobic conditions were compared by using elementary mode analysis in this work. The aerobic conditions seem to be more favorable for succinate production and the maximum succinate yield was 1.0 mol/mol. Although increase of the oxygen concentration would reduce the succinate yield, controlling the molar fraction of oxygen under 0.65 mol/mol would be beneficial for increasing the succinate productivity. According to the elementary mode analysis, the rational design was obtained for improving the succinate production by genetic modification under aerobic and anaerobic conditions, respectively. The results are considered useful for further investigation on the succinate metabolism of *E. coli*. The information also is beneficial for the efficient production of succinate from glycerol by *E. coli*.

Appendices

A. Models Used in This Study

A.1. Reactions and Enzymes Involved in Figure 1 (see Table 1).

A.2. The Anaerobic Metabolic Network Model Input File Used for the Program METATOOL.

(i) ENZREV (reversible reactions)

R10r R11r R14r R15r R16r R23r R26r R27r R28r R29r R31r R32r R33r R34r R35r R71r R86r.

(ii) ENZIRREV (irreversible reactions)

R1 R2 R3 R6 R7 R13 R17 R21 R22 R24 R30 R40 R41 R42 R43 R44 R50 R51 R53 R60 R70 R80 R81 R82 R83 R84 R85.

(iii) METINT (internal metabolite declaration)

Glycerol DHA DHAP GA-3-P propanediol PG PEP Pyruvate Acetyl-CoA CoASH Oxaloacetate Citrate Isocitrate a-Ketoglutarate Succinate Fumarate Malate Glyoxylate Glucose-6-P Fructrose-6-P Fructrose-16-P Ribulose-5-P Xylulose-5-P Ribose-5-P Sed-7-P Erythrose-4-P Lactate Formate Acetate Ethanol NAD NADH ATP ADP NADP NADPH CO_2.

(iv) METEXT (external metabolite declaration)

Glycerol_ext Ethanol_ext Acetate_ext CO_2_ext Lactate_ext Succinate_ext Formate_ext BIOMASS propanediol_ext.

(v) CAT

Reactions

(vi) Glycerol specific metabolisms

R1: Glycerol_ext = Glycerol.
R2: Glycerol + NAD = DHA + NADH.
R3: PEP + DHA = DHAP + Pyruvate.
R6: DHAP + 2NADH = propanediol + NAD.
R7: DHA + ATP = DHAP + ADP.

(vii) Glycolysis

R10r: DHAP = GA-3-P.
R11r: DHAP + GA-3-P = Fructrose-16-P.
R13: Fructrose-16-P = Fructrose-6-P.
R14r: Fructrose-6-P = Glucose-6-P.
R15r: GA-3-P + ADP + NAD = PG + ATP + NADH.
R16r: PG = PEP.
R17: PEP + ADP = PYR + ATP.

(viii) TCA cycle

R21: PYR + CoASH = Acetyl-CoA + FORMATE.
R22: Oxaloacetate + Acetyl-CoA = Citrate + CoASH.
R23r: Citrate = Isocitrate.

R24: Isocitrate + NADP = a-Ketoglutarate + NADPH + CO_2.
R27r: Succinate + NAD = Fumarate + NADH.
R28r: Fumarate = Malate.
R29r: Malate + NAD = Oxaloacetate + NADH.

(ix) Pentose Phosphate Pathway

R30: Glucose-6-P + 2NADP = Ribulose-5-P + 2NADPH + CO_2.
R31r: Ribulose-5-P = Xylulose-5-P.
R32r: Ribulose-5-P = Ribose_5_P.
R33r: Ribose-5-P + Xylulose-5-P = Sed-7-P + GA-3-P.
R34r: GA-3-P + Sed-7-P = Erythrose-4-P + Fructrose-6-P.
R35r: Erythrose-4-P + Xylulose-5-P = GA-3-P + Fructrose-6-P.

(x) Anapleurotic reactions

R40: PEP + CO_2 = Oxaloacetate.
R41: Oxaloacetate + ATP = PEP + ADP + CO_2 .
R42: MALATE + NADP = Pyruvate + NADPH + CO_2 .
R43: Isocitrate = Glyoxylate + Succinate.
R44: Glyoxylate + Acetyl-CoA = Malate + CoASH.

(xi) Redox-associated reactions

R50: Acetyl-CoA + 2NADH = Ethanol + 2NAD + CoASH.
R51: Pyruvate + NADH = Lactate + NAD.
R53: Acetyl-CoA + ADP = Acetate + CoASH + ATP.

(xii) Biomass formation

R60: 0.0206Glucose-6-P + 0.0072Fructrose-6-P + 0.0627Ribose-5-P + 0.0361 Erythrose-4-P + 0.0129GA-3-P + 0.1338PG + 0.0720PEP + 0.2861Pyruvate + 0.2930Acetyl-CoA + 0.1481 Oxaloacetate + 0.1078 a-Ketoglutarate + 1.6548 NADPH + 1.7821ATP + 0.3548 NAD = 2.87 BIOMASS + 1.6548 NADP + 0.2930 CoASH + 0.1678 CO_2 + 1.7821 ADP + 0.3548 NADH.

(xiii) Oxidative phosphorylation/maintenance energy

R70: ATP = ADP.
R71r: NADPH + NAD = NADH + NADP.

(xiv) Membrane transport reactions

R80: Lactate = Lactate_ext.
R81: Formate = Formate_ext.
R82: Acetate = Acetate_ext.
R83: Ethanol = Ethanol_ext.
R84: Succinate = Succinate_ext.
R85: propanediol = propanediol_ext.
R86r: CO_2 = CO_2_ext.

A.3. The Aerobic Metabolic Network Model Input File Used for the Program METATOOL.

(i) ENZREV (reversible reactions)

R10r R11r R14r R15r R16r R23r R26r R27r R28r R29r R31r R32r R33r R34r R35r R71r R86r.

(ii) ENZIRREV (irreversible reactions)

R1 R4 R5 R13 R17 R20 R22 R24 R25 R30 R40 R41 R42 R43 R44 R50 R51 R53 R60 R70 R72 R80 R81 R82 R83 R84 R87.

(iii) METINT (internal metabolite declaration)

Glycerol Glycerol-3-P DHAP GA-3-P PG PEP Pyruvate Acetyl-CoA CoASH Oxaloacetate Citrate Isocitrate a-Ketoglutarate Succinyl-CoA Succinate Fumarate Malate Glyoxylate Glucose-6-P Fructrose-6-P Fructrose-16-P Ribulose-5-P Xylulose-5-P Ribose-5-P Sed-7-P Erythrose-4-P Lactate Acetate Ethanol NAD NADH ATP ADP NADP NADPH CO_2 O_2.

(iv) METEXT (external metabolite declaration)

Glycerol_ext Ethanol_ext Acetate_ext CO_2_ext Lactate_ext Succinate_ext BIOMASS O_2_ext.

(v) CAT

Reactions

(vi) Glycerol specific metabolisms

R1: Glycerol_ext = Glycerol.
R4: Glycerol + ATP = Glycerol-3-P + ADP.
R5: Glycerol-3-P + NAD = DHAP + NADH.

(vii) Glycolysis

R10r: DHAP = GA-3-P.
R11r: DHAP + GA-3-P = Fructrose-16-P.
R13: Fructrose-16-P = Fructrose-6-P.
R14r: Fructrose-6-P = Glucose-6-P.
R15r: GA-3-P + ADP + NAD = PG + ATP + NADH.
R16r: PG = PEP.
R17: PEP + ADP = Pyruvate + ATP.

(viii) TCA cycle

R20: Pyruvate + CoASH + NAD = Acetyl-CoA + CO_2 + NADH.
R22: Oxaloacetate + Acetyl-CoA = Citrate + CoASH.
R23r: Citrate = Isocitrate.
R24: Isocitrate + NADP = a-Ketoglutarate + NADPH + CO_2.
R25: a-Ketoglutarate + NAD + CoASH = Succinyl-CoA + NADH + CO_2.
R26r: Succinyl-CoA + ADP = Succinate + ATP + CoASH.
R27r: Succinate + NAD = Fumarate + NADH.

R28r: Fumarate = Malate.
R29r: Malate + NAD = Oxaloacetate + NADH.

(ix) Pentose Phosphate Pathway

R30: Glucose-6-P + 2NADP = Ribulose-5-P + 2NADPH + CO_2.
R31r: Ribulose-5-P = Xylulose-5-P.
R32r: Ribulose-5-P = Ribose_5_P.
R33r: Ribose-5-P + Xylulose-5-P = Sed-7-P + GA-3-P.
R34r: GA-3-P + Sed-7-P = Erythrose-4-P + Fructrose-6-P.
R35r: Erythrose-4-P + Xylulose-5-P = GA-3-P + Fructrose-6-P.

(x) Anapleurotic reactions

R40: PEP + CO_2 = Oxaloacetate.
R41: Oxaloacetate + ATP = PEP + ADP + CO_2.
R42: MALATE + NADP = Pyruvate + NADPH + CO_2.
R43: Isocitrate = Glyoxylate + Succinate.
R44: Glyoxylate + Acetyl-CoA = Malate + CoASH.

(xi) Redox-associated reactions

R50: Acetyl-CoA + 2NADH = Ethanol + 2NAD + CoASH.
R51: Pyruvate + NADH = Lactate + NAD.
R52: Pyruvate = CO_2 + Acetate.
R53: Acetyl-CoA + ADP = Acetate + CoASH + ATP.

(xii) Biomass formation

R60: 0.0206Glucose-6-P + 0.0072Fructrose-6-P + 0.0627Ribose-5-P + 0.0361 Erythrose-4-P + + 0.0129GA-3-P + 0.1338PG + 0.0720PEP + 0.2861Pyruvate + 0.2930Acetyl-CoA + 0.1481 Oxaloacetate + 0.1078 a-Ketoglutarate + 1.6548 NADPH + 1.7821ATP + 0.3548 NAD = 2.87 BIOMASS + 1.6548 NADP + 0.2930 CoASH + 0.1678 CO_2 + 1.7821 ADP + 0.3548 NADH.

(xiii) Oxidative phosphorylation/maintenance energy:

R70: ATP = ADP.
R71r: NADPH + NAD = NADH + NADP.
R72: NADH + 2ADP + 1/2O_2 = NAD + 2ATP.

(xiv) Membrane transport reactions

R80: Lactate = Lactate_ext.
R82: Acetate = Acetate_ext.
R83: Ethanol = Ethanol_ext.
R84: Succinate = Succinate_ext.
R86r: CO_2 = CO_2_ext.
R87: O_2_ext = O_2.

A.4. Degree of Reduction of Considered Substrates and Products (see Table 2).

Nomenclature

Acetyl-CoA:	Acetyl-coenzyme A
ADP:	Adenosine diphosphate
ATP:	Adenosine triphosphate
DHA:	Dihydroxyacetone
DHAP:	Dihydroxyacetone phosphate
Erythrose-4-P:	Erythrose-4-phosphate
Fructose-6-P:	Fructose-6-phosphate
GA-3-P:	Glyceraldehyde-3-phosphate
Glucose-6-P:	Glucose-6-phosphate
NAD:	Nicotinamide adenine dinucleotide
NADH:	Nicotinamide adenine dinucleotide
1,2-PDO:	1,2-propanediol
PEP:	Phosphoenolpyruvate
PG:	Phosphoglycerate
Ribose-5-P:	Ribose-5-phosphate
Ribulose-5-P:	Ribulose-5-phosphate
Sed-7-P:	Sedoheptulose-7-P
Succinyl-CoA:	Succinyl-coenzyme A
Xylulose-5-P:	Xylulose-5-phosphate.

References

[1] S. S. Yazdani and R. Gonzalez, "Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry," *Current Opinion in Biotechnology*, vol. 18, no. 3, pp. 213–219, 2007.
[2] Y. Xu, H. Liu, W. Du, Y. Sun, X. Ou, and D. Liu, "Integrated production for biodiesel and 1,3-propanediol with lipase-catalyzed transesterification and fermentation," *Biotechnology Letters*, vol. 31, no. 9, pp. 1335–1341, 2009.
[3] Z. Chen, H. Liu, and D. Liu, "Regulation of 3-hydroxypropionaldehyde accumulation in Klebsiella pneumoniae by overexpression of dhaT and dhaD genes," *Enzyme and Microbial Technology*, vol. 45, no. 4, pp. 305–309, 2009.
[4] Z. Chen, H.-J. Liu, J.-A. Zhang, and D.-H. Liu, "Cell physiology and metabolic flux response of Klebsiella pneumoniae to aerobic conditions," *Process Biochemistry*, vol. 44, no. 8, pp. 862–868, 2009.
[5] S. Shams Yazdani and R. Gonzalez, "Engineering Escherichia coli for the efficient conversion of glycerol to ethanol and co-products," *Metabolic Engineering*, vol. 10, no. 6, pp. 340–351, 2008.
[6] D. Rittmann, S. N. Lindner, and V. F. Wendisch, "Engineering of a glycerol utilization pathway for amino acid production by Corynebacterium glutamicum," *Applied and Environmental Microbiology*, vol. 74, no. 20, pp. 6216–6222, 2008.
[7] I. R. Booth, "Glycerol and methylglyoxal metabolism," in *EcoSal—Escherichia coli and Salmonella: Cellular and Molecular Biology*, R. Curtis III, et al., Ed., ASM Press, Washington, DC, USA, 2005.
[8] Y. D. Kwon, S. Y. Lee, and P. Kim, "Influence of gluconeogenic phosphoenolpyruvate carboxykinase (PCK) expression on succinic acid fermentation in Escherichia coli under high bicarbonate condition," *Journal of Microbiology and Biotechnology*, vol. 16, no. 9, pp. 1448–1452, 2006.

[9] H. Lin, K.-Y. San, and G. N. Bennett, "Effect of Sorghum vulgare phosphoenolpyruvate carboxylase and Lactococcus lactis pyruvate carboxylase coexpression on succinate production in mutant strains of Escherichia coli," *Applied Microbiology and Biotechnology*, vol. 67, no. 4, pp. 515–523, 2005.

[10] S. H. Hong and S. Y. Lee, "Metabolic flux analysis for succinic acid production by recombinant Escherichia coli with amplified malic enzyme activity," *Biotechnology and Bioengineering*, vol. 74, no. 2, pp. 89–95, 2001.

[11] W. B. Freedberg and E. C. C. Lin, "Three kinds of controls affecting the expression of the glp regulon in Escherichia coli," *Journal of Bacteriology*, vol. 115, no. 3, pp. 816–823, 1973.

[12] G. Sweet, C. Gandor, R. Voegele et al., "Glycerol facilitator of Escherichia coli: cloning of glpF and identification of the glpF product," *Journal of Bacteriology*, vol. 172, no. 1, pp. 424–430, 1990.

[13] Y. Dharmadi, A. Murarka, and R. Gonzalez, "Anaerobic fermentation of glycerol by Escherichia coli: a new platform for metabolic engineering," *Biotechnology and Bioengineering*, vol. 94, no. 5, pp. 821–829, 2006.

[14] A. Murarka, Y. Dharmadi, S. S. Yazdani, and R. Gonzalez, "Fermentative utilization of glycerol by Escherichia coli and its implications for the production of fuels and chemicals," *Applied and Environmental Microbiology*, vol. 74, no. 4, pp. 1124–1135, 2008.

[15] N. E. Altaras and D. C. Cameron, "Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli," *Applied and Environmental Microbiology*, vol. 65, no. 3, pp. 1180–1185, 1999.

[16] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, "Comparison of network-based pathway analysis methods," *Trends in Biotechnology*, vol. 22, no. 8, pp. 400–405, 2004.

[17] C. H. Schilling, D. Letscher, and B. O. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 229–248, 2000.

[18] S. Schuster, T. Dandekar, and D. A. Fell, "Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering," *Trends in Biotechnology*, vol. 17, no. 2, pp. 53–60, 1999.

[19] J. O. Krömer, C. Wittmann, H. Schröder, and E. Heinzle, "Metabolic pathway analysis for rational design of L-methionine production by Escherichia coli and Corynebacterium glutamicum," *Metabolic Engineering*, vol. 8, no. 4, pp. 353–369, 2006.

[20] R. Carlson, D. Fell, and F. Srienc, "Metabolic pathway analysis of a recombinant yeast for rational strain development," *Biotechnology and Bioengineering*, vol. 79, no. 2, pp. 121–134, 2002.

[21] N. C. Duarte, B. ∅. Palsson, and P. Fu, "Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae," *BMC Genomics*, vol. 5, article 63, 2004.

[22] J. C. Liao and M.-K. Oh, "Toward predicting metabolic fluxes in metabolically engineered strains," *Metabolic Engineering*, vol. 1, no. 3, pp. 214–223, 1999.

[23] R. Carlson and F. Srienc, "Fundamental Escherichia coli biochemical pathways for biomass and energy production: creation of overall flux states," *Biotechnology and Bioengineering*, vol. 86, no. 2, pp. 149–162, 2004.

[24] N. Vijayasankaran, R. Carlson, and F. Srienc, "Metabolic pathway structures for recombinant protein synthesis in Escherichia coli," *Applied Microbiology and Biotechnology*, vol. 68, no. 6, pp. 737–746, 2005.

[25] Q. Zhang and Z. Xiu, "Metabolic pathway analysis of glycerol metabolism in Klebsiella pneumoniae incorporating oxygen regulatory system," *Biotechnology Progress*, vol. 25, no. 1, pp. 103–115, 2009.

[26] U. Sauer, F. Canonaco, S. Heri, A. Perrenoud, and E. Fischer, "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli," *Journal of Biological Chemistry*, vol. 279, no. 8, pp. 6613–6619, 2004.

[27] R. Carlson and F. Srienc, "Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions," *Biotechnology and Bioengineering*, vol. 85, no. 1, pp. 1–19, 2004.

[28] J. S. Edwards and B. O. Palsson, "Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions," *BMC Bioinformatics*, vol. 1, article 1, 2000.

[29] A. Hasona, Y. Kim, F. G. Healy, L. O. Ingram, and K. T. Shanmugam, "Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of Escherichia coli on xylose," *Journal of Bacteriology*, vol. 186, no. 22, pp. 7593–7600, 2004.

[30] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, WH Freeman, 5th edition, 2002.

[31] A. von Kamp and S. Schuster, "Metatool 5.0: fast and flexible elementary modes analysis," *Bioinformatics*, vol. 22, no. 15, pp. 1930–1931, 2006.

[32] T. Pfeiffer, I. Sánchez-Valdenebro, J. C. Nuño, F. Montero, and S. Schuster, "METATOOL: for studying metabolic networks," *Bioinformatics*, vol. 15, no. 3, pp. 251–257, 1999.

[33] C. T. Trinh and F. Srienc, "Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol," *Applied and Environmental Microbiology*, vol. 75, no. 21, pp. 6696–6705, 2009.

[34] H. Lin, G. N. Bennett, and K.-Y. San, "Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield," *Metabolic Engineering*, vol. 7, no. 2, pp. 116–127, 2005.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 415148, 10 pages
doi:10.1155/2010/415148

Research Article

Optimal Fluxes, Reaction Replaceability, and Response to Enzymopathies in the Human Red Blood Cell

A. De Martino,[1,2] D. Granata,[2] E. Marinari,[2] C. Martelli,[3] and V. Van Kerrebroeck[2]

[1] *CNR Institute for Physico-Chemical Processes (IPCF), Rome Sapienza Unit, Roma, Italy*
[2] *Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy*
[3] *Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA*

Correspondence should be addressed to A. De Martino, andrea.demartino@roma1.infn.it

Received 30 November 2009; Accepted 14 May 2010

Academic Editor: Jamey Young

Characterizing the capabilities, key dependencies, and response to perturbations of genome-scale metabolic networks is a basic problem with important applications. A key question concerns the identification of the potentially most harmful reaction knockouts. The integration of combinatorial methods with sampling techniques to explore the space of viable flux states may provide crucial insights on this issue. We assess the replaceability of every metabolic conversion in the human red blood cell by enumerating the alternative paths from substrate to product, obtaining a complete map of he potential damage of single enzymopathies. Sampling the space of optimal steady state fluxes in the healthy and in the mutated cell reveals both correlations and complementarity between topologic and dynamical aspects.

1. Introduction

Understanding metabolic activity from the underlying genotype is one of the most addressed problems in computational biology. Of particular interest is the issue of the identification of the reactions that are indispensable for an organism to survive, grow or perform a specific function in a given growth medium or, conversely, of the potentially most harmful knock-outs or enzymopathies. Several experimental protocols are able to assess the essentiality of gene products (and hence of the corresponding metabolic reactions), ranging from individual knock-outs to transposon mutagenesis and RNA interference [1–5]. Computational approaches on the other hand might provide important clues on the system-level organization by investigating genome-scale network reconstructions.

The functional modularity of metabolic networks suggests that topological aspects may provide a key to identify a class of essential pathways [6, 7]. However the metabolic genotype only constitutes the frame on the top of which the dynamic phenotype is built. The essentiality of a metabolic pathway will in general depend on both structural considerations based on the network reconstruction from genomic information, and on the "model of metabolism" defined on it, for example, on the corresponding steady state fluxes. In *E.coli*, phenotypical essentiality of metabolic genes has been associated with a reduced allowed variability of the corresponding fluxes, suggesting that dynamically stiff reactions may constitute an evolutionarily robust backbone of metabolism conserved over different species [8].

Here we attempt a more thorough integration of topological and dynamical views to obtain a more comprehensive insight into a metabolic network's organization, efficiency, and ability to respond to perturbations. We will first associate the essentiality of a reaction with a measure of its topological replaceability by enumerating the alternative paths from substrate to product along the network edges, with the rationale that from a purely structural viewpoint more replaceable reactions are less likely to be crucial nodes of the network. Then we will validate and compare the essentiality map thus obtained with the metabolic phenotype resulting from the definition of a general constraint-based model for metabolic flux prediction. We shall see that dynamical and structural measures of essentiality may offer complementary views of a reaction network's robustness.

We carry out our analysis on the metabolic network of the human red blood cell (hRBC), one of the most studied complexes in systems biology, from the earliest mathematical models of single biochemical pathways [9, 10] to the currently available genome-scale reconstructions [11]. The reason for this choice lies essentially in its limited size. On the one hand, it allows to compute reaction replaceabilities *exactly* by a suitable modification of Johnson's algorithm for counting loops in a directed graph [12]. On the other, it allows for the efficient application of various sampling methods to the space of viable flux states [8, 13]. The latter is a vital ingredient to address many important properties of erythrocytes. Indeed for some organisms under certain conditions it is reasonable to assume that the metabolic activity is aimed at maximizing a subset of the metabolic reactions (or a function of them) associated with a certain biological function. In such cases the relevant flux configuration can be computed by standard optimization algorithms. For example, *E. coli*'s metabolism has been shown to maximize biomass production under evolutionary pressure [14], but after a genetic knockout it responds with a minimum rearrangement of fluxes [15]. While the production of the cofactors ensuring the maintenance of osmotic balance and the release of oxygen may be argued to be their metabolic goal, erythrocytes do not generically allow for such a simplification. Information-rich directions in flux space must be retrieved by coupling the underlying constraints on fluxes with other types of analyses. Much understanding has indeed been obtained from the uniform sampling of feasible states [13, 16, 17] and by functional studies, like the computation of extreme pathways [18], of metabolic regulatory structures [19, 20] and of metabolic pools [21]. These aspects combined make hRBCs a key benchmark for both theories of metabolism and computational tools.

It is worth noting that the detailed structural information we derive (i.e., the full map of alternative paths for each substrate/product pair) cannot be retrieved by other methods. Unluckily, computation times still prevent scaling the approach we employ up to networks larger than a few hundred nodes. More refined algorithms are currently being developed to overcome this limitation.

2. Approach

2.1. Structural Analysis. Given a reaction network, we want to compute, for any pair of metabolites a and b that are, respectively, substrate and product in a reaction i (this situation will be indicated by $a \xrightarrow{i} b$), the number $\mathcal{N}^{(i)}_{a\to b}(\ell)$ of alternative pathways, excluding reaction i, of length ℓ allowing for the conversion $a \to b$, see Figure 1. The rationale is that a reaction performing a metabolite conversion $a \xrightarrow{i} b$ for which $\mathcal{N}^{(i)}_{a\to b}(\ell)$ (or, more properly, $\sum_\ell \mathcal{N}^{(i)}_{a\to b}(\ell)$) is large will be more easily substituted, in case of an enzymopathy or a knockout, than one for which the above quantity is small.

Finding paths connecting two points of a directed network is a long-studied problem in computer science. The focus is usually on locating the shortest paths or the fastest way to find any path. Enumerating *all* the distinct paths between two vertices is however a less confronted issue. In our case it is crucial to avoid overcounting, for example, due to self-intersecting paths. Therefore we shall resort to an exhaustive algorithm. We will identify the substitutive paths using the following trick: for each pair (a, b) of metabolites such that $a \xrightarrow{i} b$, revert i fictitiously. This results in a new graph where an auxiliary edge $b \xrightarrow{i} a$ replaced the edge $a \xrightarrow{i} b$, see again Figure 1. Counting the number of alternative reaction chains producing b from a then comes down to computing the number of directed cycles, that is, non self-intersecting directed closed paths along the edges of the new graph, passing through the fictitious edge $b \xrightarrow{i} a$. Thanks to the limited size of the hRBC network it is possible to solve this enumeration problem exactly via Johnson's algorithm [12], briefly described in the following section. $\mathcal{N}^{(i)}_{a\to b}(\ell)$ can now be trivially inferred. For simplicity, ℓ will denote here the number of reactions in the alternative pathway ($\ell = 4$ in Figure 1).

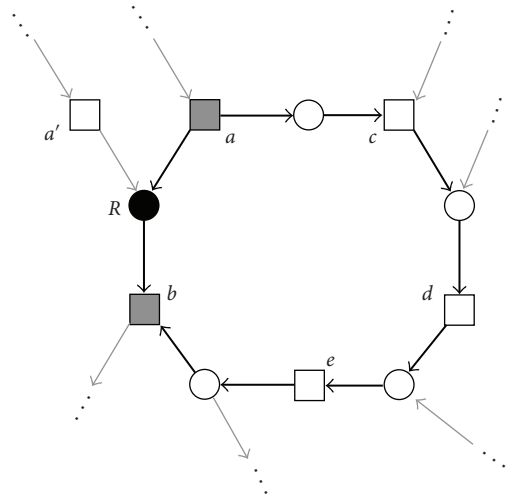


Figure 1: Bipartite graph representation of a reaction network, with circles (resp. squares) denoting reactions (resp. metabolites). Here, reaction R uses metabolites a and a' as substrates to produce metabolite b. If R is removed, the conversion of a to b is still permitted by the alternative pathway $a \to c \to d \to e \to b$. When R is fictitiously reversed, this chain forms a directed loop of length 5 reactions, formed by R and by a path passing through $\ell = 4$ other reactions.

2.2. Flux Analysis. The space of viable fluxes will be defined through a constraint-based approach which relies on more general assumptions than flux-balance analysis (FBA, [22]). FBA is the standard method to model steady-state reaction networks where mass balance constraints are imposed to every metabolite. For a reaction network with N reactions and M metabolites, let us denote by $\mathbf{A}$ and $\mathbf{B}$, respectively, the $M \times N$ matrices of output and input stoichiometric coefficients. The stoichiometric matrix is given by $\mathbf{S} = \mathbf{A} - \mathbf{B}$. Letting $\boldsymbol{\nu} = (\nu_i)_{i=1}^N$ denote a vector of fluxes (with properly chosen bounds $\nu_i^{\min} \le \nu_i \le \nu_i^{\max}$), the concentrations

$\mathbf{c} = (c^a)_{a=1}^M$ of metabolites vary in time according to $\dot{\mathbf{c}} = \mathbf{S}\boldsymbol{\nu} - \mathbf{u}$, where $\mathbf{u} = (u^a)_{a=1}^M$ stands for the net cellular uptake of metabolite a ($u^a > 0$ if a is a global output of metabolism, $u^a < 0$ if a is consumed by the organism, $u^a = 0$ if a is mass-balanced). Assuming a steady state, the concentrations are constant in time (i.e., $\dot{\mathbf{c}} = 0$) and vectors $\boldsymbol{\nu}$ satisfying $\mathbf{S}\boldsymbol{\nu} = \mathbf{u}$, or

$$(\mathbf{A} - \mathbf{B})\boldsymbol{\nu} = \mathbf{u}, \tag{1}$$

represent flux configurations ensuring that each metabolite meets its production or consumption constraints at fixed concentrations. As N is typically larger than M, the system is underdetermined and feasible flux states form a convex set of dimension N-rank($\mathbf{S}$) embedded in the N-dimensional space of fluxes. In absence of a selection criterion that allows to pick one solution out of this set (as e.g., a maximum biomass principle), a uniform sampling of the solution space should be carried out. When N is sufficiently small (as for hRBCs), this can be achieved effectively, albeit at a considerable computational cost, by Monte Carlo methods [13, 16] or by message-passing procedures [17].

Here we will consider a different but related flux scheme based on Von Neumann's (VN) model of reaction networks [8]. In the VN framework, one fixes the environment through a small set of intakes on nutrients and defines a self-consistent flux problem where the network chooses, given a target growth rate, how much of the nutrients to use and which metabolites are globally produced. Mass balance then emerges as a property of the solutions for some metabolites.

The equations describing the VN model have been studied by statistical mechanics methods in [23, 24]. For an intuitive derivation, note that the quantities $\mathbf{A}\boldsymbol{\nu}$ and $\mathbf{B}\boldsymbol{\nu}$ represent, respectively, the total output and the total input of each metabolite for a given flux vector $\boldsymbol{\nu}$. Then a flux vector such that $\mathbf{A}\boldsymbol{\nu} \geq \rho\mathbf{B}\boldsymbol{\nu}$, with some constant $\rho > 0$, describes a network state where metabolites are being produced at a rate at least equal to ρ, since for each of them the total output is at least ρ times the total input. It is simple to see that as ρ increases the volume of such flux vectors shrinks continuously (for $\rho = 0$ every flux vector is a solution). In particular, there exists a value $\rho^\star$ of ρ, representing the maximum metabolic production rate compatible with the stoichiometric constraints, above which no suitable flux vectors exist. The presence of conserved metabolic pools [25] implies $\rho^\star = 1$ [26], so that in metabolic networks optimal steady state fluxes correspond to the solutions of

$$(\mathbf{A} - \mathbf{B})\boldsymbol{\nu} \geq \mathbf{0}. \tag{2}$$

The solutions of (2) do not coincide with those of (1) even for $\mathbf{u} = \mathbf{0}$. Interestingly, a finite volume of (optimal) flux states turns out to satisfy the above constraints [8]. This trait is at odds with both the behavior of the solutions of (2) for a random reaction network (where a single solution survives at $\rho^\star$ [24]) and with the optimization that is usually coupled to FBA (where typically a single flux state maximizes the objective function), and points to the robustness of metabolic phenotypes. For *E.coli*, in particular, the solutions of (2) have been shown to reproduce both the large-scale organization of fluxes and the individual measured rates. In addition, fluxes with the smallest solution-to-solution fluctuations, representing the most susceptible parts of the network, turn out to be strongly correlated with *E.coli*'s phenomenologically essential genes [8]. The main technical advantage in using the VN scheme lies in the fact that its solution space can be sampled uniformly at very modest computational costs even for genome-scale models. The algorithm allowing for this, which has been recently applied to sample *E.coli*'s solution space [8], is detailed in the following section. Its running times for hRBCs are negligible.

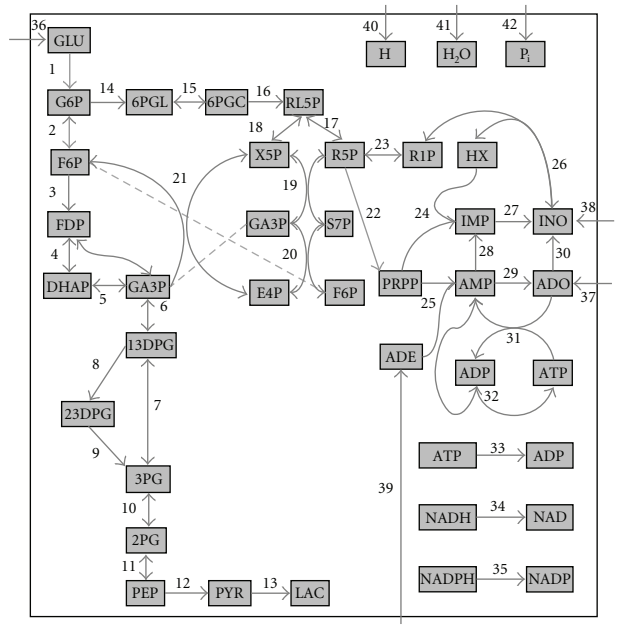


Figure 2: Scheme of the hRBC metabolic network used in our analisys. Squares correspond to metabolites, numbers to reactions (see Table 1).

3. Methods

3.1. Reconstructed Network. We consider the hRBC metabolic network studied in [16], a map of which is shown in Figure 2; Table 1 lists reactions and the corresponding abbreviations. The network comprises three main pathways, namely, glycolysis (reactions 1–13), the pentose phosphate (PP) pathway (14–21) and the adenosine metabolism, with a total of $M = 39$ metabolites linked by $N = 59$ reactions: 49 internal reactions (34 of which come from the splitting of 17 reversible processes), 3 auxiliary fluxes to maintain the osmotic equilibrium and the redox state of the cell (ATPase, NADHase, NADPHase) and 7 uptake reactions to guarantee the intake of the necessary nutrients (GLU, ADE, ADO, INO), and of the cytosol elements (H_2O, H, P_i). The forward and backward parts of reversible reactions are treated separately throughout this study, both in the structural and in the flux analysis.

Table 1: List of reactions considered in this work, including the corresponding number in the map of Figure 2, the abbreviation and the process. The 7 uptake fluxes, numbered 36 to 42, are as shown in Figure 2.

No.	Abbreviation	Chemical reaction
1	HK	GLU + ATP → G6P + ADP + H
2	PGI	G6P ↔ F6P
3	PFK	F6P + ATP → FDP + ADP + H
4	ALD	FDP ↔ GA3P + DHAP
5	TPI	DHAP ↔ GA3P
6	GAPDH	GA3P+NAD+ P_i ↔ 13DPG+NADH+H
7	PGK	13DPG + ADP ↔ 3PG + ATP
8	DPGM	13DPG → 23DPG + H
9	DPGase	23DPG + H2O → 3PG + P_i
10	PGM	3PG ↔ 2PG
11	EN	2PG ↔ PEP + H2O
12	PK	PEP + ADP + H → PYR + ATP
13	LDH	PYR + NADH + H ↔ LAC + NAD
14	G6PDH	G6P + NADP → 6PGL + NADPH + H
15	PGL	6PGL + H_2O ↔ 6PGC + H
16	PDGH	6PGC + NADP → RL5P + NADPH + CO_2
17	RPI	RL5P ↔ R5P
18	XPI	RL5P ↔ X5P
19	TKI	X5P + R5P ↔ S7P + GA3P
20	TA	GA3P + S7P ↔ E4P + F6P
21	TKII	X5P + E4P ↔ F6P + GA3P
22	PRPPsyn	R5P + ATP → PRPP + AMP
23	PRM	R1P ↔ R5P
24	HGPRT	PRPP + HX + H_2O → IMP + 2P_i
25	AdPRT	PRPP + ADE + H_2O → AMP + 2P_i
26	PNPase	INO + Pi ↔ HX + R1P
27	IMPase	IMP + H_2O → INO + P_i + H
28	AMPDA	AMP + H_2O → IMP + NH_3
29	AMPase	AMP + H_2O → ADO + P_i + H
30	ADA	ADO + H_2O → INO + NH_3
31	AK	ADO + ATP → ADP + AMP
32	ApK	2ADP ↔ ATP + AMP
33	ATPase	ATP → ADP + P_i
34	NADHase	NADH → NAD + H
35	NADPHase	NADPH → NADP + H

3.2. Structural Analysis. Structural vulnerabilities are identified by analyzing the loop structure of a modified metabolic reaction network, created from the original one by inverting—in turn—the direction of the single reaction for which we want to compute the replaceability, as explained in Figure 1. The fastest known exact algorithm (for the worst case scenario) of this cycle enumeration problem for a directed graph was introduced by Johnson [12]. We shall now shortly describe its key ideas, referring to [12] for a pseudocode.

Given a directed graph with n vertices and e edges, the algorithm is designed to build non self-intersecting paths from a root vertex r to itself, loading them onto stacks. The main ingredients allowing for an optimal exploration of the graph are (a) a smart choice of the root vertex, and (b) an efficient method to avoid duplicating cycles and repeating searches on the same portions of the graph. To achieve this, vertices are initially ordered in a lexicographic sequence, and the algorithm only selects as roots those nodes that are the "least" vertex (in the initial ordering) of at least one cycle. The algorithm described in [27] guarantees to find such vertices in $O(n + e)$ operations. Moreover, to avoid self-intersections, each time a node is loaded onto a stack it is also given a "blocked" status. It was proven by Johnson that if a vertex v stays blocked as long as every path from v to the root vertex r intersects the current path at a vertex

other than r, the algorithm outputs all cycles exactly once. By sufficiently delaying the unblocking of each of these vertices and by keeping track of the portions of the graph that have been searched holding the current stack, the maximum time that can elapse between two consecutive cycle outputs can be reduced to $O(n+e)$. The same holds for the time window before the first cycle is delivered and for the one after the output of the last cycle. Hence, the total time needed to list the, say, c cycles of the graph is $O((n+e)(c+1))$. In our case, each fictitious reaction reversal generates a new graph, so that computing the complete substitutability map for a network of N reactions requires a time of the order $O(N(n+e)(c+1))$. For practical reasons, we perform this analysis on the bipartite metabolic network (as in Figure 1) rather than the reduced network of Figure 2. This implies that in our case $n = N + M$.

One can in principle consider different measures of replaceability of a metabolic conversion $a \xrightarrow{i} b$. The quantity $\mathcal{R}^{(i)}_{a\to b} = \sum_\ell \mathcal{N}^{(i)}_{a\to b}(\ell)$, counting the total number of paths alternative to i from a to b of any length, is perhaps the most obvious option. Taking into account the fact that, typically, longer detours can be less convenient than shorter ones from an energetic viewpoint one could instead consider ℓ-weighted functions like $\mathcal{W}^{(i)}_{a\to b} = \sum_\ell \exp(-\ell)\mathcal{N}^{(i)}_{a\to b}(\ell)$, with the caveat that shorter pathways might require more ATP than longer ones. $\mathcal{R}$-based and $\mathcal{W}$-based rankings of metabolic conversions are rather different. They are fully available from http://chimera.roma1.infn.it/SYSBIO. To focus on the basics, here we limit ourselves to identifying three key reaction groups that are *independent* of the replaceability measure used:

(a) the group of reactions such that each substrate-product pair involved in them can be substituted (this is putatively the part of the network that is most robust to enzymopathies);

(b) the group of reactions that cannot be substituted, corresponding to the most harmful enzymopathies;

(c) the group of reversible reactions that are only replaceable in one direction, corresponding to the situation in which a conversion $a \leftrightarrow b$ can only be substituted in one direction in case of a knockout.

All essentiality maps we show relate to this classification. Note that, for topological reasons, intakes are not replaceable.

3.3. Flux Analysis. Optimal flux vectors, that is, solutions of (2), are computed by the algorithm introduced in [24] based on [28]. The idea is to modify fluxes iteratively until all inequalities in (2) are satisfied. Specifically, for a fixed $0 \le \rho < \rho^\star$ (with $\rho^\star = 1$ in our case) define $\Xi(\rho) = \mathbf{A} - \rho\mathbf{B}$ and let $\boldsymbol{\xi}^a(\rho)$ denote the rows of $\Xi(\rho)$, for $a \in \{1,\dots,M\}$. Let also, for each iteration step t, $\boldsymbol{\nu}(t)$ be the flux vector at step t and

$$m(t) = \arg\min_a \boldsymbol{\xi}^a(\rho)\cdot\boldsymbol{\nu}(t), \tag{3}$$

At each t, the algorithm runs as follows. If $\boldsymbol{\xi}^{m(t)}(\rho)\cdot\boldsymbol{\nu}(t) < 0$, update fluxes according to

$$\nu_i(t+1) = \max\left\{0, \nu_i(t) + \xi_i^{m(t)}(\rho)\right\}, \tag{4}$$

and iterate in t. Else, if $\boldsymbol{\xi}^{m(t)}(\rho)\cdot\boldsymbol{\nu}(t) \ge 0$ stop, that is, $\boldsymbol{\nu}(t)$ is a solution.

Convergence to a solution is rigorously ensured for all $0 \le \rho < \rho^\star$, and $\rho^\star$ can be approximated with the desired resolution by iterating the above process for increasing values of ρ [24]. To guarantee that solutions are well defined one can either resort to setting fixed upper bounds on ν_i's or, as we do, impose a linear constraint of the form $\sum_i \nu_i(t) = N$ on the solutions (this is equivalent to singling out one flux as the reference unit for the other fluxes). It is convenient to initialize the algorithm with a random vector $\boldsymbol{\nu}(0)$. Different initial points generate trajectories to different solutions at $\rho^\star$ and the sampling of the solution space thus obtained turns out to be uniform [8].

Contrary to FBA, the solution space of VN's model is generically not a polytope. Indeed much useful information can be retrieved from its shape. As a means to characterize it we employ the average overlap between different optimal flux vectors, defined as follows. Let $\boldsymbol{\nu}_\alpha$ and $\boldsymbol{\nu}_\beta$ denote two distinct solution vectors of (2) and, for each flux i, let

$$q_{\alpha\beta}(i) = \frac{2\nu_{i\alpha}\nu_{i\beta}}{\nu_{i\alpha}^2 + \nu_{i\beta}^2}. \tag{5}$$

This quantity, called the "overlap" between solutions α and β, equals 1 if flux i takes on the same value in solutions α and β and decreases as the values differ more and more. Averaging $q_{\alpha\beta}(i)$ over different pairs of solutions provides a measure of the allowed variability of flux i (smaller variability corresponds to larger average overlap), complementary to the standard deviation of the resulting flux distribution. The complexity of the solution space can then be roughly understood by distinguishing narrower directions with larger overlap or less variable fluxes from broader ones. It is reasonable to think that a cell will be more sensitive to perturbations (e.g., knockouts) of fluxes with larger overlap. Analyzing the susceptibility of the solution space to perturbations along the directions identified by different fluxes then allows to extract a list of the potentially more deleterious perturbations, in analogy with previous work on *E.coli* [8].

3.4. Response to Enzymopathies. In order to test the hRBC network against enzymopathies, we can focus on two types of perturbations. One can first employ a structural criterion: the knockout of a metabolic conversion $a \to b$ that is less easily "substituted" is more likely to be deleterious for the cell than the knockout of a highly replaceable conversion. As said above, we concentrate here on a coarse-grained view of replaceability based on classifying reactions into the groups (a), (b), and (c) defined above, with groups (b) and (c) containing potentially essential reactions. The second criterion is based on fluxes: fluxes with smaller allowed variability (i.e., larger overlap) in the healthy cell are more likely to be essential links of the network than fluxes whose

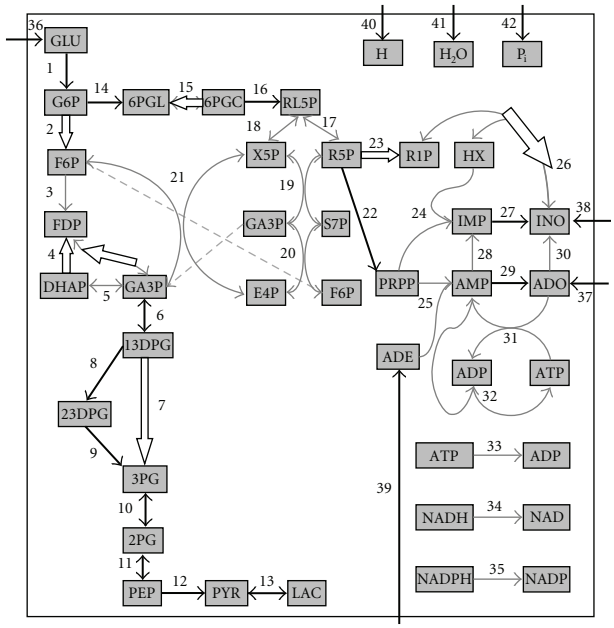


Figure 3: hRBC's structural reaction replaceability map. Thick black arrows denote unreplaceable reactions (group (b) above); thick white-filled arrows denote reversible reactions that can be replaced only in the direction indicated by the arrow (group (c)); all other reactions are fully replaceable (group (a)).

value can be changed over a larger range without losing optimality.

As is to be expected, the essentiality maps produced in these ways have a large degree of similarity, and reactions in the group (b) discussed above coincide with the physiologically most critical parts of hRBC's metabolism. The simplest way to simulate an enzymopathy on flux i is to constrain its value below a certain upper bound $\overline{v}_i$. Deficiencies can be partial, that is, of a smaller degree, the closer $\overline{v}_i$ is to the upper limit of the allowed range in the healthy cell, or total if $\overline{v}_i = 0$. Such constraints cause in principle a modification of the solution space along the direction i which in turn cascades on the entire volume, modifying the optimal states of the metabolic network.

4. Results

4.1. Structural Analysis. The substitutability map derived from the loop analysis is displayed in Figure 3. (For the sake of simplicity we exclude the highly replaceable currency exchange fluxes from this discussion.) The most replaceable core of the network lies in the PP pathway (reactions 17–21), which constitutes the main source of NADPH, the key metabolite that in erythrocytes limits the accumulation of peroxides protecting the cell from hemolysis. The high reliability coming with replaceability partly explains the reason why this group of reactions plays a central role not just as an auxiliary pathway for glycolysis, see the following analysis of fluxes. Unreplaceable reactions are instead lined up along glycolysis (numbers 1,6,8–13), in the bridge between glycolysis and the PP-pathway (14 and 16) or in auxiliary modules (22, 27, 29; the ADE → AMP conversion in 25 is also not replaceable being directly linked to the ADE uptake). The physiologically most deleterious knockouts (HK, PK, and G6PDH) all belong to this group. For instance, deficiency in the level of G6PDH is the basis of different types of hemolytic anemias, including favism, and is also linked to malaria resistance [29]. Finally, there is a group of reversible reactions (numbers 2, 4, 7, 15, 23, 26) that can be replaced only in one direction. Note however that the last three of these could still be replaced in case of an enzymopathy if a proper medium is selected. For instance, if reaction 15 is removed, it could be substituted by an alternative chain of reactions provided 6PGC is externally supplied. This is instead not possible for reaction 4 and possibly 23 (depending on the directionality of reaction 26), as a knockout in these cases would necessarily result in a net production of FDP and R1P.

4.2. Flux Analysis. The flux distribution corresponding to optimal states in the healthy and enzyme deficient hRBC are displayed reaction by reaction in Figure 4, obtained by sampling 10000 solutions of (2), while a pictorial representation of the optimal flux states is given in Figure 5. For the healthy cell (black line in Figure 4 and top left panel in Figure 5) the large flux backbone is formed by the second part of glycolysis (crucial for ATP, NADH, and 23DPG production) and the PP pathway (NADPH production). The latter gives a substantial contribution to the former, not just as salvage way. The adenosine metabolism shows instead lower flux values. In addition to GLU, which is the fundamental substrate for hRBCs, the INO uptake plays an important role as an alternative way to the PP pathway. It is worth stressing that these solutions imply a net production of 23DPG, the crucial regulator for oxygen release, which is obtained without any imposed constraint. This picture is strongly reminiscent of the first eigenpathway obtained by extreme pathways analysis in [19], though the thermodynamic constraints and production requirements used in [19], including one on 23DPG, are more strict than the self-consistent analysis presented here. Comparing the distributions with FBA studies on the same system [16], one notices instead a general rearrangement of fluxes in the network apart from glycolysis. A close inspection reveals that such a rearrangement is mostly quantitative, as preferred reaction directions are generically preserved, the noteworthy exception being the RPI flux, that in the VN solution strengthens the PP pathway with respect to the FBA solution. This scenario is not surprising in view of the basic difference between FBA and the VN approach. It should be kept in mind however that the a priori constraints on flux variability are quite more strict in FBA than they are in the VN model, and the flux distribution appear to be particularly sensitive to the assumed upper and lower bounds for the fluxes.

In Figure 6 we report the overlap map of the hRBC. Comparing this with Figure 3 one sees that the large overlap backbone (signaling dynamically stiff fluxes) coincides to

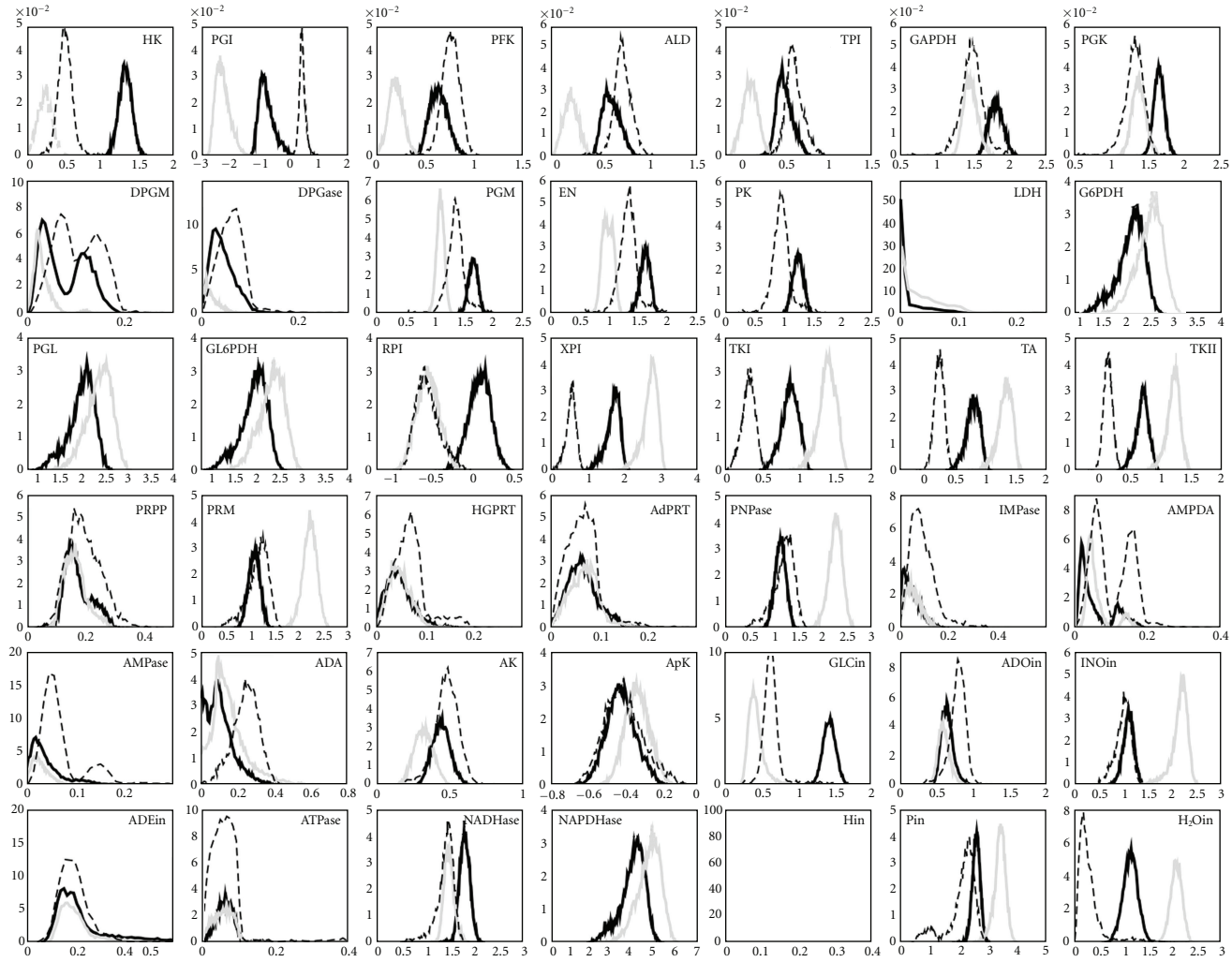


Figure 4: Distributions of differential fluxes (measured for each reaction by the difference between the direct and the reverse flux, when possible) for the healthy cell (thick black lines) and for the hRBC with knockout PK (thick gray lines) e G6PDH (dashed lines).

a large degree with the structurally most vulnerable parts of the network. Note that the overlap of reactions 2, 4, 15 and 26 is larger in the direction that cannot be replaced, further pointing to a higher susceptibility, and that currency reactions (31–35) belong to the most constrained part of the network. Revealingly, however, topological and dynamical characterizations prove to be complementary in some cases. This is seen, for example, from reaction 3, which is flux-constrained but also highly replaceable, so that the damage due to removal is limited even in presence of a small allowed dynamical range. (A similar picture holds for reaction 23.)

To conclude, we remind that in our framework uptake fluxes are optimized variables not fixed by boundary conditions. In the optimal state five of the uptakes have a limited allowed variability, implying rather severe constraints on the cell's environment.

4.3. Response to Enzymopathies. We have simulated the most studied enzymopathies by constraining the flux of the corresponding reaction. Generically speaking, the hRBC metabolism displays a large resilience against partial perturbations. Indeed, we have observed appreciable differences in relevant cellular functions compared to the nondeficient case only under full enzyme deficiencies, as also observed in [11] within a standard FBA optimization approach. Even under the most serious enzyme deficiencies the network appears to be able to maintain the production of ATP, NADH and NADPH almost constant, see also [30]. We focus here on PK and G6PDH deletions. As shown in Figure 4, the alterations in the flux distributions are not particularly striking and indeed we do not observe global flux rearrangements on the network's scale. The G6PDH enzymopathy appears to only cause local changes, confirming the structural predictions, the overlap calculations and also in agreement with clinical observations [31]. The response to PK knockout is instead more marked. The synoptic analysis of Figure 5 shows that in general the response to the perturbation consisted in a drop of the GLU uptake, and in a reduction of the glycolytic flux, while the Rapoport-Leubering shunt (reactions 8-9) for the production of 23DPG remains

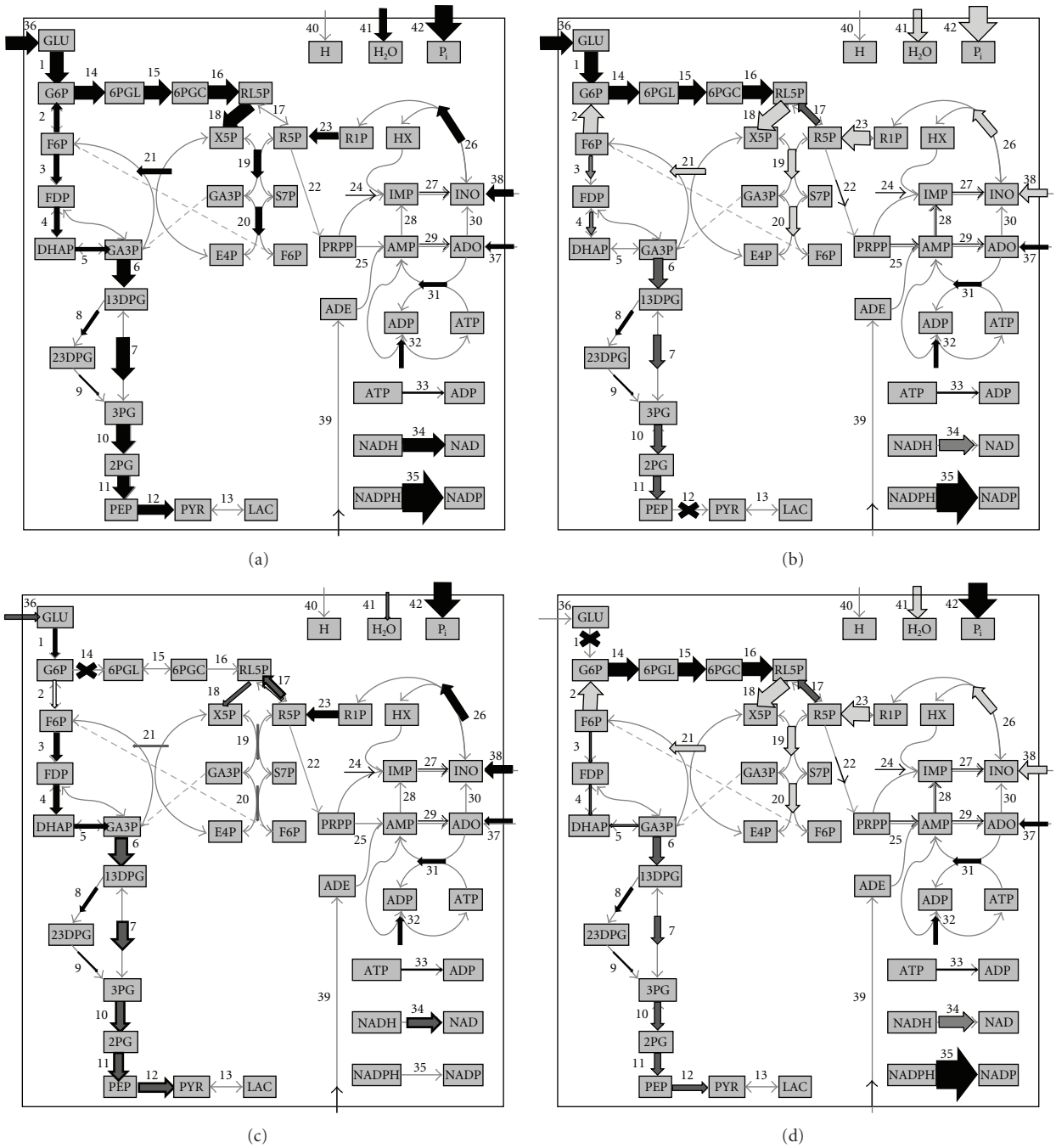


Figure 5: Maps of fluxes in the healthy hRBC (a) and flux rerouting in three full enzymopathies: PK (b), G6PDH (c) and HK (d). The thick arrows are dark grey if the reaction rates diminish with respect to the healthy cell, light grey if they are higher, black if they remains the same. The widths of the arrows are proportional to the mean value of the corresponding distributions.

particularly stable, as does the adenosine metabolism. For the glycolytic deficiencies PK and HK we further observe an increase of the INO uptake to sustain the PP pathway and allow for the second part of glycolysis, and with it the production of ATP and NADH, to take place. Detailed flux configurations corresponding to the next most severe enzymopathies (HK, EN, PGK and PGM) are available from http://chimera.roma1.infn.it/SYSBIO.

5. Final Remarks

In this work we compare two robustness measures for biochemical networks, one based on structural properties (the reaction replaceability), the other based on dynamical stiffness (the overlaps). The former can be exactly assessed by enumerating the alternative paths joining substrates and products of a given reaction in a network. The latter depends

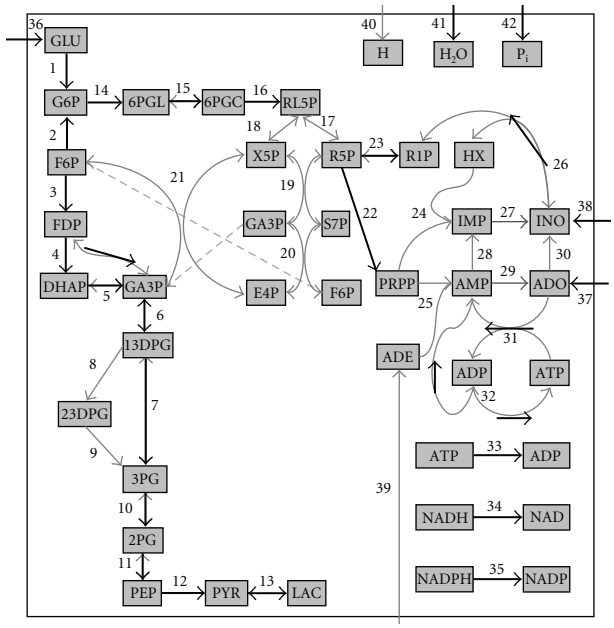


Figure 6: Large overlap backbone of hRBC: thick black arrows mark reactions with an overlap larger than 0.9 in an optimal healthy cell.

on both the network topology and the model defined on it. Within VN's frame, we found that unreplaceable reactions mostly correspond to processes with a smaller allowed flux variability. In such directions, reaction removals as well as constraints on the fluxes are expected to be generically harmful. Reactions with limited (but non zero) replaceability tend to have instead smaller overlap, so that while the reaction is difficult to substitute still its flux can be largely adjusted. In an evolutionary perspective [32], the former pathways appear as "frozen", and perturbations at these nodes will require large-scale flux rearrangement, while a mutation affecting the latter group may be neutral and could be preserved across generations. Interestingly, some reactions have both a large overlap and a large replaceability. These, albeit structurally robust, are dynamically constrained and should be considered as essential pathways of the metabolism as well. Integrating dynamical and structural characterizations may thus provide a rather complete picture of the emerging network robustness. The fact that topological and dynamical essentiality may not coincide could also prove to be important in view of the present challenges to understand the dynamical basis of topological modularity [33].

Extended flux state sampling was achieved here by an algorithm that is easily scalable to larger networks, see [8]. Exhaustive structural analysis instead was made possible by the small size of hRBC's metabolic network, which has served here as a model system to test basic concepts and algorithms. The use of the same procedure on a larger network, such as *E. coli*'s, is likely to be prevented by CPU time growth. However, message-passing algorithms, designed specifically to solve combinatorial optimization or counting problems—albeit approximately—on graphical models, may be a suitable replacement [34].

Acknowledgment

The authors would like to express their pleasure to thank S. Jain for pointing them to [7] and D. Segrè for important comments and suggestions. This work was supported by the IIT (Italian Institute of Technology) through the Seed Project DREAM.

References

[1] K. Kobayashi, S. D. Ehrlich, A. Albertini et al., "Essential Bacillus subtilis genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 8, pp. 4678–4683, 2003.

[2] C. M. Sassetti, D. H. Boyd, and E. J. Rubin, "Comprehensive identification of conditionally essential genes in mycobacteria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 22, pp. 12712–12717, 2001.

[3] B. J. Akerley, E. J. Rubin, V. L. Novick, K. Amaya, N. Judson, and J. J. Mekalanos, "A genome-scale analysis for identification of genes required for growth or survival of Haemophilus influenzae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 2, pp. 966–971, 2002.

[4] S. Y. Gerdes, M. D. Scholle, J. W. Campbell et al., "Experimental determination and system level analysis of essential genes in Escherichia coli MG1655," *Journal of Bacteriology*, vol. 185, no. 19, pp. 5673–5684, 2003.

[5] R. S. Kamath, A. G. Fraser, Y. Dong et al., "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi," *Nature*, vol. 421, no. 6920, pp. 231–237, 2003.

[6] R. Mahadevan and B. ∅. Palsson, "Properties of metabolic networks: structure versus function," *Biophysical Journal*, vol. 88, no. 1, pp. L07–L09, 2005.

[7] A. Samal, S. Singh, V. Giri, S. Krishna, N. Raghuram, and S. Jain, "Low degree metabolites explain essential reactions and enhance modularity in biological netwoks," *BMC Bioinformatics*, vol. 7, article 118, 2006.

[8] C. Martelli, A. De Martino, E. Marinari, M. Marsili, and I. P. Castillo, "Identifying essential genes in Escherichia coli from a metabolic. optimization principle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 8, pp. 2607–2611, 2009.

[9] T. A. Rapoport, R. Heinrich, and S. M. Rapoport, "The regulatory principles of glycolysis in erythrocytes in vivo and in vitro. A minimal comprehensive model describing steady states, quasi-steady states and time-dependent processes," *Biochemical Journal*, vol. 154, no. 2, pp. 449–469, 1976.

[10] H.-G. Holzhutter, G. Jacobasch, and A. Bisdorff, "Mathematical modelling of metabolic pathways affected by an enzyme deficiency. A mathematical model of glycolysis in normal and pyruvate-kinase-deficient red blood cells," *European Journal of Biochemistry*, vol. 149, no. 1, pp. 101–111, 1985.

[11] N. Jamshidi and B. ∅. Palsson, "Systems biology of the human red blood cell," *Blood Cells, Molecules, and Diseases*, vol. 36, no. 2, pp. 239–247, 2006.

[12] D. B. Johnson, "Finding all the elemtary circuits of a directed graph," *SIAM Journal on Computing*, vol. 4, p. 77, 1975.

[13] S. J. Wiback, I. Famili, H. J. Greenberg, and B. ∅. Palsson, "Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space," *Journal of Theoretical Biology*, vol. 228, no. 4, pp. 437–447, 2004.

[14] R. U. Ibarra, J. S. Edwards, and B. ∅. Palsson, "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," *Nature*, vol. 420, no. 6912, pp. 186–189, 2002.

[15] D. Segrè, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, pp. 15112–15117, 2002.

[16] N. D. Price, J. Schellenberger, and B. ∅. Palsson, "Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies," *Biophysical Journal*, vol. 87, no. 4, pp. 2172–2186, 2004.

[17] A. Braunstein, R. Mulet, and A. Pagnani, "Estimating the size of the solution space of metabolic networks," *BMC Bioinformatics*, vol. 9, article 240, 2008.

[18] S. J. Wiback and B. ∅. Palsson, "Extreme pathway analysis of human red blood cell metabolism," *Biophysical Journal*, vol. 83, no. 2, pp. 808–818, 2002.

[19] N. D. Price, J. L. Reed, J. A. Papin, S. J. Wiback, and B. ∅. Palsson, "Network-based analysis of metabolic regulation in the human red blood cell," *Journal of Theoretical Biology*, vol. 225, no. 2, pp. 185–194, 2003.

[20] C. L. Barrett, N. D. Price, and B. ∅. Palsson, "Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition," *BMC Bioinformatics*, vol. 7, article 132, 2006.

[21] K. J. Kauffman, J. D. Pajerowski, N. Jamshidi, B. ∅. Palsson, and J. S. Edwards, "Description and analysis of metabolic connectivity and dynamics in the human red blood cell," *Biophysical Journal*, vol. 83, no. 2, pp. 646–662, 2002.

[22] K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," *Current Opinion in Biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.

[23] A. De Martino and M. Marsili, "Typical properties of optimal growth in the von Neumann expanding model for large random economies," *Journal of Statistical Mechanics*, no. 9, pp. 19–27, 2005.

[24] A. De Martino, C. Martelli, R. Monasson, and I. Pérez Castillo, "Von Neumann's expanding model on random graphs," *Journal of Statistical Mechanics*, no. 5, Article ID P05012, 2007.

[25] I. Famili and B. ∅. Palsson, "The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools," *Biophysical Journal*, vol. 85, no. 1, pp. 16–26, 2003.

[26] A. De Martino, C. Martelli, and F. A. Massucci, "On the role of conserved moieties in shaping the robustness and production capabilities of reaction networks," *Europhysics Letters*, vol. 85, no. 3, Article ID 38007, 2009.

[27] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM Journal on Computing*, vol. 1, no. 2, pp. 146–160, 1972.

[28] W. Krauth and M. Mezard, "Learning algorithms with optimal stability in neural networks," *Journal of Physics*, vol. 20, no. 11, pp. L745–L752, 1987.

[29] K. B. Storey, Ed., *Functional Metabolism*, Wiley-Liss, Nashville, Tenn, USA, 2004.

[30] S. Durmuş Tekir, T. Çakir, and K. Ö. Ülgen, "Analysis of enzymopathies in the human red blood cells by constraint-based stoichiometric modeling approaches," *Computational Biology and Chemistry*, vol. 30, no. 5, pp. 327–338, 2006.

[31] G. Jacobasch, "Biochemical and genetic basis of red cell enzyme deficiencies," *Baillière's Clinical Haematologyy*, vol. 13, no. 1, pp. 1–20, 2000.

[32] D. Vitkup, P. Kharchenko, and A. Wagner, "Influence of metabolic network structure and function on enzyme evolution," *Genome Biology*, vol. 7, no. 5, article R39, 2006.

[33] R. P. Alexander, P. M. Kim, T. Emonet, and M. B. Gerstein, "Understanding modularity in molecular networks requires dynamics," *Science signaling*, vol. 2, no. 81, p. pe44, 2009.

[34] V. Van Kerrebroeck and E. Marinari, "Ranking vertices or edges of a network by loops: a new approach," *Physical Review Letters*, vol. 101, no. 9, Article ID 098701, 2008.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 642420, 14 pages
doi:10.1155/2010/642420

Review Article

A Metabolic Model of Human Erythrocytes: Practical Application of the E-Cell Simulation Environment

Ayako Yachie-Kinoshita,[1,2] Taiko Nishino,[1,3] Hanae Shimo,[1,4] Makoto Suematsu,[1,2] and Masaru Tomita[1,3,4]

[1] *Institute for Advanced Biosciences, Keio University, 403-1, Daihoji, Tsuruoka, Yamagata 997-0017, Japan*
[2] *Biochemistry and Integrative Medical Biology, School of Medicine, Keio University, Shinanomachi 35, Shinjuku, Tokyo 160-8582, Japan*
[3] *System Biology Program, Graduate School of Media and Governance, Keio University, Endo 5322, Fujisawa, Kanagawa 252-8520, Japan*
[4] *Faculty of Environment and Information Studies, Keio University, Endo 5322, Fujisawa, Kanagawa 252-8520, Japan*

Correspondence should be addressed to Ayako Yachie-Kinoshita, ayakosan@sfc.keio.ac.jp

Received 1 January 2010; Accepted 19 May 2010

Academic Editor: Maciek Antoniewicz

The human red blood cell (RBC) has long been used for modeling of complex biological networks, for elucidation of a wide variety of dynamic phenomena, and for understanding the fundamental topology of metabolic pathways. Here, we introduce our recent work on an RBC metabolic model using the E-Cell Simulation Environment. The model is sufficiently detailed to predict the temporal hypoxic response of each metabolite and, at the same time, successfully integrates modulation of metabolism and of the oxygen transporting capacity of hemoglobin. The model includes the mechanisms of RBC maintenance as a single cell system and the functioning of RBCs as components of a higher order system. Modeling of RBC metabolism is now approaching a fully mature stage of realistic predictions at the molecular level and will be useful for predicting conditions in biotechnological applications such as long-term cold storage of RBCs.

1. Introduction

Systems level behaviors occur as a consequence of synergistic interactions between individual components that can function as single systems, but can also affect the dynamic state of other components. To understand how components interact in biological systems, which exhibit nonlinear behaviors, not only the network structure of the system, but also all relevant components of the system need to be integrated in a quantitative manner. Systems biology has emerged to address these problems, using a combination of computational and experimental analyses [1, 2].

Kinetic and dynamic modeling is one of the most useful approaches in systems biology [3]. The E-Cell project was launched in 1996, early in the systems biology and -omics technology era, with the aim of developing whole-cell-scale mathematical models [4, 5]. Although a number of bio-simulation-specific platforms have been released recently [6], the approach of E-Cell in a fully object-oriented fashion is unique and allows for multitimescale/multialgorithm simulations [7, 8].

When a dynamic mathematical model is obtained, multiple analyses can be conducted to elucidate the fundamental design principles of a biological system [9, 10]. Such models enable researchers to examine experimentally intractable systems; for example, maintenance of homeostasis in organisms *in vivo* can be mimicked *in silico*, with large approximations, but not *in vitro*. Additionally, experimentally costly analyses such as comprehensive sensitivity analysis of an enzyme activity in response to perturbation of all components of the systems' networks can be modeled. Furthermore, a successful model that can represent the dynamic behavior of

the target system in the intended environment can be used to predict how the system responds to an external physiological stimulus or to modification of each component in the system, which would be extremely useful for development of biotechnological applications.

The metabolic network in human red blood cells (RBCs) has been the subject of mathematical modeling for the past three to four decades. As discussed below, a number of detailed models of RBC metabolism have been developed with various levels of abstraction, and many studies analyzing these models have produced a deeper understanding of the regulatory properties of metabolic pathways and provided novel insights for implementation of both mathematical analysis and metabolic engineering strategies, which can then be applied to many other cell types.

Our model of RBC metabolism using the E-Cell Simulation Environment focuses on the interrelationship between metabolism and the functional aspects of RBC physiology, including factors affecting allosteric transitions of hemoglobin and recent findings regarding the assembly of intracellular proteins at the plasma membrane. Our model successfully reproduced the temporal response to hypoxia, previously measured by metabolome analysis [11]. The model was then used to predict the metabolic status of RBCs in long-term cold storage, with the goal of optimizing the storage conditions [12].

2. The E-Cell System as a Cell Simulation Environment

Modeling of biological systems requires suitable abstraction of the system considering the amount, size, and speed of reactions, as well as many other factors. This process includes making decisions as to whether continuous processes should be broken down into discrete steps or, if treated as continuous processes, using deterministic or stochastic rules for modeling the process. The E-Cell System provides the simulation platform for use of these calculations to model separately, and/or in combination, multialgorithm/multitimescale simulations [7].

E-Cell models have three fundamental object classes: "variable", with the option of either molar concentration units or value units, "process", for writing operations on the variables, and "system" for identifying logical and/or physical compartments with/without volume that contain the variables and processes. This object-oriented approach enables intuitive description and prevents mistakes in modeling because of the one-to-one correspondence between each chemical process and reaction process in the E-Cell model. At the same time, once the reaction-module ("process" in the E-Cell System) has been created and defined, the module is easily reusable not only in the same model but also in another models. Similarly, the calculation algorithm itself can be modified or extended as a module and switched easily by rewriting one line of the model file, for example, from a deterministic to a stochastic model.

The system is written in C^{++} to maximize calculation speed and the frontends are easily scriptable and extensible using the Python language. Using Python scripts, users can program the rules for simulation sessions as well as the simulation conditions themselves in a given session; for example, initial parameter settings, time-dependent or concentration-dependent perturbations, or the output form of the simulation results. Some widely used methods of ODE model-based mathematical analysis have been already provided in the E-Cell Simulation Environment. These include sensitivity analysis, bifurcation analysis, Metabolic Control Analysis (MCA), and real genetic algorithms for parameter optimization.

In the case of large-scale pathways such as whole RBC metabolism, object-oriented modeling is necessary to ensure accuracy. In addition, its extensibility is also very helpful for parametric tuning of the model.

However, due to the lack of a user-friendly Graphical User Interface (GUI) for the display of modeling, *in silico* experimentation, and simulation results, the E-Cell System has been difficult to use, especially for biologists. Recently, the E-Cell IDE (Integrated Development Environment), a GUI -based simulation toolkit for Windows, was developed in order to allow non-expert users to edit, run, and analyze the E-Cell model (Figure 1). The pathway editor allows users to generate new pathways or customize existing pathway maps, and to set values/concentrations, reactions, and parameters directly using the GUI toolkit. This system can also read and write models in SBML (Systems Biology Markup Language) format [13], which is the most common markup language for making the models compatible with other cell simulators, for example, Copasi [14], DBSolve [15], Virtual Cell [16], Systems Biology Workbench (SBW [17]), and XPPAUT (http://www.math.pitt.edu/~bard/xpp/xpp.html). The E-Cell IDE provides a visual representation of the dynamics of the simulation on the pathway map by varying the size and width of each node or edge, respectively. The GUI tools enable the user to conduct classical mathematical analyses, such as parameter estimation, using a real number genetic algorithm and metabolic control analysis. Users can also perform CUI (Command-line User Interface)-based simulations and analyses for more complex or large-scale operations, such as situations in which the automation of many simulations or parametric computing is needed.

Development of the next-generation E-Cell System (E-Cell Version 4) has started by aiming to simulate biological events at the molecular level, so that we can assess "cellular space" issues such as molecular fluctuation, localization, and crowding, while keeping complete compatibility and combinability with the current version (E-Cell Version 3). A detailed description, the current status, and the future vision of the E-Cell Simulation Environment are provided in recent works [7] and the project web page [8].

3. Human RBC Metabolism: A Long History of Dynamic Simulation

Human RBCs lose their mitochondria and are entirely dependent on glycolysis to produce ATP. The glucose

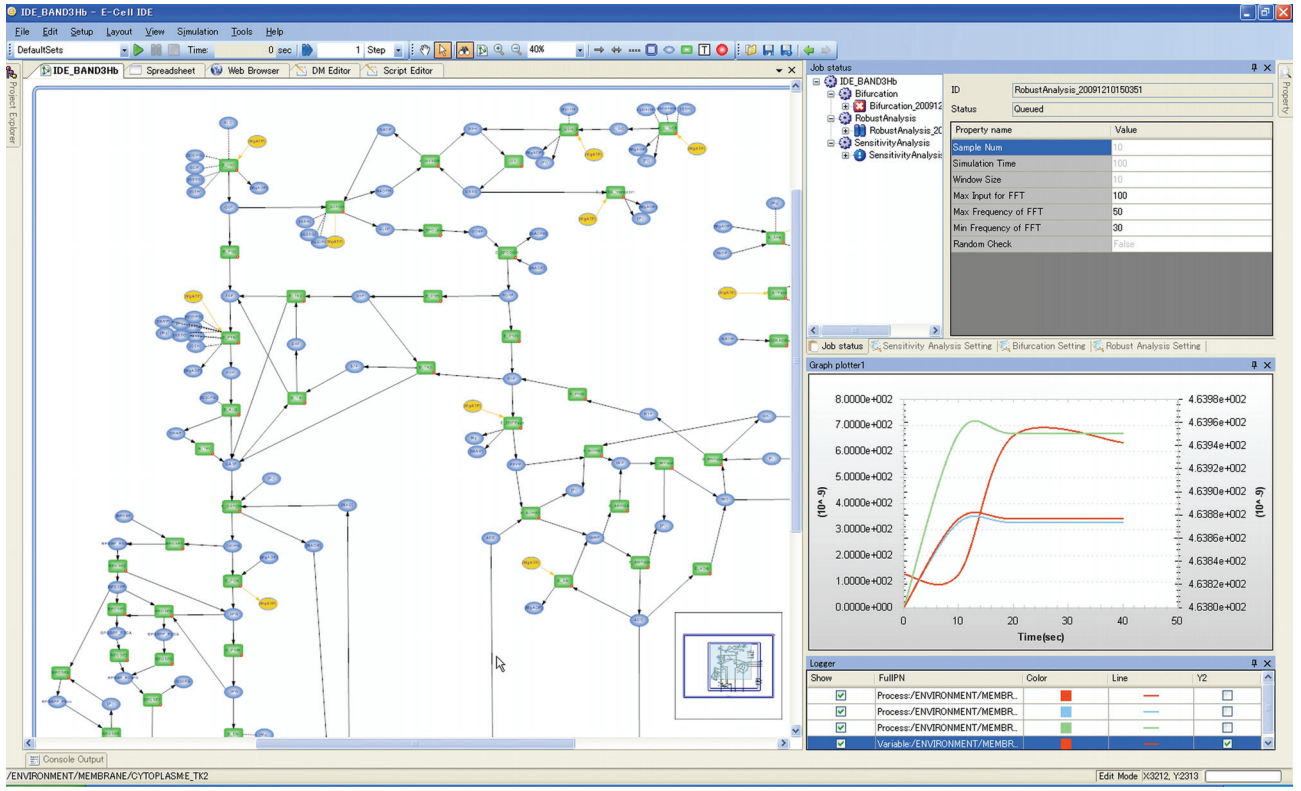


Figure 1: View of the on-screen E-Cell GUI (E-Cell IDE). The E-Cell IDE has a user-friendly pathway editor (left) and a GUI-based mathematical analysis tool kit (right top). The simulated dynamics is visualized in the tracer of the time histories of variables/processes (bottom right), as well as in the pathway map (left).

transporter GLUT-1, expressed in RBCs, has a high affinity for extracellular glucose and remains saturated under physiological concentrations of glucose. Hexokinase (HK), which catalyzes the initial step of glycolysis in RBCs, also has a low Km (high affinity) for intracellular glucose. Thus, the rate of initiation of glycolysis is ensured even when plasma glucose concentration is low. ATP is required by ion pumps to prevent cell swelling and is also used in many other endergonic enzymatic processes. ATP is a necessary substrate for two of the initial rate-limiting steps in glycolysis (HK, PFK), but excess ATP can downregulate glycolysis through the inhibition of PFK. Most of the glucose entering a cell is converted to lactate through glycolysis (in our model, 87.5% under normoxic steady-state conditions), with the remainder entering pentose-phosphate pathways (in our model, 12.5%). This route provides reduced nicotinamide adenine dinucleotide phosphate (NADPH), which prevents oxidation of the cell directly and indirectly through the conversion of glutathione from its oxidative form back to its active reduced form.

One of the most characteristic pathways in RBCs is the Rapoport–Luebering cycle, which generates a high concentration of 2,3-diphosphoglycerate(2,3-BPG), a diversion of glycolysis from 1,3-disphosphoglycerate(1,3-BPG) that prevents excess ATP production by bypassing the process catalyzed by phosphoglycerate kinase (PGK). The increase in 2,3-BPG facilitates the release of oxygen from hemoglobin to tissues. In this manner, although the binding and release of gas molecules (oxygen, carbon dioxide, and so on) through hemoglobin requires no energy, the regulation of glycolytic flux, which is important for maintaining oxygen transport capacity through the maintenance of adequate levels of ATP and 2,3-BPG.

A single RBC can be assumed to be a closed system enclosed by a plasma membrane. Human RBCs circulate for as long as 120 days [18], and, thus, the metabolism in the cell should be robust in physiological situations. This simplicity and robustness, as well as the abundance of accumulated knowledge regarding metabolic enzymes, have made RBC metabolism a suitable subject for mathematical modeling and system level analysis of the metabolic/biological pathways. There is a long history of construction of metabolic models of human RBCs, in which RBC metabolism is described with simultaneous ordinary differential equations with different levels of detail, depending on the focus of the model. These models have very different system level properties (for comparative analysis of several RBC glycolysis models, see [19]).

The first challenge in modeling human RBC metabolism was constructing a linear glycolytic model (by Rapoport and Heinrich [20]), which was intended to test whether a linear theory suffices for a description of the steady state under several experimental conditions, and to better understand the crossover structure. The model also contributed to the

discovery of the framework in "flux control theorem" [21]. Ataullakhanov et al. expanded the glycolytic model to include the pentose phosphate pathway, in which ATP and ADP are treated as parameters, to predict the dependence of glycolytic flux on ATP content [22].

An extension of the glycolytic model to include the reactions of synthesis and degradation of adenine nucleotides was provided by Schauer et al., and suggested that adenylate metabolism, the functional role of which remained poorly understood at the time, may serve to improve the stabilization of the energy charge [23].

Brumen and Heinrich were the first to attempt combining the models of energy metabolism with the models of volume regulation [24], and presented a metabolic osmotic model of RBCs [25]. Reactions synthesizing the adenylate pool, osmoregulation, electroneutrality, and ion transport were later incorporated and proved that changes in RBC volume are associated with glycolysis.

Holzhutter et al. used their mathematical model, which included glycolysis, the pentose phosphate pathway, and the 2,3-BPG shunt, for analysis of the pathology associated with pyruvate kinase deficiency [26]. They showed the effects of severely lowering the activities of several enzymes in the model and compared their results with experimental data from patients with deficiencies in these enzymes.

Joshi and Palsson constructed a model to provide a framework for the integration and interpretation of the extensive biochemical data on enzymes and metabolites by means of a comprehensive mathematical model of RBC metabolism [27, 28]. The model examines three different interdependent characteristics of RBC: the properties of the red cell membrane, the kinetics of transmembrane fluxes of chemical constituents of RBCs and plasma, and the thermodynamic formulation of the osmotic states. The model was extended to include the pH dependence of enzyme activities and mechanisms of volume regulation, namely electroneutrality and osmotic balances, accounting for the cell's interaction with the environment as well as cell metabolism [29]. They used their comprehensive model to propose emerging mathematical frameworks for metabolic analysis such as top-down analysis for revealing metabolic pools [30–32].

Mulquiney and Kuchel developed a precise model for the Rapoport–Lubering (2,3-BPG) shunt, as well as glycolysis and the pentose phosphate pathway, based on enzyme kinetics derived from their NMR assays [33, 34]. Their model also includes a detailed description of magnesium equilibrium and binding of metabolites to oxyhemoglobin (oxyHb) [35].

The first E-Cell RBC model that was published presented an analysis of the pathology associated with hereditary G6PD deficiency [36]. This model was subsequently expanded by incorporating the Joshi and Palsson model and also by the inclusion of the GSH synthesis pathway and GSSG transport system, and suggested that normally inactive pathways may have an essential role in abnormal conditions such as enzyme deficiencies.

Using E-Cell, we also constructed a model of the two metHb-reduction pathways in RBCs by expanding the Mulquiney and Kuchel model [37]. The model assessed the mechanism that switches between NADPH dependent and NADH dependent pathways of metHb reduction. The former pathway has a high response rate to hemoglobin oxidation with a limited reducing flux, and the latter has a low response rate with a high-capacity flux, correlated to the supply of NADH and NADPH from central energy metabolism.

Recently, we published a model that contains not only enzymatic or ion binding reactions combined with existing models, but also includes allosteric transitions in hemoglobin in response to the partial pressure of oxygen and the binding of plasma membrane proteins to glycolytic enzymes and hemoglobin ([11], described below).

Based on the same assumptions, Hald et al. focused on the alteration of RBC metabolism caused by changes in oxygen and carbon dioxide partial pressures during circulation, and showed that changes in these gases resulted in glycolytic flux oscillations, with consequent overshoots in levels of central metabolites [38].

4. Metabolic Model of Human RBCs Using the E-Cell System

4.1. Model Construction

4.1.1. Metabolic Reactions. The metabolic model constructed with the E-Cell system is shown schematically in Figure 2, which includes enzyme reactions, transporter functions, binding interactions, and a process for determining hemoglobin allosteric transitions. For detailed descriptions of all equations, see the supporting material from our previous reports [11, 12].

The metabolic network developed in this mathematical model covers a majority of the metabolic pathways: glycolysis, the pentose monophosphate shunt, the purine salvage pathway, glutathione metabolism, and Na^+-K^+-ATPase activity coupled to the leak of Na^+ and K^+. Membrane transporters for pyruvate, lactate, adenine, adenosine, hypoxanthine, and inosine are modeled, and the concentrations of all the extracellular metabolites, as well as intracellular glucose, are fixed at physiological concentrations. The rate equations for metabolic reactions are derived from previously published experimental data, and are largely based on the Mulquiney model [23, 33] and, in part, on other previous mathematical models [27, 36]. Most enzyme rate equations were derived using the King-Altman method for each of the enzyme-catalyzed reactions, with the exception of some reactions in purine metabolism and purine transport processes, which were described using Michaelis-Menten or first-order kinetics. Following the Mulquiney model, the following reactions are modeled as pH-dependent kinetics where intracellular pH is treated as an independent parameter: key enzymatic reactions in glycolysis, such as hexokinase (HK), phosphofructokinase (PFK), glyceraldehyde phosphate dehydrogenase (GAPDH), pyruvate kinase (PK), lactate dehydrogenase (LDH), the 2,3-BPG shunt reactions (2,3-BPG shunt), and binding of metabolites to Mg^{2+} and

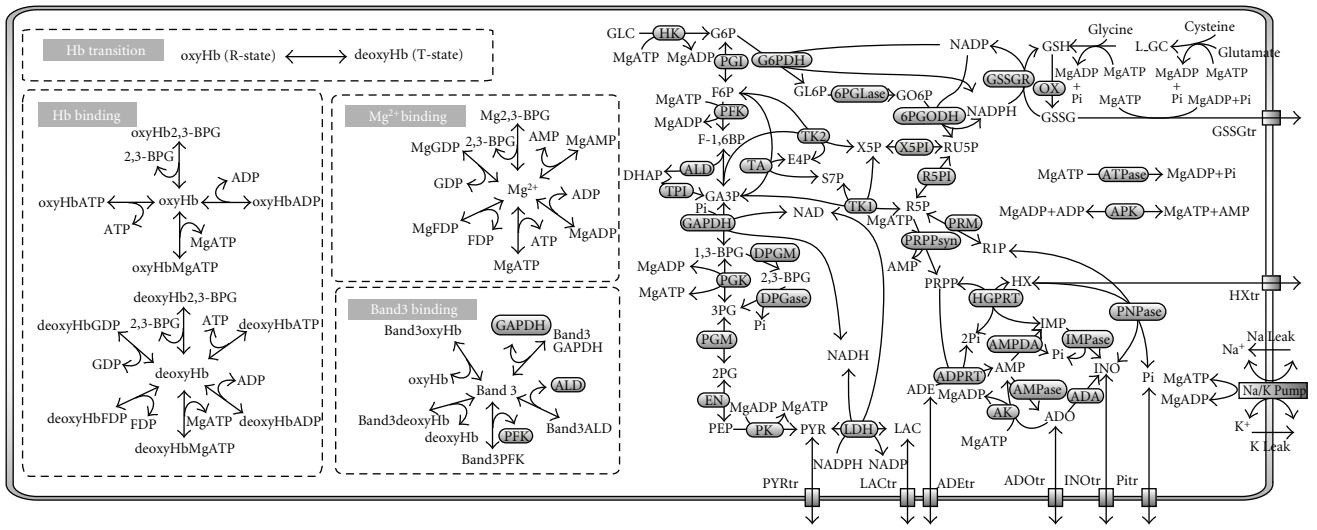


Figure 2: A Metabolic pathway map constructed using the E-Cell RBC model. The complete reaction network of the metabolism in the E-Cell RBC model. MgX/oxyHbX/deoxyHbX/band3X denotes the complex formed by compound "X" and Mg^{2+}, oxyhemoglobin, deoxyhemoglobin, or band 3, respectively. Abbreviations used in this figure: GLC, glucose; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; F1,6-BP, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GA3P, glyceraldehyde-3-phosphate; 1,3-BPG, 1,3-bisphosphoglycerate; 2,3-BPG, 2,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; LAC, lactate; GL6P, gluconolactone-6-phosphate; GO6P, gluconate-6-phosphate; RU5P, ribulose-5-phosphate; X5P, xylulose-5-phosphate; E4P, erythrose-4-phosphate; S7P, sedoheptulose-7-phosphate; R5P, ribose-5-phosphate; PRPP, 5-phosphoribosyl-1-phosphate; ADE, adenine; IMP, inosine monophosphate; R1P, ribose-1-phosphate; INO, inosine; ADO, adenosine; HX, hypoxanthine; AMP, adenosine monophosphate; ADP, adenosine diphosphate; ATP, adenosine triphosphate; NADP, nicotinamide adenine dinucleotide phosphate; NADPH, nicotinamide adenine dinucleotide phosphate (reduced); NAD, nicotinamide adenine dinucleotide; NADH, nicotinamide adenine dinucleotide (reduced); Ki, potassium ion; Nai, Sodium ion; Pi, inorganic phosphate; L_GC, L-glutamyl cysteine; GSH, glutathione (reduced); GSSG, glutathione (oxidized).

hemoglobin. The overall oxidization of glutathione in the cell and dephosphorylation of ATP except for Na^+/K^+-ATPase and kinases in above-mentioned major pathways are simplified into first-order reactions.

4.1.2. Hemoglobin Transition. The capacity of RBCs to deliver oxygen to tissues is fundamentally linked to the equilibrium between the two main states of hemoglobin structure, a tense state, "T" (or "deoxy"), and a relaxed state, "R" (or "oxy"), although the transition between the structures is not a simple two-state mechanism. The ability of oxygen to bind to hemoglobin is allosterically regulated by intracellular components such as 2,3-BPG and ATP. An increase in 2,3-BPG stabilizes the T-state of Hb, thereby facilitating O_2 dissociation from the Hb. Since T-state hemoglobin has a higher affinity for 2,3-BPG and ATP than the R-state, stabilization of hemoglobin in the T-state by increasing the amount of 2,3-BPG binding reduces the amount of free 2,3-BPG and ATP, leading to a further acceleration of metabolism. Under these circumstances, consideration of the hemoglobin transition may exert considerable effects on RBC metabolism, especially on glycolysis, and vice versa.

The equation for the kinetics of oxygen saturation of hemoglobin, derived by Dash and Bassingthwaighte [39], was used to calculate the ratio of oxy-(R-) hemoglobin to total hemoglobin. The saturation kinetics is described by the Hill equation, with dependencies on the levels of pO_2, pCO_2, intracellular pH, concentration of 2,3-BPG, and temperature. In our model, pO_2, pCO_2, pH, and temperature are independent variables. Once the ratio between oxy-(R-) and deoxy-(T-) hemoglobin (oxygen saturation of hemoglobin, S_{HbO2}) is determined under given conditions at a certain step in the simulation, the velocity of a reversible conversion of free T-state hemoglobin to free R-state hemoglobin in the step is derived as

$$\text{velocity}_{(T\to R)} = k \cdot \left(T_{\text{total}} \cdot \frac{S_{\text{HbO2}}}{1 - S_{\text{HbO2}}} - R_{\text{total}} \right), \quad (1)$$

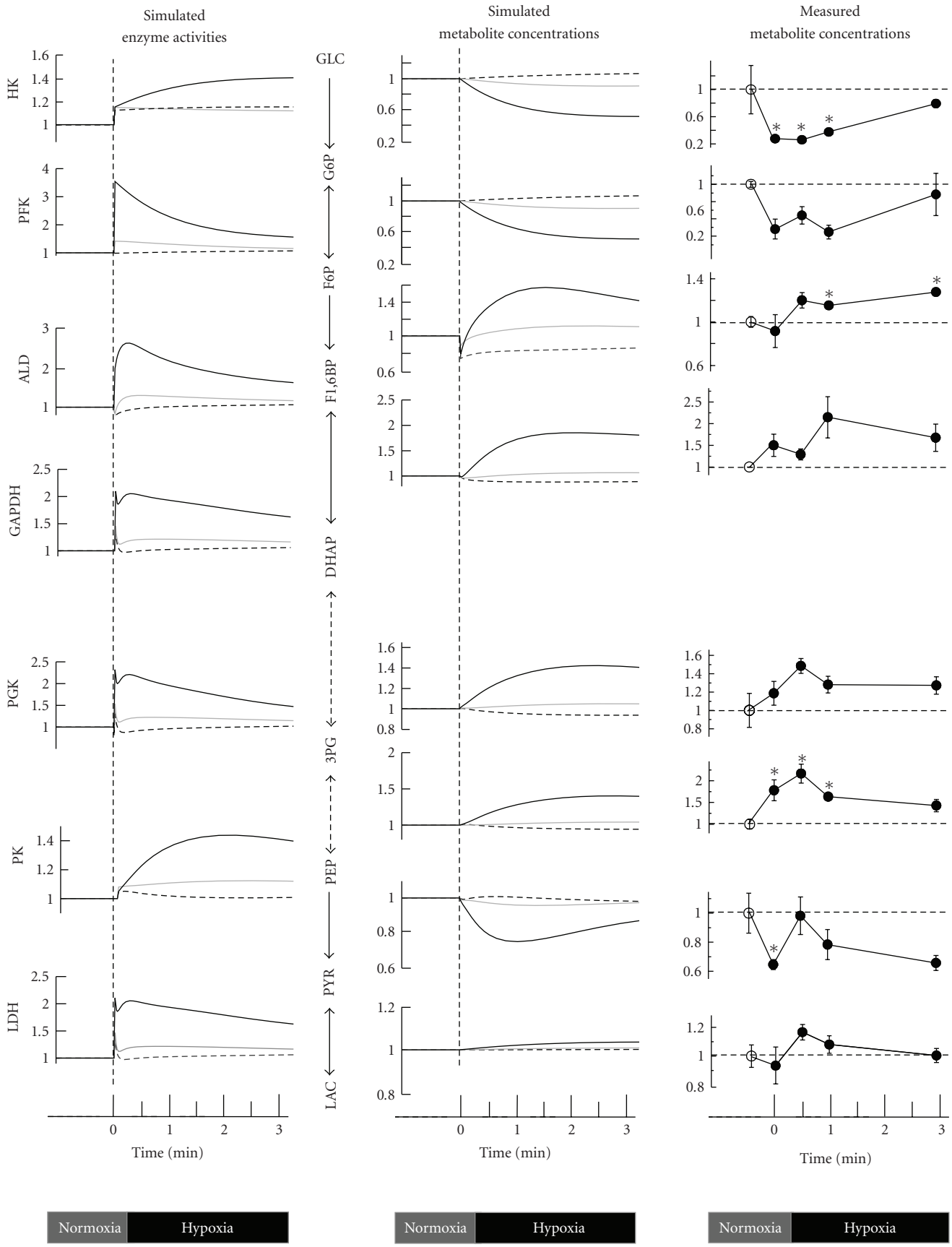
where S_{HbO2} ranges from 0 to 1, T_{total} and R_{total} are the sum of all complexed forms and the free form of each state of hemoglobin, and k denotes the scaling constant, which is set to 1200 to be equal to those in other binding reactions. In our model, it is assumed that this transition occurs only between free form R/T-hemoglobins, even though T_{total} and R_{total} in the equation above are the "total hemoglobin" of each state. In our model, the P_{50} value, which is the partial pressure of oxygen at half-saturation of hemoglobin, is calculated to be 25-26 mmHg under virtual cell-free conditions with 5 mM 2,3-BPG. However, if the concentrations of the free forms of T-state and R-state hemoglobin are substituted for T_{total} and R_{total}, the P_{50} is calculated to be approximately 50 mmHg, which is significantly different from the physiological value.

4.1.3. Binding between Band 3 and Glycolytic Enzymes or Hemoglobin. Band 3 is the major anion transporter in RBCs and plays a role in chloride/bicarbonate exchange, as well as an important structural role as a plasma membrane protein that contributes to stabilization of the membrane surface by forming multiprotein complexes [40]. Band 3 has multiple cytoplasmic domains, including an N-terminal cytoplasmic domain, which binds to the glycolytic enzymes PFK, ALD and GAPDH. PK and LDH are also localized to the plasma membrane when intact RBCs were fixed in their oxygenated states, but there has been no evidence of direct association with band 3 [41]. The N-terminal cytoplasmic domain of band 3 also binds to Hb, with a greater affinity for T-state hemoglobin than R-state [42]. Recent observations have shown the enzymatic activities of PFK, ALD, and GAPDH are completely blocked by binding to band 3, whereas their activities are recovered upon dissociation from band 3. No changes in PK and LDH activities have been detected [43] and the membrane docking sites of PK and LDH have yet to be identified [44]. We were the first to incorporate the binding of band 3 to glycolytic enzymes/hemoglobins into the mathematical model of major RBC metabolism [11]. The binding affinity of the metabolite-hemoglobin complex to band 3 is assumed to be the same as that of the free form of hemoglobin (e.g., deoxyHb and deoxyHb-2,3BPG). We simply used the published association constants evaluated *in vitro*.

4.2. Application of the Model I: Analysis of the RBC Response to Hypoxia. ATP is required in the initial steps of glycolysis by HK and PFK to trigger glycolytic ATP synthesis, and a fraction of the intracellular ATP is released to the extracellular space under hypoxia to elicit hypoxia-induced vasodilation, although the amount of ATP released is small, compared to intracellular levels [45]. At the same time, RBCs are known to accelerate glucose consumption in response to exposure to hypoxia, which results from acceleration of glycolysis, as judged by the increase in 2,3-BPG [46]. This change leads to further T-state Hb stabilization and increases the supply of oxygen to hypoxic tissues. Taking into account these features, researchers have hypothesized that RBCs have appropriate mechanisms for responding quickly to hypoxia to upregulate *de novo* ATP synthesis and glycolytic flux, leading to the increase in 2,3-BPG. Historically, the hypoxic acceleration of glycolysis in RBCs was thought to be induced by alterations in pH or the metabolic compensation of ATP [47, 48]. However, recent evidence from studies with intact RBCs has shown that T-state hemoglobin triggers an increase in the activities of glycolytic enzymes that interact with band 3 and plays a central role in acceleration of glucose consumption to increase the synthesis of ATP and 2,3-BPG [49]. Our model was used to elucidate the mechanistic features of the coordination and dynamics of sequential glycolytic reactions and the outcomes, in terms of alterations in levels of intracellular metabolites upon hypoxia, in particular as a result of band 3 interactions. To mimic hypoxic conditions, the pO_2 of the model, which was initially set to 100 mmHg, was reduced to 30 mmHg, a value consistent with capillary microvessels *in vivo* [50].

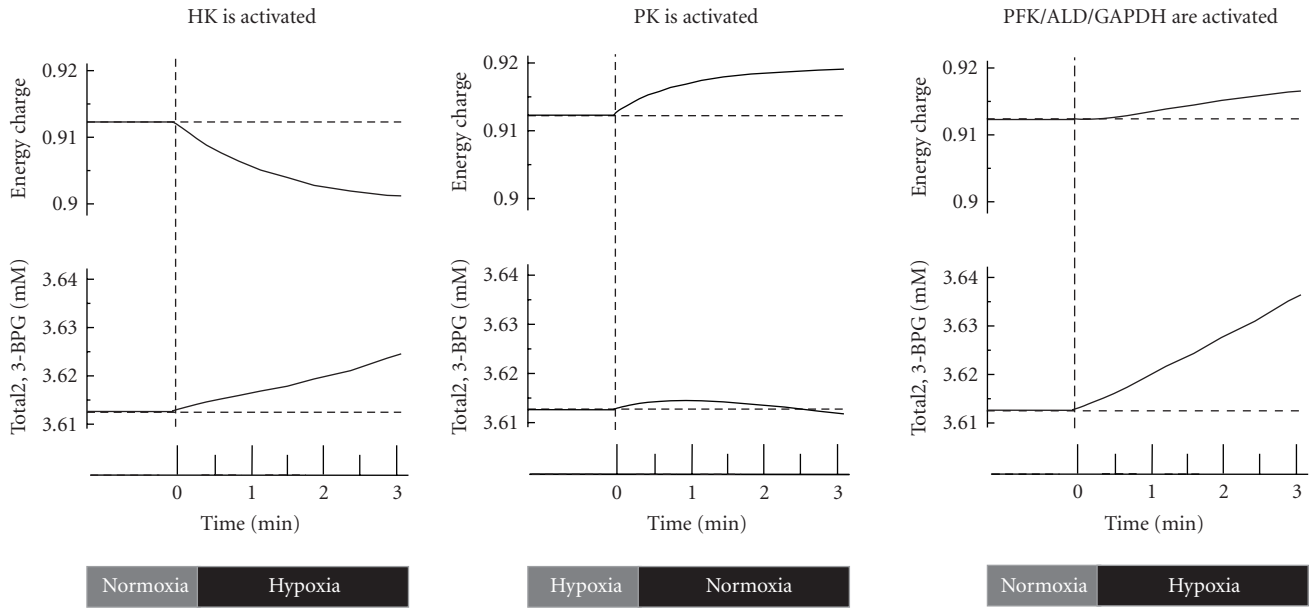
A comparison between three models in predicting temporal alterations of glycolysis during a 3-min virtual hypoxia is illustrated in Figure 3(a). Model A includes Band 3 interactions with hemoglobin and glycolytic enzymes (corresponding to the BIII(+) model described in [11]). Model B uses the same initial and normoxic steady-state conditions as model A, but omits interactions between Band 3 and hemoglobin/enzymes (corresponding to the BIII(−) model in the reference [11]). By comparing model A with model B, we can estimate the pure effect of hypoxia-induced glycolytic activation exerted by the accumulation of T-state hemoglobin. In Model C, all of the glycolytic enzymes exist in dissociated forms, even in the initial (PO2 = 100) state. Thus, the initial and steady-state conditions of model C are different from those of model A and model B. As shown in Figure 3(a), the overall activities of the glycolytic enzymes in model A are significantly accelerated relative to the other models. As expected, the activities of PFK, ALD, and GAPDH spiked immediately as a result of their release from band 3 upon hemoglobin binding, while these enzyme activities did not change significantly in model B. These differences in enzyme activities between the two models created a distinct metabolome profile: the glycolytic intermediates in model A displayed a pattern opposite to those in model B. In model A, G6P and F6P decreased by 50% and F1,6BP, DHAP, 3PG, and PEP increased by 40% versus the corresponding baseline levels (Figure 3(a)). The time-dependent alteration in levels of glycolytic intermediates predicted by model A is entirely consistent with results obtained from metabolome analyses using capillary electrophoresis mass spectrometry (CE-MS). These trends in time-variation in glycolytic fluxes/metabolite concentrations have also been reproduced in model C, where no band 3 interactions are considered. However, the alterations seen in model C exhibit much less variation than the experimental results. In both models (model A and model C), the activation of HK is caused by a decrease in the free form of 2,3-BPG, which is a strong inhibitor of HK; however, HK activity in model A exhibited greater activation. This difference appeared to result from a decrease in G6P and an increase in ATP, leading to a reduction in HK product inhibition. The distinct hypoxia-induced increase in glycolytic flux arising from hemoglobin transition and the consequent changes in band 3-interactions have been verified by several separate experiments. Messana et al. showed that the hypoxia-induced increase in glucose consumption disappeared in RBCs treated with 4,4′-diiso-thiocyanostilbene-2,2′-disulfonate (DIDS), which acts as an anion exchange inhibitor targeting band 3 [51]. We demonstrated that CO pretreatment of RBCs to stabilize hemoglobin in the R-state attenuated the hypoxia-induced acceleration in the conversion of ^{13}C-glucose into ^{13}C-lactate [11]. Furthermore, a recent study by Lewis et al. provided direct evidence for the role of band 3 in mediating metabolic shifts under more physiological conditions in intact RBCs, in which the metabolic fluxes were measured using (1)H-(13)C NMR and RBCs were treated with pervanadate, a reagent that blocks the interaction between band 3 and glycolytic enzymes [49].

Through the activation of glycolysis, the energy charge is greater in the band 3-implemented model than the



(a)

Figure 3: Continued.



(b)

Figure 3: Temporal alterations in glycolysis in response to hypoxia. (a) Predicted alterations in glycolytic enzyme activities under hypoxia and comparison of time-courses of glycolytic intermediates between the simulations and metabolome measurements using CE-MS. The simulation results of model A (including Band 3 interactions, corresponding to the solid red line in Figure 2 of [11]), model B (using the same initial steady-state conditions as A but neglecting further Band 3 interactions, corresponding to the dotted blue line in Figure 2 of [11]), and model C (with all enzymes dissociated from Band 3) are represented by the solid black (model A), dotted black (model B), and solid gray (model C) lines, respectively. See the main text for discussion of the models. In the metabolome data collected by CE-MS, closed circles indicate ratios of hypoxic metabolite concentrations to normoxic control concentrations, which are represented with open circles. Values are the means ± SE of four separate experiments. Asterisks: $P < .05$ versus the steady-state baseline values. (b) Results of model analysis to determine whether or not increasing the activity of a particular enzyme (HK, PK, or PFK+ALD+GAPDH) by 2-fold can promote a simultaneous increase in the energy charge and in 2,3-BPG under conditions of hypoxia.

other models. The basal energy charge under normoxic steady-state conditions was predicted to be 0.912, which is comparable to that reported in previous studies (ranging from 0.86 [52] to 0.935 [53]), and the value rose to 0.917 after three minutes in hypoxia. At the same time, the total amount of 2,3-BPG in model A was also predicted to be greater than the other two models, which is likely to contribute to a rapid increase in the formation of hemoglobin-2,3-BPG complexes that consequently lead to the release of residual hemoglobin-bound oxygen from the RBC. The calculated amount of oxygen released by increasing 2,3-BPG in three minutes of hypoxia was 6 μmol per L of RBC volume.

Another key finding of this simulation study is that PFK activation is a crucial step for the upregulation of both energy charge and 2,3-BPG generation, while activation of the other so-called rate-limiting enzymes, at the initial (e.g., HK) or final (e.g., PK) steps of the glycolytic pathway, fails to satisfy these requirements. HK activation resulted in a decrease in energy charge and an increase in 2,3-BPG, while PK activation increased energy charge without stimulating 2,3-BPG generation (Figure 3(b)).

Through these analyses, our model has demonstrated a pivotal role for band 3-intracellular protein interactions in enhancing activation of glycolysis in RBCs as part of a metabolic response to hypoxia, and in the consequent increase in both cell energetics and oxygen-carrying capacity. Taken together with the recent findings that the oxygenation-dependent assembly of glycolytic enzymes on the membrane is conserved in mammalian erythrocytes even in the absence of the intracellular band 3 binding site [44], the hypoxia induced glycolytic activation may be necessary for RBC viability over the long term and/or for temporally appropriate oxygen supply to the tissues.

4.3. Application of the Model II: Prediction of Metabolism of RBCs in Long-Term Storage. Whereas the kinetic metabolic models of human RBCs have led the simulation studies in the systems biology era, there has been no practical application of the RBC metabolic models for biotechnological use, at least to our knowledge. We intend to use our RBC model as a virtual experiment for optimization of RBC storage conditions. In the field of emergency medicine, it is critical to store RBCs in such a manner that their viability and capacity for oxygen delivery are retained after transfusion. These characteristics are strongly correlated with the intracellular metabolic status. Long-term cold storage reduced intracellular ATP and 2,3-BPG, which causes a reduction in deformability, oxygen-carrying capacity, and reduces energy sources for intracellular processes. In Japan, a mannitol-adenine-phosphate (MAP) solution is commonly used for storing

RBCs (RC-MAP). In this solution, approximately 50% of ATP is retained after 42 days in storage, but 2,3-BPG is almost completely depleted after 2 weeks [54]. However, the large metabolic network of the RBC that underlies the depletion of these two metabolic indicators has not been considered.

To develop an "RC-MAP model" that can represent the long-term dynamics of cold MAP-stored RBC metabolism, concentrations of glucose and adenine of the basal model were set to the values in MAP-solution, and the extracellular sodium ion and inorganic phosphate concentrations were changed by calculating the content of sodium phosphate, sodium chloride and sodium citrate in the solution. The intracellular pH of RBCs in RC-MAP is lowered due to the large addition of citrate and gradually decreases further over time in storage due to the accumulation of lactate [55]. Thus, the initial pH and pH variation during cold storage were set according to the reported time-course pH data in the literature [54], where the decrease in intracellular pH is fitted by first-order kinetics. As the cold temperature stabilizes hemoglobin in the R-state, all hemoglobin was designated R-state in the RC-MAP model.

In the process of parameter estimation, we focused on those enzymatic activities that may be altered by physical changes in the storage conditions relative to the circulating conditions, the cold temperature and the low pH. All chemical reactions, including enzymatic reactions, chemical-binding reactions, and active transport processes, should be significantly reduced but not completely stopped at 4°C. In particular, the Na^+/K^+-ATPase pump is very sensitive to lowered temperature [56]. Moreover, there are many reports that enzymes involved in purine metabolism, including adenosine deaminase, adenosine monophosphate phosphohydrolase (AMPase), inosine monophosphatase, and adenosine monophosphate deaminase are optimized or activated at relatively low pH, such as in the RC-MAP storage conditions, but adenosine kinase and hypoxanthine-guanine phosphoryl-transferase are not. From these knowledge, we grouped all reaction activities into three groups: Na^+/K^+ pump activity, purine metabolism enzyme activities, and all remaining enzymatic or binding activities. The three groups of reaction activities are then treated as "adjustable parameters" for the parameter estimation.

As a result of the GA (Genetic Algorithm) analysis, the three adjustable parameters were determined and their estimated values were supported by previous knowledge, as follows. (1) Na^+/K^+ pump activity was 0.6% of the basal values, since it decreases to 0.2–0.8% of the level at body temperature when the temperature is decreased to 5°C in most mammalian erythrocytes [56]. (2) The activity of purine metabolism enzymes was 24% of the basal model, which is considerably higher than that of other enzymatic processes. Furthermore, the loss of intracellular adenine within three weeks was reproduced only by the model when this parameter was 20–30% [12]. (3) The activity level of the other reactions was estimated to be 3.5%, since, in general, enzyme activities at 4°C are reduced to 1–5% of those at the normal body temperature [57]. The dynamics of the estimated RC-MAP model during cold storage was then compared with the experimentally measured glycolytic intermediates in MAP-stored RBCs using capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS), as previously described [12, 58] (Figure 4(b)).

In CE-TOFMS measurements, PYR, LAC, HX, and S7P were significantly increased, and all glycolytic intermediates, with the exception of PYR and LAC, were markedly decreased after 49 days. These measured alterations of intermediates were qualitatively reproduced in an estimated model that was fitted to the reported RC-MAP data. The models using random sets of adjustable parameters failed to predict these final increases and/or decreases in the concentrations of the glycolytic intermediates. However, the dynamics of the intermediates, such as the extraordinarily large increases in F1,6-BP and DHAP in the first week or the initial stagnation of PYR, could not be predicted by the estimated model. These gaps may be the result of both the level of simplicity of setting the adjustable parameters and the difference in experimental conditions in our lab relative to commercial RC-MAP usage. The early large peaks of F1,6-BP and DHAP did appear when the initial pH was set to lower values, but some mechanism of reduction in PK activity was needed to reproduce the initial stagnation of PYR (see the supporting material in [12]). More detailed refinement of the settings of the RC-MAP model using comprehensive metabolome data will be necessary to increase the predictive power of the model.

Based on the sensitivity analysis of the RC-MAP model, 2,3-BPG levels may be maintained when hemoglobin is shifted to the T-state. Reducing enzyme activities in purine metabolism would be very effective for maintaining ATP, whereas the ATP concentration control of enzymes in purine metabolism is small. Another interesting prediction of the model is that a slight activation of HK and PFK during storage can maintain both ATP and 2,3-BPG, despite the fact that all prior studies of blood storage have aimed to reduce enzymatic activities. These factors can serve as a possible target of the next round of experiments for improving RBC storage conditions.

However, many physicochemical aspects have yet to be incorporated in the model, such as the maintenance of cellular homeostasis through regulating intracellular pH and cell volume in connection with ion balance. In this study, the effects of additives in the MAP solution, such as mannitol and sodium chloride, as osmoregulatory substrates that prevent erythrocyte hemolysis, as well as the acidotic shift due to lactate accumulation, were ignored (or considered to be unchanged from normal circulating RBCs).

5. Discussion: Future Perspectives of RBC Metabolic Model

A mathematical model of the metabolic networks in human RBCs with precise enzyme kinetics and linkages between metabolism and oxygen pressure can provide us with a better understanding of the response to environmental stimuli that may occur *in vitro* or in storage conditions. Our model is the first to include the effects of intracellular protein-protein interactions (competitive binding to band 3 between

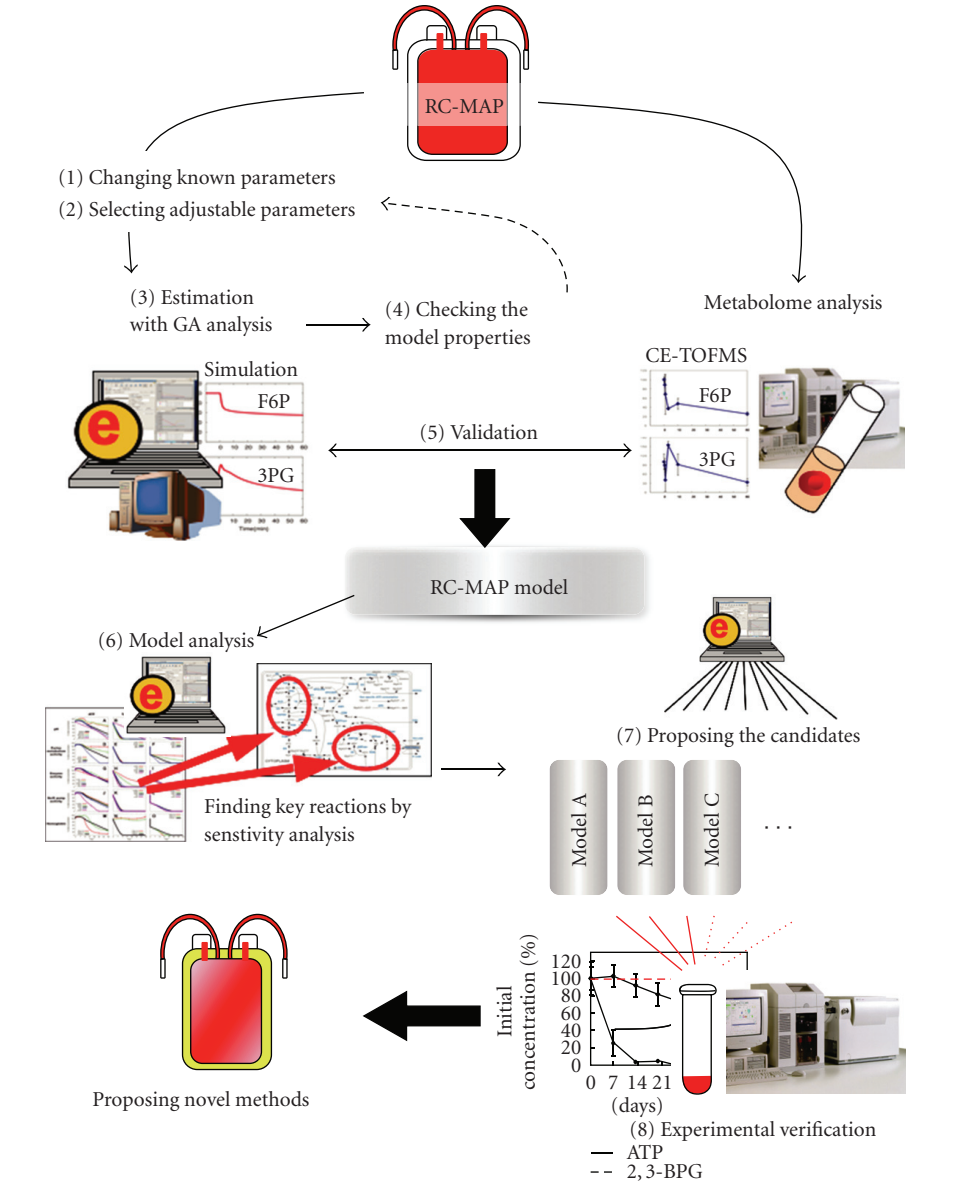


Figure 4: Continued.

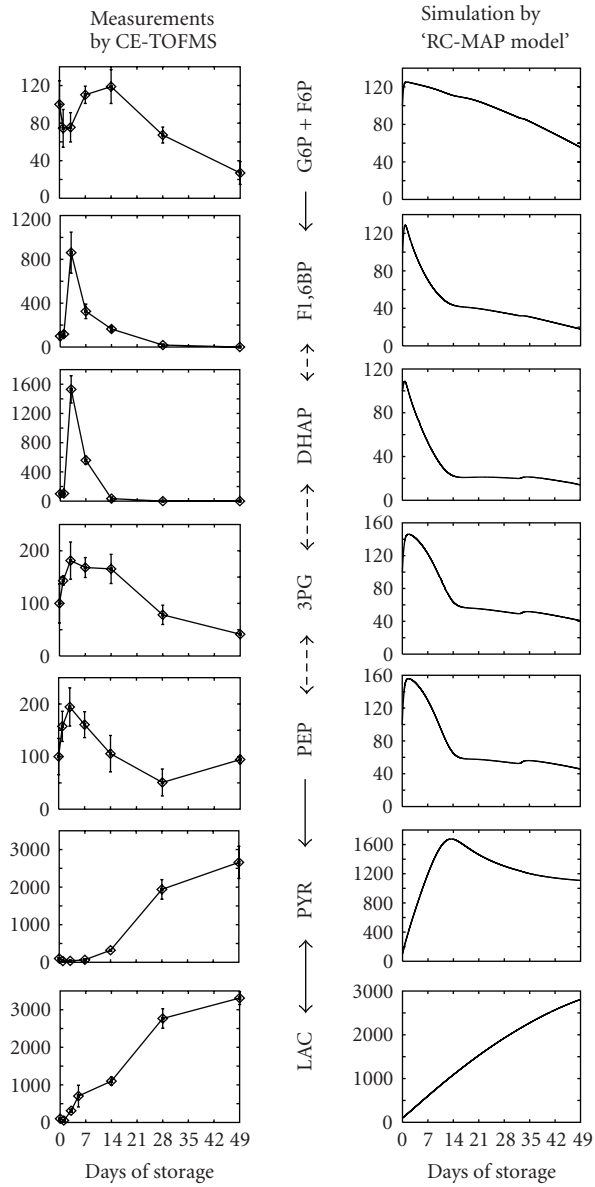


Figure 4: Modeling of metabolism in RBCs during long-term cold storage and model verification. (a) The work flow of developing and analyzing the "RC-MAP model". (1) The values of external parameters in our E-Cell RBC model (basal model), which assumes normal circulating conditions, are modified to meet the conditions of the cold-MAP solution. (2) Adjustable parameters, which are thought to be changed by storage conditions, are determined. (3) Adjustable parameters are estimated using experimental data on metabolic alterations in RC-MAP. (4) The relevance and robustness of the parameter choices are checked. (5) The dynamics of metabolism are compared between the RC-MAP simulation and experimental treatment. (6) A sensitivity analysis of the model is conducted to find key reactions, metabolites, and parameters to maintain the energetics and oxygen-carrying capacity of stored RBCs. (7) Candidate components for optimized storage conditions are determined by computationally testing various combinations of factors that can be modified experimentally. (8) Finally, the candidate models are validated by a metabolome analysis. In (3), we employed the classical read number genetic algorithm within the E-Cell Simulation Environment to fit the model to the reported time-course data of ATP and 2,3-BPG concentration changes in RBCs held in cold RC-MAP for 49 days. (b) A comparison of time-related changes in ATP and 2,3-BPG levels between reported experiments (Shiba et al., 1991) and the "RC-MAP model" is presented. The time-course of ATP (solid black) and 2,3-BPG (broken black) in RC-MAP at 4°C for 49 days in (a) previously reported data and (b) the prediction of the estimated model derived by the Genetic Algorithm. Experimental values are shown as the mean ± S.D. of 19 separate experiments. Values are percentages of the initial concentrations. (c) Measured (left) and simulated (right) time-dependent alterations in glycolytic intermediates. In CE-TOFMS measurements, the RBC samples were suspended in cold-MAP for 49 days under laboratory conditions. Data represent the means ± S.D. of five separate experiments. "G6P + F6P" indicates the sum of the concentrations of G6P and F6P. Values are represented as percentages of the initial concentration of each metabolite.

hemoglobin and glycolytic enzymes) on RBC metabolism and to show the relevance of these interactions to the supply of ATP and 2,3-BPG over time.

Several factors must be considered in order for the model to provide an insight into the possible physiological importance of these intracellular events. First, a precise estimation of the energy demand of the cell, such as the energy required to maintain its biconcave shape, should be made [59]. Among all previously published mathematical models of RBC metabolism, the processes utilizing ATP are oversimplified into first-order kinetics with respect to the ATP concentration, with the exception of Na^+-K^+ pump ATPase activity in Jasimidi and Palsson's model [27, 28], and in our models [11, 36], despite the fact that both the glycolytic flux and the concentration of ATP are largely controlled by the ATP demand [19].

Secondly, it is important to examine the external effects of RBC metabolism, that is, the modulation of oxygen carrying capacity by altering the hemoglobin allostery by intracellular ATP and 2,3-BPG. Including these features requires knowing "when" and "to what extent" the environmental oxygen demand is raised, in order to obtain information at a systems level of behavior. In other words, the model would need to make a connection between the efficiency of carrying gas molecules in the RBC and the underlying metabolism, as well as the behavior of the RBC population in response to capillary geometry and blood flow. Bassingthwaights' group has begun to construct a multiscale model in their physiome project: A model of blood-tissue/tissue-capillary exchange of oxygen, carbon dioxide, which includes exchange of bicarbonate and hydrogen ion for considering Bohr and Haldane effects in RBC, intraerythrocytic adsorption of CO_2 and O_2 on hemoglobin, and extraerythrocytic tissue-dependent gas consumptions [58, 60]. Furthermore, a novel aspect of RBC metabolism is the recently observed release of intracellular ATP into the extracellular space in response to hypoxia, although the actual amount of ATP released is very small compared to levels within the cell [45]. The released ATP regulates blood flow by binding to P_{2Y} purinergic receptors on the luminal surface of the endothelium, initiating the signaling events that result in vasodilation. A theoretical test of the contribution of the hypoxia-dependent ATP release by RBCs to an increase in vessel diameter in upstream arterioles has been carried out using a simplified theoretical blood-flow model [61]. An understanding of the physiological importance of temporal alterations in RBC metabolism could be accomplished by using these theoretical frameworks to develop models that connect the detailed metabolism in the RBC to higher level processes outside the cell.

Another intriguing addition to the model would be the incorporation of spatial or intracellular locus information into the metabolic model. Because RBCs do not contain any bound organelles in the cytoplasm, the intracellular system has been modeled as homogeneous in space. However, recent observations that glycolytic enzymes form a macromolecular complex [43] suggest the importance of considering spatial effects, even in models that focus on metabolic reactions. Additionally, another cytoplasmic domain of band 3 (C-terminal) binds carbonic anhydrase, which plays an important role in RBCs by catalyzing the hydration of carbon dioxide [62]. This interaction may improve the efficiency of Cl^-/HCO_3^- exchange and, at the same time, may change the local pH significantly. Since the interactions between intracellular proteins and band 3 are strongly dependent on pH and ionic conditions [63], the local pH may control glycolytic flux both directly and indirectly. The intracellular compartmentalization and the diffusion of the glycolytic enzyme complexes and hemoglobin need to be taken into consideration for precise estimation of the functional requirements of the gas (oxygen or carbon dioxide) dependent assembly of these macromolecules, especially for the juxtamembrane localization of PK and LDH (whose activities are not masked by binding to the plasma membrane).

Acknowledgments

The authors thank Kazunari Kaizu and Takeshi Sakurada for helpful comments. This paper was supported by a grant from CREST, the Japan Science and Technology Agency (JST); a grant from Yamagata prefecture, Japan; and a Grant-in-Aid for Young Scientists from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Research and Development of the Next-Generation Integrated Simulation of Living Matter, which is part of the Development and Use of the Next-Generation Supercomputer Project of MEXT. A. Y-Kinoshita and T. Nishino are supported by the "Global-COE Program for Human Metabolomic Systems Biology".

References

[1] H. Kitano, "Computational systems biology," *Nature*, vol. 420, no. 6912, pp. 206–210, 2002.

[2] B. Di Ventura, C. Lemerle, K. Michalodimitrakis, and L. Serrano, "From in vivo to in silico biology and back," *Nature*, vol. 443, no. 7111, pp. 527–533, 2006.

[3] R. Steuer, T. Gross, J. Selbig, and B. Blasius, "Structural kinetic modeling of metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 32, pp. 11868–11873, 2006.

[4] M. Tomita, K. Hashimoto, K. Takahashi et al., "E-CELL: software environment for whole-cell simulation," *Bioinformatics*, vol. 15, no. 1, pp. 72–84, 1999.

[5] M. Tomita, "Whole-cell simulation: a grand challenge of the 21st century," *Trends in Biotechnology*, vol. 19, no. 6, pp. 205–210, 2001.

[6] R. Alves, F. Antunes, and A. Salvador, "Tools for kinetic modeling of biochemical networks," *Nature Biotechnology*, vol. 24, no. 6, pp. 667–672, 2006.

[7] K. Takahashi, S. N. Vel Arjunan, and M. Tomita, "Space in systems biology of signaling pathways—towards intracellular molecular crowding in silico," *FEBS Letters*, vol. 579, no. 8, pp. 1783–1788, 2005.

[8] E-Cell Project, http://www.e-cell.org/.

[9] J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle III, and J. Doyle, "Robustness of cellular functions," *Cell*, vol. 118, no. 6, pp. 675–685, 2004.

[10] E. Kilpp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems Biology in Practice: Concepts, Implementation and Application*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2005.

[11] A. Kinoshita, K. Tsukada, T. Soga et al., "Roles of hemoglobin allostery in hypoxia-induced metabolic alterations in erythrocytes: simulation and its verification by metabolome analysis," *Journal of Biological Chemistry*, vol. 282, no. 14, pp. 10731–10741, 2007.

[12] T. Nishino, A. Yachie-Kinoshita, A. Hirayama, T. Soga, M. Suematsu, and M. Tomita, "In silico modeling and metabolome analysis of long-stored erythrocytes to improve blood storage methods," *Journal of Biotechnology*, vol. 144, no. 3, pp. 212–223, 2009.

[13] M. Hucka, A. Finney, H. M. Sauro et al., "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, no. 4, pp. 524–531, 2003.

[14] S. Hoops, R. Gauges, C. Lee et al., "COPASI: a complex pathway simulator," *Bioinformatics*, vol. 22, no. 24, pp. 3067–3074, 2006.

[15] O. D. Goryanin and F. Tobin, "Applications of whole cell and large pathway mathematical models in the pharmaceutical industry," in *Metabolic Engineering in the Post Genomic Era*, pp. 321–356, Horizon Bioscience, Oxford, UK, 2004.

[16] I. I. Moraru, J. C. Schaff, B. M. Slepchenko et al., "Virtual cell modelling and simulation software environment," *IET Systems Biology*, vol. 2, no. 5, pp. 352–362, 2008.

[17] H. M. Sauro, M. Hucka, A. Finney et al., "Next generation simulation tools: the systems biology workbench and BioSPICE integration," *OMICS*, vol. 7, no. 4, pp. 355–372, 2003.

[18] H. S. Hsieh and E. R. Jaffé, "Electrophoretic and functional variants of NADH-methemoglobin reductase in hereditary methemoglobinemia," *Journal of Clinical Investigation*, vol. 50, no. 1, pp. 196–202, 1971.

[19] F. B. du Preez, R. Conradie, G. P. Penkler, K. Holm, F. L. J. van Dooren, and J. L. Snoep, "A comparative analysis of kinetic models of erythrocyte glycolysis," *Journal of Theoretical Biology*, vol. 252, no. 3, pp. 488–496, 2008.

[20] T. A. Rapoport, R. Heinrich, G. Jacobasch, and S. Rapoport, "A linear steady state treatment of enzymatic chains: a mathematical model of glycolysis of human erythrocytes," *European Journal of Biochemistry*, vol. 42, no. 1, pp. 107–120, 1974.

[21] R. Heinrich and T. A. Rapoport, "Mathematical analysis of multienzyme systems. II. Steady state and transient control," *BioSystems*, vol. 7, no. 1, pp. 130–136, 1975.

[22] F. I. Ataullakhanov, V. M. Vitvitsky, A. M. Zhabotinsky, et al., "The regulation of glycolysis in human erythrocytes. The dependence of the glycolytic flux on the ATP concentration," *European Journal of Biochemistry*, vol. 115, no. 2, pp. 359–365, 1981.

[23] M. Schauer, R. Heinrich, and S. M. Rapoport, "Mathematical modelling of glycolysis and of the adenine nucleotide metabolism of human erythrocytes. II. Simulation of the adenine nucleotide breakdown after glucose depletion," *Acta Biologica et Medica Germanica*, vol. 40, no. 12, pp. 1683–1697, 1981.

[24] D. C. Tosteson and J. F. Hoffman, "Regulation of cell volume by active cation transport in high and low potassium sheep red cells," *The Journal of General Physiology*, vol. 44, pp. 169–194, 1960.

[25] M. Brumen and R. Heinrich, "A metabolic osmotic model of human erythrocytes," *BioSystems*, vol. 17, no. 2, pp. 155–169, 1984.

[26] H.-G. Holzhutter, G. Jacobasch, and A. Bisdorff, "Mathematical modelling of metabolic pathways affected by an enzyme deficiency. A mathematical model of glycolysis in normal and pyruvate-kinase-deficient red blood cells," *European Journal of Biochemistry*, vol. 149, no. 1, pp. 101–111, 1985.

[27] A. Joshi and B. O. Palsson, "Metabolic dynamics in the human red cell. Part I. A comprehensive kinetic model," *Journal of Theoretical Biology*, vol. 141, no. 4, pp. 515–528, 1989.

[28] A. Joshi and B. O. Palsson, "Metabolic dynamics in the human red cell. Part II. Interactions with the environment," *Journal of Theoretical Biology*, vol. 141, no. 4, pp. 529–545, 1989.

[29] N. Jamshidi, J. S. Edwards, T. Fahland, G. M. Church, and B. O. Palsson, "Dynamic simulation of the human red blood cell metabolic network," *Bioinformatics*, vol. 17, no. 3, pp. 286–287, 2001.

[30] N. D. Price, J. L. Reed, J. A. Papin, S. J. Wiback, and B. O. Palsson, "Network-based analysis of metabolic regulation in the human red blood cell," *Journal of Theoretical Biology*, vol. 225, no. 2, pp. 185–194, 2003.

[31] N. Jamshidi and B. O. Palsson, "Top-down analysis of temporal hierarchy in biochemical reaction networks," *PLoS Computational Biology*, vol. 4, no. 9, Article ID e1000177, 2008.

[32] N. Jamshidi and B. O. Palsson, "Using in silico models to simulate dual perturbation experiments: procedure development and interpretation of outcomes," *BMC Systems Biology*, vol. 3, article 44, 2009.

[33] P. J. Mulquiney, W. A. Bubb, and P. W. Kuchel, "Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: *in vivo* kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using ^{13}C and ^{31}P NMR," *Biochemical Journal*, vol. 342, no. 3, pp. 567–580, 1999.

[34] P. J. Mulquiney and P. W. Kuchel, "Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement," *Biochemical Journal*, vol. 342, no. 3, pp. 581–596, 1999.

[35] P. J. Mulquiney and P. W. Kuchel, "Model of the pH-dependence of the concentrations of complexes involving metabolites, haemoglobin and magnesium ions in the human erythrocyte," *European Journal of Biochemistry*, vol. 245, no. 1, pp. 71–83, 1997.

[36] Y. Nakayama, A. Kinoshita, and M. Tomita, "Dynamic simulation of red blood cell metabolism and its application to the analysis of a pathological condition," *Theoretical Biology and Medical Modelling*, vol. 2, article 18, 2005.

[37] A. Kinoshita, Y. Nakayama, T. Kitayama, and M. Tomita, "Simulation study of methemoglobin reduction in erythrocytes: differential contributions of two pathways to tolerance to oxidative stress," *FEBS Journal*, vol. 274, no. 6, pp. 1449–1458, 2007.

[38] B. Hald, M. F. Madsen, S. Danø, B. Quistorff, and P. G. Sørensen, "Quantitative evaluation of respiration induced metabolic oscillations in erythrocytes," *Biophysical Chemistry*, vol. 141, no. 1, pp. 41–48, 2009.

[39] R. K. Dash and J. B. Bassingthwaighte, "Blood HbO_2 and $HbCO_2$ dissociation curves at varied O^2, CO_2, pH, 2,3-DPG and temperature levels," *Annals of Biomedical Engineering*, vol. 32, no. 12, pp. 1676–1693, 2004.

[40] L. L. Peters, R. A. Shivdasani, S.-C. Liu et al., "Anion exchanger 1 (band 3) is required to prevent erythrocyte membrane surface loss but not to form the membrane skeleton," *Cell*, vol. 86, no. 6, pp. 917–927, 1996.

[41] H. Chu and P. S. Low, "Mapping of glycolytic enzyme-binding sites on human erythrocyte band 3," *Biochemical Journal*, vol. 400, no. 1, pp. 143–151, 2006.

[42] A. Tsuneshige, K. Imai, and I. Tyuma, "The binding of hemoglobin to red cell membrane lowers its oxygen affinity," *Journal of Biochemistry*, vol. 101, no. 3, pp. 695–704, 1987.

[43] M. E. Campanella, H. Chu, and P. S. Low, "Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 7, pp. 2402–2407, 2005.

[44] M. E. Campanella, H. Chu, N. J. Wandersee et al., "Characterization of glycolytic enzyme interactions with murine erythrocyte membranes in wild-type and membrane protein knockout mice," *Blood*, vol. 112, no. 9, pp. 3900–3906, 2008.

[45] M. L. Ellsworth, T. Forrester, C. G. Ellis, and H. H. Dietrich, "The erythrocyte as a regulator of vascular tone," *American Journal of Physiology*, vol. 269, no. 6, pp. H2155–H2161, 1995.

[46] C. Lenfant, J. Torrance, E. English et al., "Effect of altitude on oxygen binding by hemoglobin and on organic phosphate levels," *Journal of Clinical Investigation*, vol. 47, no. 12, pp. 2652–2656, 1968.

[47] N. Hamasaki, T. Asakura, and S. Minakami, "Effect of oxygen tension on glycolysis in human erythrocytes," *Journal of Biochemistry*, vol. 68, no. 2, pp. 157–161, 1970.

[48] I. Rapoport, H. Berger, S. M. Rapoport, R. Elsner, and G. Gerber, "Response of the glycolysis of human erythrocytes to the transition from the oxygenated to the deoxygenated state at constant intracellular pH," *Biochimica et Biophysica Acta*, vol. 428, no. 1, pp. 193–204, 1976.

[49] I. A. Lewis, M. E. Campanella, J. L. Markley, and P. S. Low, "Role of band 3 in regulating metabolic flux of red blood cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 44, pp. 18515–18520, 2009.

[50] K. Suganuma, K. Tsukada, M. Kashiba et al., "Erythrocytes with T-state-stabilized hemoglobin as a therapeutic tool for postischemic liver dysfunction," *Antioxidants and Redox Signaling*, vol. 8, no. 9-10, pp. 1847–1855, 2006.

[51] I. Messana, M. Orlando, L. Cassiano et al., "Human erythrocyte metabolism is modulated by the O^2 -linked transition of hemoglobin," *FEBS Letters*, vol. 390, no. 1, pp. 25–28, 1996.

[52] S. V. Komarova, E. V. Mosharov, V. M. Vitvitsky, and F. I. Ataullakhanov, "Adenine nucleotide synthesis in human erythrocytes depends on the mode of supplementation of cell suspension with adenosine," *Blood Cells, Molecules, and Diseases*, vol. 25, no. 3-4, pp. 170–179, 1999.

[53] T. Shimizu, N. Kono, H. Kiyokawa et al., "Erythrocyte glycolysis and its marked alteration by muscular exercise in type VII glycogenosis," *Blood*, vol. 71, no. 4, pp. 1130–1134, 1988.

[54] M. Shiba, T. Mura, T. Masuyama, et al., "Preparation and preservation of red blood cell concentrates in MAP solution by quadruple bag system," *Japanese Journal of Transfusion Medicine*, vol. 37, pp. 404–410, 1991.

[55] C. F. Högman, F. Knutson, and H. Lööf, "Storage of whole blood before separation: the effect of temperature on red cell 2,3 DPG and the accumulation of lactate," *Transfusion*, vol. 39, no. 5, pp. 492–497, 1999.

[56] J. C. Ellory and J. S. Willis, "Kinetics of the sodium pump in red cells of different temperature sensitivity," *Journal of General Physiology*, vol. 79, no. 6, pp. 1115–1130, 1982.

[57] J. A. MacDonald and K. B. Storey, "Temperature and phosphate effects on allosteric phenomena of phosphofructokinase from a hibernating ground squirrel (Spermophilus lateralis)," *FEBS Journal*, vol. 272, no. 1, pp. 120–128, 2005.

[58] R. K. Dash and J. B. Bassingthwaighte, "Simultaneous blood-tissue exchange of oxygen, carbon dioxide, bicarbonate, and hydrogen ion," *Annals of Biomedical Engineering*, vol. 34, no. 7, pp. 1129–1148, 2006.

[59] T. Betz, M. Lenz, J.-F. Joanny, and C. Sykes, "ATP-dependent mechanics of red blood cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 36, pp. 15320–15325, 2009.

[60] J. B. Bassingthwaighte, G. M. Raymond, J. D. Ploger, L. M. Schwartz, and T. R. Bukowski, "GENTEX, a general multiscale model for in vivo tissue exchanges and intraorgan metabolism," *Philosophical Transactions of the Royal Society A*, vol. 364, no. 1843, pp. 1423–1442, 2006.

[61] J. C. Arciero, B. E. Carlson, and T. W. Secomb, "Theoretical model of metabolic blood flow regulation: roles of ATP release by red blood cells and conducted responses," *American Journal of Physiology*, vol. 295, no. 4, pp. H1562–H1571, 2008.

[62] J. W. Vince and R. A. F. Reithmeier, "Carbonic anhydrase II binds to the carboxyl terminus of human band 3, the erythrocyte Cl^-/HCO_3^- exchanger," *Journal of Biological Chemistry*, vol. 273, no. 43, pp. 28430–28437, 1998.

[63] R. E. Weber, W. Voelter, A. Fago, H. Echner, E. Campanella, and P. S. Low, "Modulation of red cell glycolysis: interactions between vertebrate hemoglobins and cytoplasmic domains of band 3 red cell membrane proteins," *American Journal of Physiology*, vol. 287, no. 2, pp. R454–R464, 2004.