

A Novel Method for the Integration of Stochastic Petri Net Simulation and Transcriptomic Data Applied to a Metabolic Pathway

Carvalho LM¹, Meirelles GV², Pereira GAG^{1*} and Carazzolle MF¹

¹Laboratory of Genomics and BioEnergia, Department of Genetics and Evolution, Microbiology and Immunology, São Paulo, Brazil

²National Laboratory of Biosciences, National Center for Research in Energy and Materials, Campinas, SP, Brazil

Abstract

Over the years, methods capable of integrating data from omics, such as transcriptomics, proteomics and metabolomics have emerged in Systems Biology, principally the use of networks to integrate omics information. In particular, the role of each biological pathway aims to understand the intermolecular interactions. While there are theoretical and experimental strategies to investigate biological pathways involved in cellular metabolism, computational modeling methods allow for a better understanding of them. Here we propose a new method to connect transcriptomic data with simulation approach using stochastic Petri Net (PN) metabolic networks. This new approach was developed based on well-studied theoretical gene expression modeling while trying to assimilate dynamic ordinary systems to a stochastic model function. The developed method was used to perform stochastic PN simulations of ethanol fermentation by *Saccharomyces cerevisiae* considering glucose and xylose as carbon sources. Lastly, we developed the PeNTIOS software, which is capable of converting *Saccharomyces cerevisiae* metabolic pathways and transcriptomic data into SBML format. The generated files can be readily imported into PN simulation programs. Our results show that the reconstruction of stochastic PN systems with transcriptomics data is a promising method that can generate new insights about biological experiments, as shown through our case study with the xylose-fermenting yeast.

Keywords: Stochastic petri net; Transcriptomics; Metabolic simulation; Biological networks; Ethanol 2G; Yeast; Metabolomics

Introduction

The large-scale generation of biological information on metabolism and gene regulation led to the development of new bioinformatics methods based on mathematical modeling which, coupled with multi-omics approaches, have enabled a deeper understanding of cellular systems. In this context, systems biology emerges as a field of study capable of integrating omics data and metabolic networks to extract information about biological responses and phenotypes.

The standard analysis in metabolic network simulation is Flux Balance Analysis (FBA) [1], an approach that studies metabolism through the reconstruction of metabolic pathways at the genomic scale and the integration of omics data [2]. This mathematical representation uses linear programming to solve reaction equations and to calculate possible flows using a metabolic model. In order to apply this methodology, one must have a completely validated model supported by experimental data, such as fermentation profiles and cellular growth. Moreover, transcriptomic and metabolomic data can be used as additional constraints to the model and to give support for the evaluation step.

Many other methodologies have been used for modeling complex biological systems. Examples include the ordinary differential equations (DOE), Boolean networks and Bayesian models. The Petri net (PN), introduced by Carl Petri in early 1960s [3], emerges as a potential mathematical approach for modeling biological systems [4] and metabolic networks [5]. There are various Petri Net models, such as classical Petri Net, Continuous Petri Net, Colored Petri Net (CPT) and Stochastic Petri Net (SPN). SPN preserves the conditions of qualitative networks with the addition of an exponential probability distribution (delay function) to each transition trigger action. The modeling of stochastic networks can be generated by adding the transition probability estimated from previously available data. Considering the complexity of biological systems, the use of stochastic modeling

represents an improvement over FBA, but it also brings the challenge of choosing the appropriate delay function.

In the present work we describe a new method that uses transcriptomic data to estimate the delay function in SPN and applies it to metabolic network simulations. *Saccharomyces cerevisiae* considering glucose and xylose as carbon sources, a case study for both the first and the second-generation ethanol production processes [6]. Lastly, we developed the PeNTIOS software, which is capable of converting *Saccharomyces cerevisiae* metabolic pathways and transcriptomic data into SBML format. The generated files can be readily imported into PN simulation programs.

Methods and Implementation

In a metabolic PN model, the network representation consists of states, transitions and tokens, normally represented by circles, rectangles and black dots, respectively. State S represents the substance (metabolites, cofactors, etc) and transition T represents the biochemical reaction, where the token-pass is activated only by the transition T, since the state immediately above contains a token. Edges connect a state to transition and represent the flux of the reaction promoted by token passing [7]. PN always reach the next state, unless if there is not any token to fire, resulting in the end of simulation. Other advantage in Petri Net simulation is the use of logical nodes. This type of notation

***Corresponding author:** Pereira GAG, Department of Genetics, State University of Campinas, São Paulo, Brazil, Tel: +551935216237; E-mail: gongalo@unicamp.br

Received December 06, 2018; **Accepted** December 26, 2018; **Published** January 01, 2019

Citation: Carvalho LM, Meirelles GV, Pereira GAG, Carazzolle MF (2018) A Novel Method for the Integration of Stochastic Petri Net Simulation and Transcriptomic Data Applied to a Metabolic Pathway. J Comput Sci Syst Biol 12: 9-14. doi:10.4172/jcsb.1000293

Copyright: © 2018 Carvalho LM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

is useful to represent multiple nodes in network that cause the same trigger during simulation.

SPN can be expressed as a stochastic probability function $F(t)$ of a PN transition using an exponential delay function given by $F(t)=1-e^{-\lambda t}$, which has an expected distribution value of $\frac{1}{\lambda}$. In terms of PN transition, different values of λ change the probability function (Figure 1A) and generate different final states to the network simulations. Thus, the correct choice of λ values for each transition is an essential step for simulations to converge the results into a real biological system.

Thus, our method is based on a gene expression model [8], in which the mRNA is produced from DNA at a rate k_m and degraded at a rate γ_m , as shown in equation 1:

$$\frac{d[mRNA]}{dt} = k_m - \gamma_m[mRNA] \quad (1)$$

The deterministic analysis of the model reveals a temporal variation of mRNA formation over time in function of k_m and γ_m ratio. The analytical solution of $mRNA(t)$ and its final stationary state are observed in the equations 2 and 3, respectively, where the value represents the stationary state of this system.

$$mRNA(t) = c_1 e^{-(\gamma_m)t} + \frac{k_m}{\gamma_m} \quad (2)$$

$$\frac{d[mRNA]}{dt} = 0 \Rightarrow mRNA = \frac{k_m}{\gamma_m} \quad (3)$$

In a biological system, considering that at any given moment the metabolic flux of the cells is constant as a function of time (called steady state), we can assume that the transcription and degradation rates, and its respective ratio $\frac{k_m}{\gamma_m}$, will also be constant in the steady state. In this condition, we can consider that the ratio $\frac{k_m}{\gamma_m}$ can be approximated by mRNA expression data obtained from transcriptome analysis, such as RPKM, FPKM, TPM or normalized microarray [9].

In this context, our method proposes a positive correlation between λ values and normalized gene expression data obtained from transcriptome analysis. This insight was achieved by comparing the graph of the delay function over the time (Figure 1A) and the graph of the theoretical deterministic model of inverse mRNA production over time (Figure 1B).

Based on the comparison of the functions steady state, we associated the expected value $1/\lambda$ of delay function with the plateau of the inverse of the equation (2) achieved in $\frac{1}{\frac{k_m}{\gamma_m}}$, implicating in $\lambda = \frac{k_m}{\gamma_m}$ and, therefore, associating the λ value of the delay function with normalized gene expression data.

Since the normalized expression values range from 0 to large positive values, it is necessary to rescale the range to values around 1. Thus, we applied a simple median normalization by dividing the obtained values by the median value of the expression data, resulting in a transition function for a gene i , $F_i(t)$, defined as, where $FPKM_i$ is

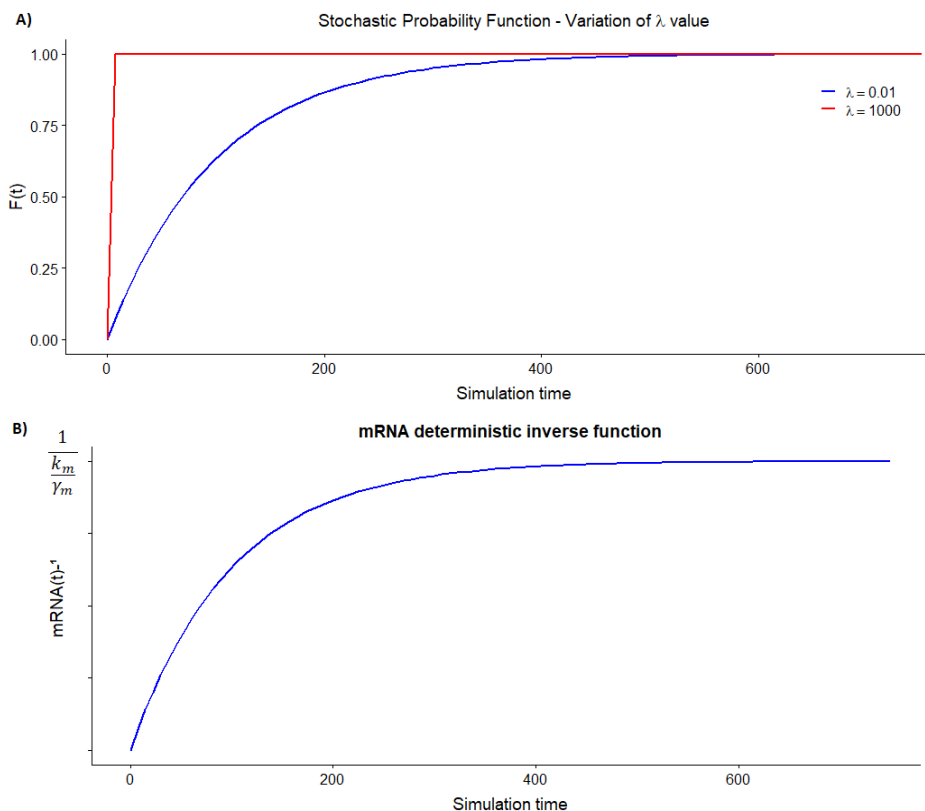


Figure 1: A) Behavior of $F(t)=1-e^{-\lambda t}$ using low (blue) and high (red) values of λ and B) mRNA deterministic inverse function $mRNA^{-1}(t)$ over the time.

the expression value of gene i in the condition of interest and FPKMs represents the arrays expression of all genes in that condition.

Although this new method can be applied to any organism, we focused our analyses on the metabolic pathways of *Saccharomyces cerevisiae*, with a case study of the consumption of glucose and xylose and their conversion into ethanol. For this purpose, we implemented the PeNTIOS software (<https://github.com/lmigueel/PeNTIOS>) in Perl language to automatically generate a SBML file for the PN network simulations, including the calculation of the λ value extracted from transcriptomic data. PeNTIOS requires as inputs a (1) pathway name from the *Saccharomyces* Genome Database (SGD) (example: glycolysis)

and (2) a tab-separated text file containing names and expression data (RPKM, FPKM, TPM, CPM or normalized microarray) for all genes in the condition of interest. PeNTIOS performs the download of metabolic pathways from SGD (HTML format) and makes a conversion into SBML format parsing the reactions. To attribute the correct gene name for each reaction ID, each reaction is downloaded from the HTML page using a reaction ID and parsed to recover the gene name. After that, gene expression data from each gene identified in metabolic pathway is extracted from transcriptomic data, normalized by median and converted to SBML math format for simulation (Figure 2). Furthermore, all the simulation can be performed using Snoopy2 v1.21 [10] in Windows/Linux.

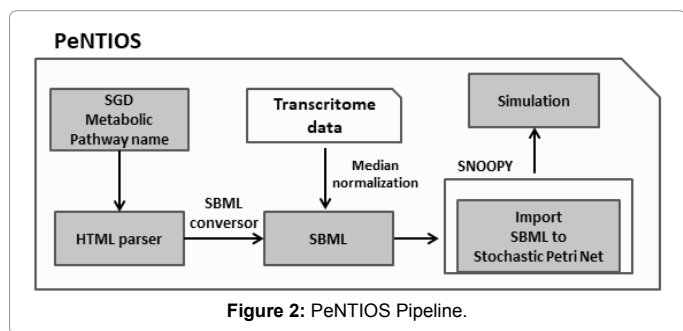


Figure 2: PeNTIOS Pipeline.

Example of Usage

In order to validate our method, we focused on the analysis of ethanol production by *Saccharomyces cerevisiae* having glucose and xylose as carbon sources. Although yeast glucose uptake and its conversion into ethanol has been well studied through several experimental techniques and models [11,12], xylose fermentation remains an unresolved problem that requires the development of new genetically modified strains and presents metabolism bottlenecks. Thus, several strategies have been employed to study this system, including transcriptomic and metabolomic approaches [12,13].

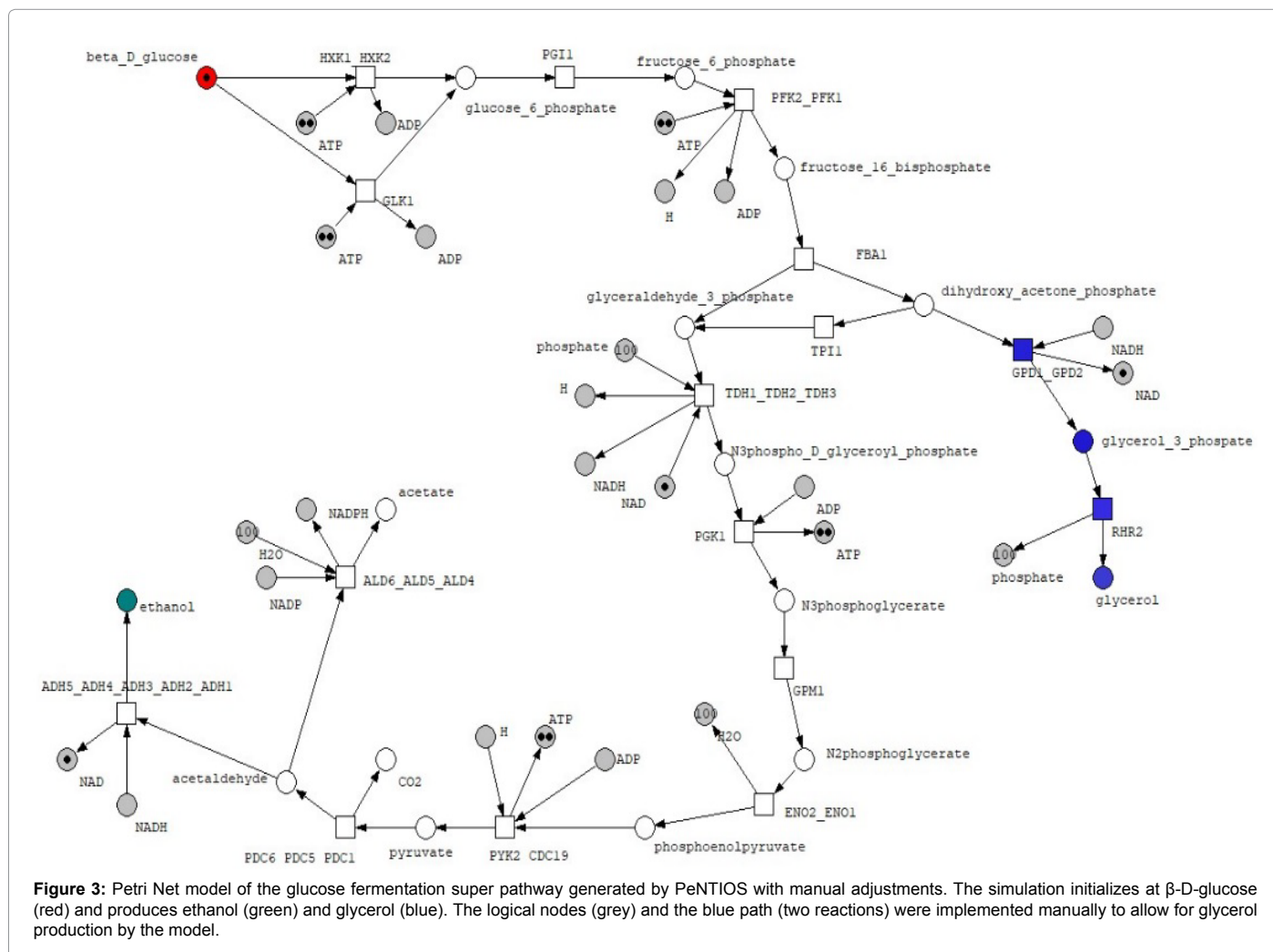


Figure 3: Petri Net model of the glucose fermentation super pathway generated by PeNTIOS with manual adjustments. The simulation initializes at β -D-glucose (red) and produces ethanol (green) and glycerol (blue). The logical nodes (grey) and the blue path (two reactions) were implemented manually to allow for glycerol production by the model.

Ethanol production from glucose

PeNTIOS was used to convert the super pathway of glucose fermentation (glycolysis, glycerol biosynthesis and alcoholic fermentation pathways) and a publicly available transcriptome dataset (FPKM values) [11] into a SBML file containing all metabolic reactions, cofactors and λ value calculated by rescaled FPKM values. After the PeNTIOS conversion, all cofactors nodes, such as ATP and NADH, were manually converted to logical nodes. Also, the cofactor was modified from NAD-dependent to NADP-dependent in the acetate production reactions, as previously described [14], and two other reactions for acetate formation (ALD4 and ALD5) were excluded as they involve mitochondrial genes (details in Supplementary Figure 1). Finally, as shown in Figure 3, the initial conditions of the tokens in SPN simulation were configured to 1, 2, 1 for glucose, ATP and NAD, respectively, in order to allow for glucose consumption in the first steps of the metabolic pathway. The simulation was performed using Gillespie model over 1000 iterations during 60 ms. We compared the scenario where all the transition rates were defined as $\lambda=1$, symbolizing the simulation without any external information, against the scenario where all transition rates were corrected by the transcriptome data.

In the first scenario, the simulation resulted in ethanol and glycerol as the main products at a proportion of 50% each (Figure 4A). On the other hand, in the second scenario, the simulation resulted in the production of ethanol and glycerol at a ratio of 9:1 (Figure 4B), showing a good correlation with experimental results (Figure 4C). These results reveal the potential of the new method for recovering the essence of biological systems through computational simulation supported by transcriptomic data.

Ethanol production from xylose

The production of ethanol using xylose as a carbon source is commonly referred to as second generation (2G) ethanol technology. It is a promising process which can increase production and reduce the costs associated with the first-generation (1G) ethanol production from glucose. The 2G process is based on the deconstruction of lignocellulosic biomass to release fermentable sugars (mainly glucose

and xylose) [6]. Since wild type yeast cannot consume xylose, it is necessary to perform genetic modifications that enable xylose consumption, such as the insertion of the exogenous xylose pathway genes xylose reductase (XR) and xylitol dehydrogenase (XDH) [15,16]. These specific modifications cause an unbalanced redox in the yeast cell, leading to xylitol accumulation and ultimately hindering the production of ethanol.

In order to evaluate the performance of our new integration method in the study of xylose-fermenting yeast, we created a SPN containing a xylose consumption pathway by insertion of the XR and XDH genes and the pentose phosphate pathway (PPP) into the already described model (Supplementary Figure 1). The stochastic simulation was then performed using as input transcriptomic data from a study where glucose and xylose consumption and their conversion into ethanol by a genetically modified yeast were investigated [12]. The simulation was started with 5 tokens of glucose or xylose, 10 tokens of ATP, 5 tokens of NAD⁺ and 5 tokens of NADP⁺. The Gillespie model was used over 1000 iterations during 60 ms. The resulting number of tokens of ethanol and xylitol predicted by the simulation with glucose and xylose as carbon sources are shown in Figure 5A. A comparison of these numbers with experimental data, obtained from Zeng et al. [12], shows very similar end-products profiles (Figure 5B). The lower ethanol yield observed during the consumption of xylose when compared to glucose is related to the accumulation of xylitol, which is produced by the XR reaction and cannot be consumed by the XDH reaction due to the unavailability of the NAD cofactor.

In order to reinforce the correlation between simulation and experimental results, a publicly available metabolomic dataset from yeast genetically modified for xylose consumption was obtained from Wasylenko et al. [13]. Figure 5C shows the number of tokens of internal metabolites and the NAD/NADH ratio obtained through the simulation using glucose and xylose as carbon sources. The comparison with the retrieved metabolomic data (Figure 5D) reveals good agreement between simulation and experimental results. This is especially true for the high accumulation of intracellular xylulose 5-phosphate (X5P) and the reduced NAD/NADH ratio. The reduction

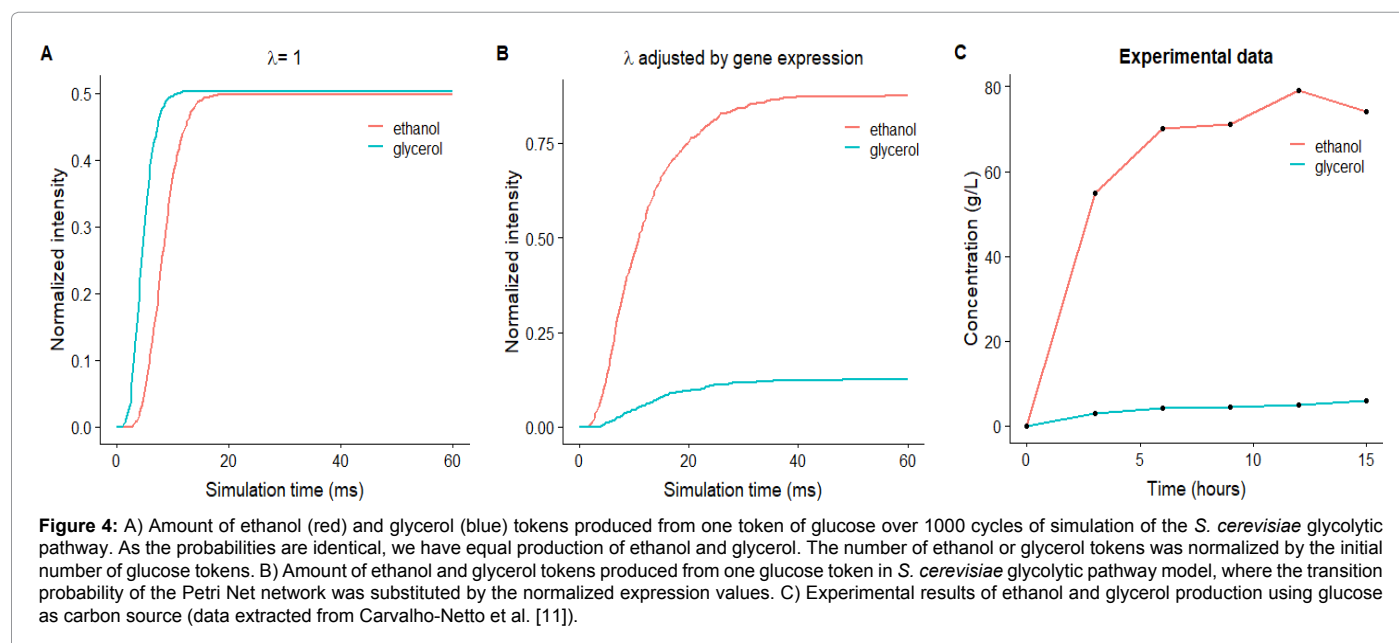
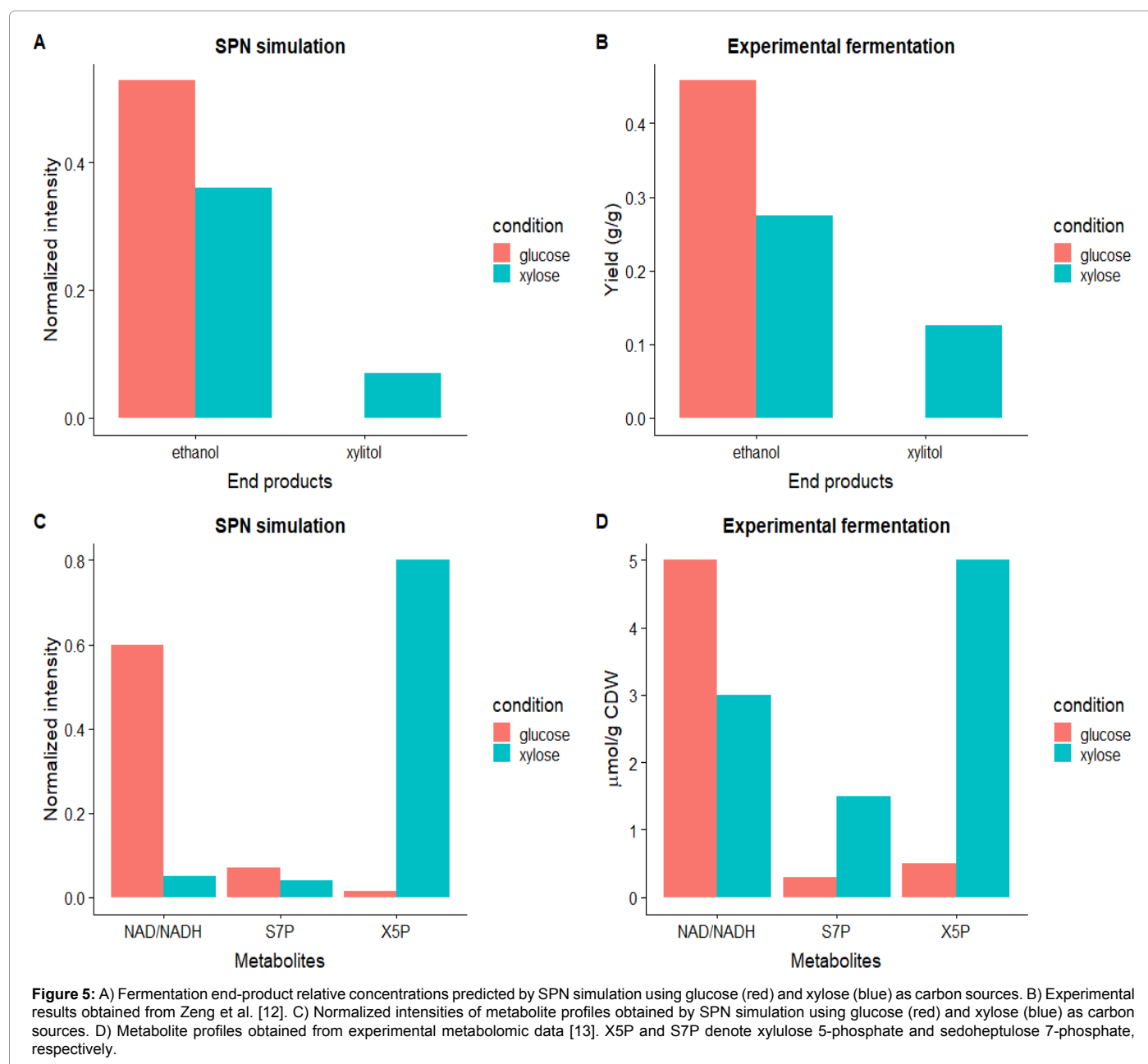


Figure 4: A) Amount of ethanol (red) and glycerol (blue) tokens produced from one token of glucose over 1000 cycles of simulation of the *S. cerevisiae* glycolytic pathway. As the probabilities are identical, we have equal production of ethanol and glycerol. The number of ethanol or glycerol tokens was normalized by the initial number of glucose tokens. B) Amount of ethanol and glycerol tokens produced from one glucose token in *S. cerevisiae* glycolytic pathway model, where the transition probability of the Petri Net network was substituted by the normalized expression values. C) Experimental results of ethanol and glycerol production using glucose as carbon source (data extracted from Carvalho-Netto et al. [11]).



of the NAD cofactor observed in the simulation and in the metabolomic data is related to xylitol production, as previously discussed. Moreover, the accumulation of xylulose 5-phosphate in the PPP during xylose consumption suggests a metabolism bottleneck that can be a target for genetic manipulation, such as the over expression of the transketolase (TKL1) gene.

The overall results showed that the here described method, which is based on the integration of transcriptomic data with Stochastic Petri Net simulation, can be used to predict end-products and internal metabolites profiles. The possibility of performing stochastic simulations guided by transcriptomic data enables the use of new strategies for interpreting transcriptome and metabolome experiments, besides generating new insights on metabolism. Although PeNTIOS has been implemented for *Saccharomyces cerevisiae*, the method developed in this work can be applied to other organisms.

References

1. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nature Biotechnology 28: 245-248.
2. Blazier AS, Papin JA (2012) Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 3: 299.
3. Petri CA (1962) In: Communication with automata. Schriften des IMM 3: 16-27.
4. Koch I, Reissig W, Schreiber F (2011) Modeling in systems biology: The Petri net approach. Computational Biology.
5. Liu F, Heiner M (2014) Petri nets for modeling and analyzing biochemical reaction networks. In Approaches in Integrative Bioinformatics: 245-272.
6. dos Santos LV, de Barros Grassi MC, Gallardo JCM, Pirolla RAS, Calderón LL, et al. (2016) Second-generation ethanol: the need is becoming a reality. Industrial Biotechnology 12: 40-57.
7. Pinney JW, Westhead DR, McConkey GA (2003) Petri Net representations in systems biology. Biochemical Society Transactions 31: 1513-1515.

8. Chen T, He HL, Church GM (1999) Modeling gene expression with differential equations. *Biocomputing* 99: 29-40.
9. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology* 17: 13.
10. Rohr C, Manwan W, Heiner M (2010) Snoopy-a unifying Petri net frame work to investigate biomolecular networks. *Bioinformatics* 26: 974-975.
11. Carvalho-Netto OV, Carazzolle MF, Mofatto LS, Teixeira PJ, Noronha MF, et al. (2015) *Saccharomyces cerevisiae* transcriptional reprogramming due to bacterial contamination during industrial scale bioethanol production. *Microbial Cell Factories* 14: 13.
12. Zeng WY, Tang YQ, Gou M, Xia ZY, Kida K (2016) Transcriptomes of a xylose-utilizing industrial flocculating *Saccharomyces cerevisiae* strain cultured in media containing different sugar sources. *AMB Express* 6: 51.
13. Wasylenko TM, Stephanopoulos G (2015) Metabolomic and ¹³C-metabolic flux analysis of a xylose-consuming *Saccharomyces cerevisiae* strain expressing xylose isomerase. *Biotechnology and Bioengineering* 112: 470-483.
14. Singh S, Brocker C, Koppaka V, Chen Y, Jackson BC, et al. (2013) Aldehyde Dehydrogenases in Cellular Responses to Oxidative/electrophilic Stress. *Free Radical Biology and Medicine* 56: 89-101.
15. Runquist D, Hahn-Hägerdal B, Bettiga M (2010) Increased ethanol productivity in xylose-utilizing *Saccharomyces cerevisiae* via a randomly mutagenized xylose reductase. *Applied and Environmental Microbiology* 76: 7796-7802.
16. Suga H, Matsuda F, Hasunuma T, Ishii J, Kondo A (2013) Implementation of a transhydrogenase-like shunt to counter redox imbalance during xylose fermentation in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* 97: 1669-1678.