2019

Applying Molecular Communication Theory to Multi-Scale Integrated Models of Biological Pathways

Zahmeeth Sakkaff

Bige Deniz Unluturk

Massimiliano Pierobon

Published in NANOCOM '19: Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing a
Communication, September 2019, Article No.: 41, Pages 1–2

Applying Molecular Communication Theory to Multi-Scale Integrated Models of Biological Pathways

Zahmeeth Sakkaff
Department of Computer Science and Engineering
University of Nebraska-Lincoln
zsayedsa@cse.unl.edu

Bige Deniz Unluturk
School of Electrical and Computer Engineering
Georgia Institute of Technology
bunluturk@gatech.edu

Massimiliano Pierobon
Department of Computer Science and Engineering
University of Nebraska-Lincoln
pierobon@cse.unl.edu

ABSTRACT

The natural communication ability of cells is explored in this paper by providing preliminary results in the estimation of the Mutual Information (MI) of signaling pathway communication channels. These results, based on an application of Molecular Communication (MC) and information theory concepts to multi-scale integrated Flux-Balance Analysis (iFBA) models are a first step to evaluate the potential of cells and their biochemical processes as a substrate for enabling engineered MC channels for the future internet of Bio-Nano Things.

KEYWORDS

Molecular Communication, Mutual Information, Cell Signaling Pathways, Flux Balance Analysis, Whole Cell Model

1 INTRODUCTION

The Internet of Bio-NanoThings (IoBNT) is envisioned as a network of genetically engineered cells and bio and nano-technology enabled devices interlinked through Molecular Communication (MC), which is an emerging paradigm where tools from wireless communications and information theory are applied to the realm of biochemical reactions [1]. The natural communication ability of cells, *i.e.*, cell signaling, is one of the best candidates as a substrate for enabling MC channels for IoBNT. To make use of cell signaling components and functionality to design MC systems, it is necessary to first characterize the information theoretic performance of these channels in their natural settings, as they enable the exchange of information between cells and their environment.

To realize the aforementioned goal, we propose in this paper an MC abstraction of the information flow underlying cell signalling, which is controlled by complex biological pathways, *i.e.*, chains of chemical reactions. For this, we detail an ***information & communication-centric*** view of a biological cell to quantify the potential of the cell signaling pathways, usually classified into signal transduction, gene regulation, and metabolic pathways [2], as shown in Fig. 1, to propagate information from the extracellular environment into the cell, and encode it into changes in the cell's behavior. The contribution in this paper complements our previous publications independently focused on signal transduction pathways [3] and metabolism [4], by developing an end-to-end view of cell signaling pathways linked by the regulation of gene expression.

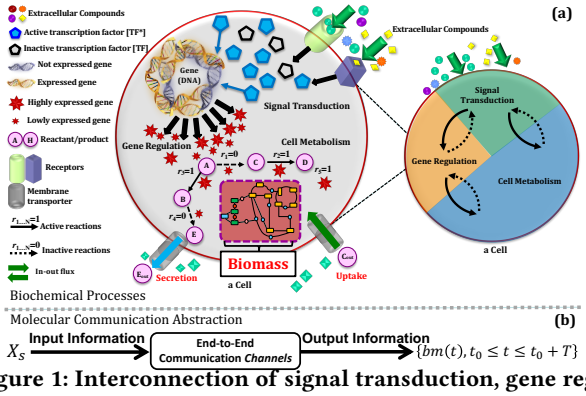


Figure 1: Interconnection of signal transduction, gene regulation and metabolic pathways.

The rest of the paper is organized as follows. In Sec. 2, we detail the aforementioned MC abstraction comprising all three cellular pathways for the first time in the literature, and we propose a computational approach to estimate the information theoretic performance of these pathways based on the application of MC and information theory concepts to multi-scale integrated Flux-Balance Analysis (iFBA) models [2]. Preliminary numerical results of this approach are presented in Sec 3. In Sec. 4, we conclude the paper.

2 MOLECULAR COMMUNICATION ABSTRACTION OF CELL SIGNALING PATHWAYS

The **biochemical processes** underlying cell signaling pathways are abstracted in this paper as communication channels that propagate information on the concentration values of *extracellular chemical compounds* (*e.g.*, sugars such as the glucose, lactose, and glucose6P at the basis of the numerical results in this paper) into the cell, ultimately resulting into the modulation of the cell growth rate, *i.e.*, the *Biomass*. This is realized through the cascade of signal transduction, gene regulation, and metabolic pathways as follows. The signal transduction pathways, according to the aforementioned extracellular information sensed through *receptors*, modulate the activation of one or more *transcription factors*, a particular type of protein, *i.e.*, a biological macromolecule that results from the expression of a DNA gene. As a result of an active transcription factor, one or more genes can be up-regulated (activation), therefore expressing the proteins at a higher rate, or otherwise down-regulated

NANOCOM '19, The Sixth Annual ACM International Conference on Nanoscale Computing and Communication, September 25–27, 2019, Dublin, Ireland

ACM ISBN 978-1-4503-6897-1/19/09…$15.00
https://doi.org/10.1145/3345312.3345495

(repression). In gene regulation pathways, networks of genes interacting through mutual activation or repression of their expression propagate the activation/repression modulated by the signal transduction pathway into the modulation of the expression rate of proteins with different functionality, such as enzymes. Metabolic pathways are chains of chemical reactions that decompose nutrient compounds uptaken from the extracellular environment by *transporter proteins* into energy, biomass, and waste (secreted by the cell into the extracellular environment), where the presence/absence of the corresponding enzymes *activates/inactivates these reactions*.

In our **proposed MC abstraction** of cell signaling pathways, the channel *input* is represented by the concentration values X_s of $s = 1, \cdots, S$ information-bearing extracellular chemical compounds at time t_0; the *signaling pathway communication channels* are modeled and simulated for an *E. coli* bacterium as in [2], by combining ordinary differential equations (ODE's), regulatory Boolean logic, and FBA on metabolic reactions; the channel *input* is represented by the resulting biomass values $bm(t)$ ranging from t_0 to the time $t_0 + T$, where the biomass reaches a steady state.

The final goal of our computational approach is the **estimation of the information theoretic performance** in terms of estimated Mutual Information (MI) $\tilde{I}_{bm}$ for each considered input chemical compound [5], expressed as

$$\tilde{I}_{bm} = \tilde{H}(X) - \tilde{H}(X|bm(t_c)), \tag{1}$$

where $\tilde{H}(.)$ and $\tilde{H}(.|.)$ denote the estimated entropy and conditional entropy, respectively, X is the input concentration of information-bearing molecules and $bm(t_c)$ is the corner concentration of the biomass where the growth trend changes from exponential to linear due to the total consumption of the input sugar, where $t_c \subset [t_0, t_0 + T]$ is the time instant when this critical point is reached.

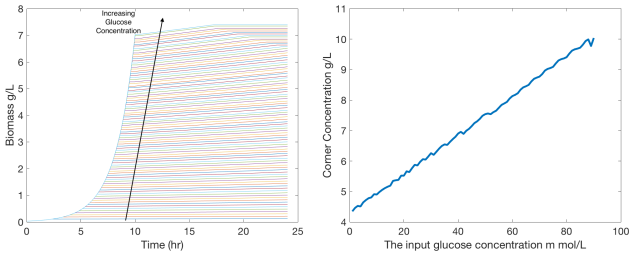
The necessary *data for the MI estimations* is obtained through iFBA simulations [2]. For each value x_i, $i = 0, \ldots, I$, of the input concentration X sampled from the range between x_{min} and x_{max}, defined here as the value below and above which the concentrations of biomass do not significantly change. The *estimated input entropy* $\tilde{H}(X)$ is computed through the histogram approach [3] as

$$\tilde{H}(X) = -\sum_{i=1}^{I} p_X(x_i) \log_2 \left(\frac{p_X(x_i)}{w_X} \right), \tag{2}$$

where $p_X(x_i) = 1/I$, *i.e.*, a uniformly distributed input [6], and w_X is the sampling interval $(x_{max} - x_{min})/I$. The *estimated conditional entropy* $\tilde{H}(X|bm(t_c))$ is computed as

$$\tilde{H}(X|bm(t_c)) = -\sum_{n=1}^{N_{bm(t_c)}} p_{bm(t_c)}(\{bm(t_c)\}_n) \sum_{s=1}^{S_{\{bm(t_c)\}_n}} p_{X|\{bm(t_c)\}_n}(x_s) \log_2 \left(\frac{p_{X|\{bm(t_c)\}_n}(x_s)}{w_{X,\{bm(t_c)\}_n}} \right), \tag{3}$$

where $\{bm(t_c)_n\}_{n=1}^{N_{bm(t_c)}}$ is the set of obtained values for the biomass $bm(t_c)$, divided into $N_{bm(t_c)}$ bins, $S_{\{bm(t_c)\}_n}$ and $w_{X,\{bm(t_c)\}_n}$ are the number and the size of histogram bins considered for the input concentration X to compute the histogram $p_{X|\{bm(t_c)\}_n}(x_s)$, where $w_{X,S_{\{bm(,t_c)\}_n}} = (x_{max} - x_{min})/S_{\{bm(t_c)\}_n}$ and x_s is a value from $\{x_i\}_{i=0}^{I}$ sampled according to the histogram. The numbers of histogram bins $N_{bm(t_c)}$ and $S_{\{bm(t_c)\}_n}$ are computed from the aforementioned simulation data according to the Doane's formula [3].



(a) **(b)**

Figure 2: (a) Biomass for different glucose concentration levels. (b) Corner concentration for glucose.

3 NUMERICAL RESULTS

The concentrations of three different extracellular chemical compounds (sugars), namely, glucose, lactose, and glucose6P, are varied in a range between $x_{min} = 0$ $[mmol/L]$ and $x_{max} = 90$ $[mmol/L]$ with 1 $[mmol/L]$ increments, where the bounds are set considering the uptake flux limits of a cell for these compounds. For each concentration level, iFBA simulations are run, and the resulting biomass is recorded. From this, the estimated MI is calculated as described above. Fig. 2-a shows the change in biomass as function of the time for each concentration level in $[x_{min}\ x_{max}]$. The output information is the corner concentration of the biomass reached due to the total consumption of sugar, shown in Fig. 2-b with respect to the input concentration. By applying the mutual information estimation method described in Sec. 2, the mutual information values are obtained as in Tab. 1. We observe that the MI of the glucose, which has higher efficiency for *E. coli* [2], has a higher MI value than the less preferred sugars lactose and glucose6P.

Table 1: Mutual information values for different sugars.

Sugar Type	$\tilde{H}(X)$	$\tilde{H}(X\|bm(t_c))$	$\tilde{H}(X) - \tilde{H}(X\|bm(t_c))$
Glucose	6.6295	5.2694	1.3601
Lactose	6.6295	5.2851	1.3444
Glucose6P	6.6295	5.8971	0.7324

4 CONCLUSION

In this paper, we introduced a computational approach to estimate the information flow in biological cells underlying the acquisition of extracellular information and its impact on the cell growth. We believe that this is the first step to enable the analysis and subsequent utilization of the biochemical processes underlying natural cellular communication mechanisms for engineering novel communication systems within the IoBNT paradigm.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants CISE CNS-1763969 and CCF-1816969.

REFERENCES

[1] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, 2015.
[2] M. W. Covert, N. Xiao, T. J. Chen, and J. R. Karr, "Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli," *Bioinformatics*, vol. 24, no. 18, pp. 2044–2050, 2008.
[3] Z. Sakkaff, A. Immaneni, and M. Pierobon, "Estimating the molecular information through cell signal transduction pathways," in *2018 IEEE SPAWC*, 2018, pp. 1–5.
[4] M. Pierobon, Z. Sakkaff, J. L. Catlett, and N. R. Buan, "Mutual information upper bound of molecular communication based on cell metabolism," in *2016 IEEE SPAWC*, 2016, pp. 1–6.
[5] A. Rhee, R. Cheong, and A. Levchenko, "The application of information theory to biochemical signaling systems," *Physical biology*, vol. 9, no. 4, p. 045011, 2012.
[6] R. Suderman, J. A. Bachman, A. Smith, P. K. Sorger, and E. J. Deeds, "Fundamental trade-offs between information flow in single cells and cellular populations," *PNAS*, vol. 114, no. 22, pp. 73–85, May 2017.