2017

End-to-end Molecular Communication Channels in Cell Metabolism: an Information Theoretic Study

Zahmeeth Sakkaff

Massimiliano Pierobon

Jennie L. Catlett

Nicole R. Buan

Christine A. Kelley

See next page for additional authors

Authors

Zahmeeth Sakkaff, Massimiliano Pierobon, Jennie L. Catlett, Nicole R. Buan, Christine A. Kelley, Mikaela Cashman, and Myra B. Cohen

End-to-end Molecular Communication Channels in Cell Metabolism: an Information Theoretic Study

Zahmeeth Sakkaff
Dept. of Comp. Sci. & Eng.
University of Nebraska-Lincoln
zsayedsa@cse.unl.com

Jennie L. Catlett
Dept. of Biochemistry
University of Nebraska-Lincoln
jcatlett@acm.org

Mikaela Cashman
Dept. of Comp. Sci. & Eng.
University of Nebraska-Lincoln
mcashman@cse.unl.edu

Massimiliano Pierobon
Dept. of Comp. Sci. & Eng.
University of Nebraska-Lincoln
pierobon@cse.unl.edu

Nicole R. Buan
Dept. of Biochemistry
University of Nebraska-Lincoln
nbuan@unl.edu

Myra B. Cohen
Dept. of Comp. Sci. & Eng.
University of Nebraska-Lincoln
myra@cse.unl.edu

Christine A. Kelley
Dept. of Mathematics
University of Nebraska-Lincoln
ckelley2@math.unl.edu

ABSTRACT

The opportunity to control and fine-tune the behavior of biological cells is a fascinating possibility for many diverse disciplines, ranging from medicine and ecology, to chemical industry and space exploration. While synthetic biology is providing novel tools to reprogram cell behavior from their genetic code, many challenges need to be solved before it can become a true engineering discipline, such as reliability, safety assurance, reproducibility and stability. This paper aims to understand the limits in the controllability of the behavior of a natural (non-engineered) biological cell. In particular, the focus is on cell metabolism, and its natural regulation mechanisms, and their ability to react and change according to the chemical characteristics of the external environment. To understand the aforementioned limits of this ability, molecular communication is used to abstract biological cells into a series of channels that propagate information on the chemical composition of the extracellular environment to the cell's behavior in terms of uptake and consumption of chemical compounds, and growth rate. This provides an information-theoretic framework to analyze the upper bound limit to the capacity of these channels to propagate information, which is based on a well-known and computationally efficient metabolic simulation technique. A numerical study is performed on two human gut microbes, where the upper bound is estimated for different environmental compounds, showing there is a potential for future practical applications.

NANOCOM '17, Washington D.C., USA, 2017

978-1-4503-4931-4/17/09…$15.00
DOI: http://dx.doi.org/10.1145/3109453.3109474

KEYWORDS

Molecular communication, information theory, cell metabolism.

1 INTRODUCTION

Biological cells are the essential building blocks of life, and their biochemical behavior, *e.g.*, how they consume nutrients, grow, reproduce, and secrete byproducts, shapes our life and our planet as we know it [13]. The possibility to control and fine tune cell behavior is one of the main objectives of not only life sciences, but also engineering and technology. Existing research has focused on several different techniques to realize this control, such as the use of light, *i.e.*, optogenetics [19], and magnetic fields, *i.e.*, magnetic nanoparticles [6]. At the cutting edge of this research stands synthetic biology, which provides tools to access the information flow inside cells by reprogramming DNA code [19]. Synthetic Biology techniques have a major drawback, namely, they necessitate the artificial modification of cells through addition of components (*e.g.*, magnetic nanoparticles) and/or synthetic genetic code, with the risk of impacting the natural cell functionalities and behaviors, often in unpredictable ways [5, 6].

In this paper, we advocate for an alternative solution based on the utilization of the natural mechanisms involved in the regulation of the cell's internal functionalities, *i.e.,* its metabolism, to realize the aforementioned goal. In particular, through these mechanisms, cells naturally adapt their biochemical behavior as a function of the chemical characteristics of the external environment, in the direction of maximum fitness to procure energy from the environment, grow, and reproduce [11]. While it is clear that by carefully tuning determinate characteristics of the environments where cells live, we have the ability to induce a specific behavior in a natural (non-engineered) cell, it is of utmost importance to understand the limits of this controllability.

Molecular Communication (MC), a recently developed discipline that studies systems where the exchange of information is realized through molecules and chemical reactions [1], is proposed in this paper as a tool to estimate the aforementioned limits. In particular, we abstract cell metabolism and its regulation as an

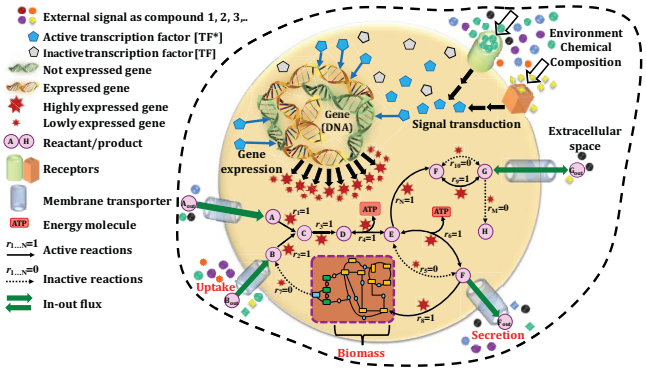

Figure 1: Schematic representation of cell metabolism and the regulation mechanisms considered in this paper.

end-to-end MC system, where information on the chemical composition of the external environment (transmitter in the end-to-end abstraction) is transmitted through a series of communication channels to determinate states in the cell's metabolism, and from these states to a specific behavior of the cell in terms of growth rate, and uptake and consumption of the chemical compounds in the environment (receiver in the end-to-end abstraction). On the basis of this abstraction, we apply information-theoretic tools to estimate an upper bound limit in terms of amount of information that can be channeled through this system. Thanks to a well-known and computationally efficient metabolic modeling and simulation technique, we are able to calculate this estimate from the knowledge of the cell DNA code, *i.e.*, genome, and the selection of the environmental factors that are under our control.

In prior work [15], we proposed a very first MC abstraction limited to a single communication channel that underlies the propagation of chemical information from the external environment to an internal cell metabolic state, and we derived proof-of-concept information-theoretic limits of this channel for a standard set of metabolic states in an *E. coli* organism. These theoretical limits were accounting for the amount of information that the cell is able to absorb from the environment, but not for the resulting changes in the cell's behavior. In this new contribution, we abstract the complete series of channels that account for how the environmental chemical information is affecting the external metabolic behavior of the cell, *i.e.*, the way the cell interacts with the environment, similarly to a complete transmitter-receiver system (end-to-end). In addition, to suggest a direct application of our study to real world problems, our new numerical results on the theoretical limits of these channels are now based on two main important human gut microbes [17], whose behavior has been linked to nutrition-related disorders such as obesity [10]. We also demonstrate that different compounds in the environment are associated with different amounts of information propagated by the end-to-end system.

The rest of this paper is organized as follows. In Section 2 we review basic concepts in the cell metabolism and its regulation, and present the end-to-end (E2E) MC system abstraction. In Section 3 we detail the information theoretic study to characterize this system in term so of steady-state mutual information. In Section 4 we discuss the data generation and relevant numerical results for the gut bacterium *B. theta* and archaeon *M. smithii*. The final section covers the conclusion and details of future avenues.

2 E2E MOLECULAR COMMUNICATION ABSTRACTION OF CELL METABOLISM

Cell metabolism is the set of chemical reactions responsible for the conversion of chemical compounds taken up from the environment into energy, cell building blocks, *i.e.*, biomass, and waste, for cell homeostasis, growth, and reproduction [13], as schematically shown in Fig. 1. We present some background material described in [15]. These chemical reactions, which transform reactant molecules into product molecules (A, B, C,... in Fig. 1), and are chained together into a complex metabolic network, are generally not spontaneous, but happen (are catalyzed) when in the presence of particular enzymes, which are special proteins synthesized (expressed) from genes written in the cells' DNA. Cells have the ability to control the rate of enzyme expression, therefore controlling the rate of the corresponding metabolic reactions, or flux, in transforming reactants into products. In particular, this control is operated by transcription factors, which are proteins that change their state between inactive [TF] and active [TF*] in function of environmental conditions (*e.g.*, presence of determinate chemical compounds in the extracellular environment) through a biochemical process named signal transduction. Depending on the particular transcription factor, a gene can be up-regulated (activation), therefore expressing the enzyme at a higher rate, or otherwise down-regulated (repression), which impacts the resulting metabolic flux of the corresponding catalyzed chemical reaction. In this paper, we consider a commonly accepted model of gene regulation, according to a logical approximation [2], where the relation between an enzyme-regulated metabolic reaction r_i and an active transcription factor $[TF^*]$ is represented by one of the following two expressions:

$$\begin{aligned} r_i &\simeq \beta H\left([TF^*] - K_d\right) \text{ if activation}, \\ r_i &\simeq \beta H\left(K_d - [TF^*]\right) \text{ if repression}; \end{aligned} \tag{1}$$

where β is the maximum expression level of the enzyme expressing $[TF]$, $H(.)$ is the Heaviside step function, equal to 1 when the argument is positive, and 0 otherwise, and K_d is the equilibrium constant of the reaction [18]. According to the logical approximation in (1), the enzyme expression, and the activity of the corresponding chemical reaction within the cell metabolism, can be either ON (= maximum enzyme expression rate and corresponding rate of the reaction) or OFF (= no expressed enzyme and absence of the corresponding chemical reaction in the cell metabolism) depending on the quantity of specific chemical compounds in the cell.

We abstract the aforementioned mechanisms underlying cell metabolism as an E2E MC system, as shown in Fig. 2, where through an information-centric approach. In our abstraction, we study the potential of cell metabolism and its regulation to propagate information from the external environment through the regulation of enzyme expression and the flux of metabolic reactions, resulting in the control of specific characteristics of cell's behavior, namely, the rate of uptake/secretion of chemical compounds to/from the environment and the rate of growth. In particular, with reference to Fig. 2, the proposed E2E MC system abstraction is composed of a series of two channels, namely, the **Enzyme Expression Regulation Channel** and the **Metabolic Reaction Network Channel**.

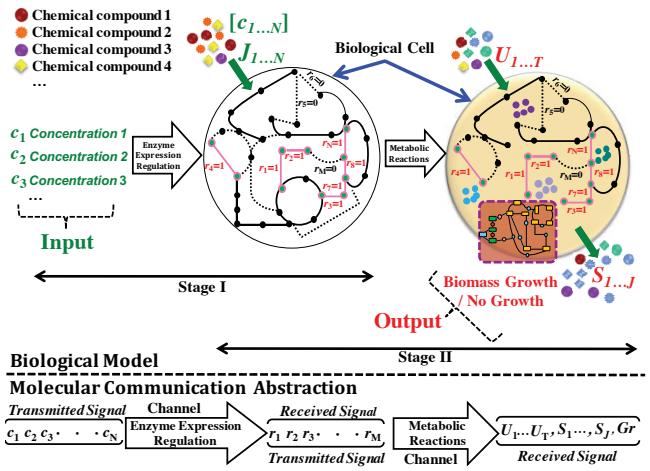

Figure 2: The proposed end-to-end molecular communication abstraction of cell metabolism.

When considering the **Enzyme Expression Regulation Channel**, identified as **Stage I** in Fig. 2, the transmitter abstracts the environment surrounding the cell, where the transmitted signal is the set of chemical compounds present in the environment that are input of the pathways that compose the cell metabolic network, the channel represents the mechanisms that regulate the expression of determinate enzymes as a function of the chemical compounds in input, and the receiver represents the cell metabolism, where the received signal is the resulting ON/OFF activity of the chemical reactions catalyzed by these enzymes. This abstraction is more formally expressed as $\{c_1, c_2, ...c_N\} \xrightarrow[\text{Regulation}]{\text{Enzyme Expression}} \{r_1, r_2, ...r_M\}$, where N is the number of chemical compounds present in the environment surrounding the cell and input of the metabolic pathway network, c_i is the concentration (number of molecules per unit volume) of the chemical compound $i = 1, ...N$, r_i is a binary value equal to 1 if the enzyme-expression-regulated reaction i is ON, and equal to 0 if the same reaction is OFF, as expressed in (1), M is the number of enzyme-expression-regulated reactions that change their state upon variations in the concentrations of input chemical compounds c_i, $i = 1, ...N$. A proof-of-concept analysis of this channel for *E. coli* cells is included in our previous publication [15].

When considering the **Metabolic Reaction Network Channel**, (**Stage II** in Fig. 2), the transmitter represents the cell itself, where the transmitted signal is the metabolic state of the cell, namely, the ON/OFF activity of the state-changing enzyme-regulated reactions. The channel represents mechanisms involved in the ON metabolic reactions that impact the consumption and production of chemical compounds within the cell. The receiver represents the cell's environment, where the received signal is the variation in the uptake and secretion of compound(s) from/to the environment itself, and the cell growth in terms of biomass production, respectively. This is more formally expressed as $\{r_1, r_2, ...r_M\} \xrightarrow[\text{Reactions}]{\text{Metabolic}} \left\{U_1, ...U_T, S_1, ..., S_J, Gr\right\}$, where $U_1...U_T$ are the fluxes, *i.e.,* the velocity of molecule concentration propagating in space (*e.g.,* from environment to cell), of metabolites taken up from the environment, $S_1...S_J$ are the fluxes of metabolites secreted by the cell into its environment, and *Growth (Gr)* represents the flux of produced compounds inside the cell that contribute to its growth (biomass).

In the rest of the paper, we detail an information-theoretic study to quantify the information flow in the E2E MC system. For this, we consider the model of Stage I from our previous work [15] along with the newly proposed Stage II model. Although the aforementioned biological processes are realized in nature by the interplay of signal transduction, gene regulation and cell metabolism, the interactions among these different mechanisms are still only partially understood [8]. This is especially true for the organisms consider, namely, the *B. theta* and the *M. smithii*. As a consequence, the study in this paper is based on a black-box abstraction of signal transduction and gene regulation, and focuses on the mechanisms underlying chemical reactions and their fluxes in cell metabolism.

3 INFORMATION-THEORETIC STUDY OF THE E2E MC METABOLIC CHANNELS

To quantify the potential of cell metabolism and its regulation to be utilized to control cell behavior, we define the steady-state mutual information I of the two stages of the aforementioned MC abstraction, as well as for the overall E2E MC system. In particular, we consider the steady-state case since our methodology is currently based on constraint-based approach, the Flux Balance Analysis (FBA), as explained later in the section, which estimates the optimal cell metabolic state and chemical reaction fluxes without taking into account the dynamic behavior of the enzyme expression regulation or the metabolic reaction network. As detailed later in and in [15], the use of FBA data in our expressions results in the computation of an upper-bound of the mutual information.

The steady-state mutual information I of Stage I quantifies the amount of information about the chemical composition of the cell's environment measured in bits that a cell is able to represent in the binary state of its enzyme-expression-regulated metabolic reactions at steady state, after any evolution of the enzyme-expression regulation channel. As detailed in [15], the steady-state mutual information for Stage I is defined as follows:

$$I(\{c_i\}_{i=1}^N; \{r_i\}_{i=1}^M) = H(\{c_i\}_{i=1}^N) - H(\{c_i\}_{i=1}^N \mid \{r_i\}_{i=1}^M), \quad (2)$$

where the expressions for the input entropy $H(\{c_i\}_{i=1}^N)$ and the conditional entropy of the input given the output $H(\{c_i\}_{i=1}^N \mid \{r_i\}_{i=1}^M)$ can be found in [15].

The steady-state mutual information I for the Stage II quantifies the amount of information contained in the internal binary cell metabolic state that can be perceived from the outside environment through the metabolic-state-modulated values of the fluxes of metabolites taken up or secreted, and biomass (growth), after any evolution in time of the fluxes of consumed or produced metabolites accounting the rate of metabolic reactions to be negligible with respect to cell growth. We defined this as

$$I(\{r_i\}_{i=1}^M; \{U_t\}_{t=1}^T, \left\{S_j\right\}_{j=1}^J, Gr) = \quad (3)$$
$$H((\{r_i\}_{i=1}^M) - H(\{r_i\}_{i=1}^M \mid \{U_t\}_{t=1}^T, \left\{S_j\right\}_{j=1}^J, Gr),$$

where the input entropy $H(\{r_i\}_{i=1}^M)$ is defined as

$$H(\{r_i\}_{i=1}^M) = -\sum_{k=1}^{K} P\left(\left\{r_{i_k}\right\}_{i=1}^M\right) \log_2 P\left(\left\{r_{i_k}\right\}_{i=1}^M\right), \quad (4)$$

where K is equal to the total number of different combinations of binary values of the enzyme-expression-regulated chemical reactions at the output of the system $\left[\{r_i\}_{i=1}^{M}\right]_k$ resulting from the all the possible values that the input chemical compound concentrations $\{c_i\}_{i=1}^{N}$ can assume, and $P(.)$ is their probability distribution.

The conditional entropy of the input given the output $H(\{r_i\}_{i=1}^{M} \mid \{U_t\}_{t=1}^{T} \left\{S_j\right\}_{j=1}^{J}, Gr)$ is then defined as follows:

$$H(\{r_i\}_{i=1}^{M} \mid \{U_t\}_{t=1}^{T}, \left\{S_j\right\}_{j=1}^{J}, Gr) = \tag{5}$$
$$-\sum_{q=1}^{Q} P\left(\left[\{U_t\}_{t=1}^{T}, \left\{S_j\right\}_{j=1}^{J}, Gr\right]_q\right)$$
$$\sum_{k=1}^{K} P\left(\left\{r_{i_k}\right\}_{i=1}^{M} \middle| \left[\{U_t\}_{t=1}^{T}, \left\{S_j\right\}_{j=1}^{J}, Gr\right]_q\right)$$
$$\log_2 P\left(\left\{r_{i_k}\right\}_{i=1}^{M} \middle| \left[\{U_t\}_{t=1}^{T}, \left\{S_j\right\}_{j=1}^{J}, Gr\right]_q\right);$$

where Q is equal to the total number of different sets of flux values at the output of the system $\{U_i\}_{i=1}^{t}, \{S_i\}_{i=1}^{j}, Growth\ (Gr)$ resulting from different combinations of the the enzyme-expression-regulated reactions, and $P(.)$ is their probability distribution [16].

To provide an *in silico* evaluation of the expressions in (2) and (3) for a specific cell, we make use of the aforementioned FBA, which is a well-known and computationally efficient mathematical method that allows us to simulate cell metabolism and its regulation through linear optimization techniques given the cell's genetic code (genome) and determinate constraints [14]. Through FBA we estimate the chemical reactions that might be active in cell metabolism given the set of compounds (and their concentration and maximum flux) present in the environment [3, 14]. In particular, through FBA we are able to obtain an estimate of the state $\left\{r_i^*\right\}_{i=1}^{M}$ of the enzyme-expression-regulated chemical reactions that results into an overall maximum biomass production, and the consequent flux values of the metabolic reactions in the cell, including those at the output, *i.e.*, $\left\{U_i^*\right\}_{i=1}^{t}, \left\{S_i^*\right\}_{i=1}^{j}, (Gr)^*$. Further details about the theory behind FBA and the step-by-step estimation are detailed in [15].

Given the optimal estimates of the chemical reaction states $\left\{r_i^*\right\}_{i=1}^{M}$ obtained through the FBA from the knowledge of the cell's genome for all the values that our input set of chemical compound concentrations $\{c_i\}_{i=1}^{N}$ can assume, we can compute the following upper bound steady-state mutual information for Stage I [15]:

$$I(\{c_i\}_{i=1}^{N}; \left\{r_i^*\right\}_{i=1}^{M}) = H(\{c_i\}_{i=1}^{N}) - H(\{c_i\}_{i=1}^{N} \mid \left\{r_i^*\right\}_{i=1}^{M}), \tag{6}$$

The relation between steady state mutual information and its upper bound can be formalized as follows [15]:

$$I(\{c_i\}_{i=1}^{N}; \left\{r_i^*\right\}_{i=1}^{M}) \geq I(\{c_i\}_{i=1}^{N}; \{r_i\}_{i=1}^{M}), \tag{7}$$

Consequently, we compute the steady-state mutual information for Stage II with the values of the chemical reaction states and flux values at the output of the system as follows:

$$I(\left\{r_i^*\right\}_{i=1}^{M}; \{U_t^*\}_{t=1}^{T}, \left\{S_j^*\right\}_{j=1}^{J}, Gr^*), \tag{8}$$

Finally, the upper bound to the E2E steady-state mutual information $I(\{c_i\}_{i=1}^{N}; \left\{U_t^*\right\}_{t=1}^{T}, \left\{S_j^*\right\}_{j=1}^{J}, Gr^*)$ is computed through (4) by substituting $\{r_i\}_{i=1}^{M}$ with $\{c_i\}_{i=1}^{N}$ and summation with integration, and the expression in (5) with the same variable substitution, and the second summation with integration.

4 DATA GENERATION AND NUMERICAL RESULTS

In this section, we present a numerical example of the information-theoretic study of the information flow in the E2E MC abstraction of cell metabolism proposed in this paper. This example is based on two microbial organisms present in the human gut, namely, the *B. theta* bacterium and *M. smithii* archeon, which are the subject of cutting-edge experimental research by co-authors of this paper and by others [12, 17]. The control of the metabolic behavior of these organisms has particular significance since changes in the abundance of intestinal *B. theta* and *M. smithii* have been linked to nutrition-related disorders such as obesity [10]. In recent work [4], we examined the correlation between changes in the growth of these organisms and the chemical composition of the environment with sampling and inference techniques from software testing, but without the communication-centric abstraction of the underlying biochemical processes presented in this paper.

Through the quantification of the information flow for these two organisms, we can infer how variations in the chemical composition of the environment can impact their metabolic behavior. *In the following, we demonstrate that this study can lead to a ranking of chemical compounds on the basis of how much mutual information through the E2E MC system can result from their variation. This can potentially lead to the reduction in the number of expensive experiments by reducing the different combinations of chemical compounds necessary to obtain and study determinate metabolic behaviors of these organisms and emergent behaviors from their metabolic interactions, which might be associated with dysbiosis.* In the long run, by quantifying their metabolic controllability through our framework, we may enhance knowledge of how organisms and the environment interact through genomic information to produce cell behaviors that ultimately have an impact on human health.

To generate data on the metabolism of the aforementioned organisms, we utilize a bioinformatics modeling pipeline and the implementation of the FBA algorithm within KBase (Department of Energy Systems Biology Knowledgebase) software application suite [9]. Given the *B. theta* and *M. smithii* genomes [7, 20], and the concentrations of compounds in the environment, we undergo a series of steps where the genome is used to construct a Genome Scale Metabolic (GEM) model, which is then at the base of the FBA optimization algorithm to estimate the resulting biomass, uptake, secretion, and all the optimized reactions in the organism pertaining to the known genome as in Fig. 3. We then use code that we developed in Matlab and Python to analyze the results and compute the expressions presented in Section 3. We provide a detailed description of this procedure in [15], where we generate data to study the Stage I of the MC system for the *E. coli* bacterium.

The Stage I upper-bound of the steady state mutual information of *B. theta* and *M. smithii* was computed by using 128 binary combinations (present/not present) of seven input compounds, *i.e.,*

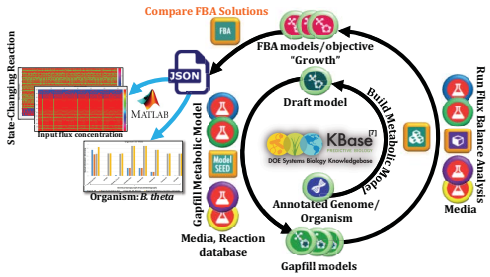

Figure 3: KBase workflow to generate *in silico* data.

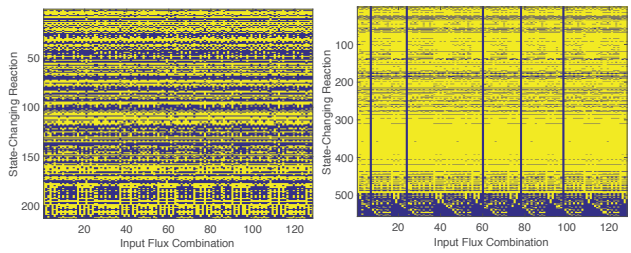

Figure 4: (left)-variations in state changing reactions for 128 input combinations for *B. theta*. (right)-variations in state changing reactions for 128 input combinations for *M. smithii*

Carbon-D-Glucose (G), Hematin (He), Formate (F), H_2, Vitamin B_{12} (B_{12}), Acetate (A), and Vitamin K (V_k). In this paper, we limited the possible input compounds to seven in order to obtain a ground truth for our study as it is a resource limiting factor if we want to move the experiments to an *in vitro* environment in the future. The results obtained for Stage I for *B. theta* and *M. smithii* are shown in Fig. 4, where the state-changing metabolic reactions in the *B. theta* and *M. smithii*, respectively, are set to ON (yellow) or OFF (blue) according to the FBA-estimated chemical reaction state, for each of the aforementioned 128 binary input flux combinations. Based on these data, we obtain the upper-bound of the steady-state mutual information of the *B. theta* Stage I using (6): $I\left(\left\{c_G, c_{He}, c_F, c_{H_2}, c_{B_{12}}, c_A, c_{V_k}\right\}; \left\{r_i^*\right\}_{i=1}^{113}\right) = 3.3$ bits. Similarly, we compute the upper-bound of the steady-state mutual information of the *M. smithii* Stage I, resulting in 4.5 bits. More details of these calculations are presented in [16]. $\left\{r_i^*\right\}_{i=1}^{113}$ and $\left\{r_i^*\right\}_{i=1}^{136}$ represent upper-bound enzyme-expression regulated unique chemical reactions for *B. theta* and *M. smithii* respective, derived from Fig. 4 (left) and Fig. 4 (right). These values represent an estimate of the maximum information that the each cell's metabolism can absorb from the chemical compounds present in the environment.

In Stage II we applied our method to compute the potential of each organism to translate changes in the internal metabolic state into differences in its interactions with the environment. To do this, we grouped the FBAs based on the similar FBA-estimated chemical reaction states, giving us 14 groups for *B. theta* and 31 groups for *M. smithii*. We first compute the upper bound of the steady-state mutual information over internal metabolic state changing reactions with respect to biomass only, resulting in 2.8 bits for *B. theta* and 3.8 bits for *M. smithii*, respectively. These values represent an estimate of the information that can flow from the state of the metabolic reactions to the output, and quantifies the controllability of these organisms in terms of rate of growth (or biomass). Furthermore, we compute the upper bound of steady-state mutual information over internal metabolic state changing reactions with respect to uptake and secretion of compounds and biomass, resulting in 3.7 bits and 3.9 bits for *B. theta* and *M. smithii* respectively. These values quantify the same information flow estimate, this time related to the joint controllability of the organisms in terms of rate of growth (biomass) and fluxes of compounds secreted or taken up.

The upper bound of the steady-state mutual information of the overall E2E MC system, from the input environmental compounds to the compound uptake/secretion and biomass, resulted in 2.7 bits for *B. theta* and 2.6 bits for *M. smithii*, respectively. These values represent an estimate of the information flow from the two aforementioned MC channels in cell metabolism, and quantify the overall maximum controllability of these organisms in terms of growth and uptake/secretion of compounds that can be operated from the external environment by a tuning the chemical composition (each of the 7 chemical compound as present/not present) of the environment. Stage II values are computed using (8). A summary of all the calculations is presented in Table 1.

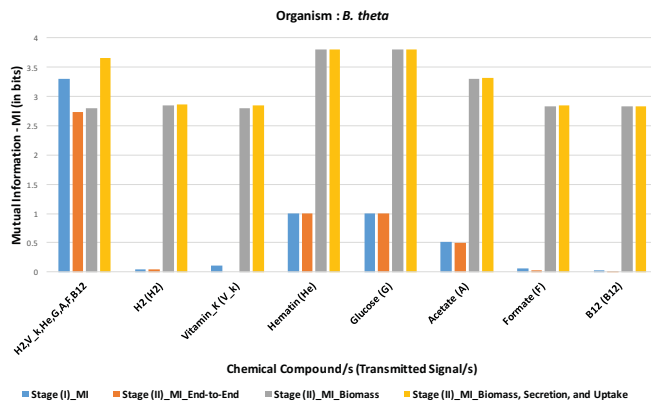

Figure 5: Upper bound to the steady-state mutual information of *B. theta*, one input compound vs. 7 compounds.

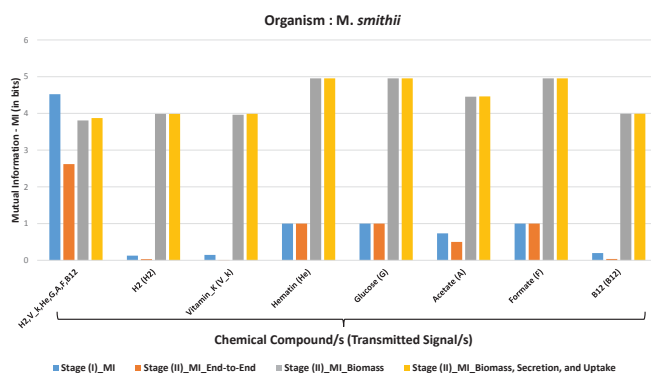

Figure 6: Same as in Fig. 5 for *M. smithii*.

To demonstrate an immediate practical application of the study presented in this paper, we show in Fig. 5 and Fig. 6 the calculation of the upper bound to the steady-state mutual information in the case where we can control the presence/no presence of only one input compound at a time in the environment, while the presence/no presence of other compounds is unknown. This calculation is performed for each of the seven compounds for Stage I, Stage II (with

Table 1: Summary of the Upper Bound of the Steady-State Mutual Information (UBSSMI) presented in this paper.

7 bit - UBSSMI - STAGE (I)					
Organisms	$H(\{c_i\}_{i=1}^{N})$		$H(\{c_i\}_{i=1}^{N} \mid \{r_i^*\}_{i=1}^{M})$		$I(\{c_i\}_{i=1}^{N}; \{r_i^*\}_{i=1}^{M})$
B. theta	$log_2(128) = 7$ bits	(N = 7)	3.7 bits	(M = 113)	3.3 bits
M. smithii	$log_2(128) = 7$ bits	(N = 7)	2.5 bits	(M = 136)	4.5 bits
7 bit - UBSSMI OVER INTERNAL METABOLIC STATE REACTIONS WITH RESPECT TO ONLY BIOMASS – STAGE (II)					
Organisms	$H(\{r_i^*\}_{i=1}^{M})$		$H(\{r_i^*\}_{i=1}^{M} \mid Gr^*)$		$I(\{r_i^*\}_{i=1}^{M}; Gr^*)$
B. theta	$log_2(14) = 3.8$ bits	(M = 14)	1 bits	(M = 8)	2.8 bits
M. smithii	$log_2(31) = 4.9$ bits	(M = 31)	1.1 bits	(M = 15)	3.8 bits
7 bit - UBSSMI OVER INTERNAL METABOLIC STATE REACTIONS WITH RESPECT TO UPTAKE AND SECRETION OF COMPOUNDS AND BIOMASS - STAGE (II)					
Organisms	$H(\{r_i^*\}_{i=1}^{M})$		$H\left(\{r_i^*\}_{i=1}^{M} \mid \{U_t^*\}_{t=1}^{T}, \{S_j^*\}_{j=1}^{J}, Gr^*\right)$		$I\left(\{r_i^*\}_{i=1}^{M}; \{U_t^*\}_{t=1}^{T}, \{S_j^*\}_{j=1}^{J}, Gr^*\right)$
B. theta	$log_2(14) = 3.8$ bits	(M = 14)	0.1 bits	(M = 13)	3.7 bits
M. smithii	$log_2(31) = 4.9$ bits	(M = 31)	1 bits	(M = 16)	3.9 bits
7 bit - UBSSMI SEVEN INPUT COMPOUNDS WITH RESPECT TO UPTAKE AND SECRETION OF COMPOUNDS AND BIOMASS – E2E					
Organisms	$H(\{c_i\}_{i=1}^{N})$		$H\left(\{c_i\}_{i=1}^{N} \mid \{U_t^*\}_{t=1}^{T}, \{S_j^*\}_{j=1}^{J}, Gr^*\right)$		$I\left(\{c_i\}_{i=1}^{N}; \{U_t^*\}_{t=1}^{T}, \{S_j^*\}_{j=1}^{J}, Gr^*\right)$
B. theta	$log_2(128) = 7$ bits	(N = 7)	4.3 bits	(M = 8)	2.7 bits
M. smithii	$log_2(128) = 7$bits	(N = 7)	4.4 bits	(M = 15)	2.6 bits

output biomass only or biomass+uptake+secretion), and the E2E system. All these values are compared with the aforementioned values where the presence/absence can be controlled for all the 7 input compounds. Fig. 5 and Fig. 6 reveal that it is possible to rank the input chemical compounds on the basis of how much mutual information can result from their variation for the different stages in *B. theta* and *M. smithii*, respectively. As a consequence, we notice that Hematin and Glucose for *B. theta* and Hematin, Glucose, and Formate for *M. smithii* are compounds that maximally affect the metabolic behavior of each organism, and in these cases particularly affect their biomass production rates, since the inclusion of uptake or secretion as additional output does not affect the mutual information values. As mentioned above, the E2E MC approach is consistent with results obtained from BioSIMP [4], which used software testing methods to identify the environmental nutritional factors required for growth of *B. theta* and *M. smithii*. The E2E MC method predicts a range of 2.6-4.5 bits of molecular information that can be transmitted and received for each cell, which agrees with the BioSIMP prediction of 3-5 factor sampling to cover 90% of the possible metabolic states of the cell.

5 CONCLUSIONS

In this paper, we have proposed a methodology to quantify the limits in the controllability of the metabolic behavior of a natural (non-engineered) biological cell by varying the chemical composition of its environment, where this behavior is represented by the rates of cell's growth, consumption of chemical compounds from the environment, and secretion of metabolic byproducts. This methodology is based on the abstraction of molecular communication and mathematical tools from information theory. In particular, we have abstracted cell metabolism, and its regulation, as an E2E communication system where information about the chemical composition of the environment is propagated by a sequence of two channels, namely, the enzyme expression regulation and the metabolic reaction network. Through the parameter of mutual information we have quantified the limits of these channels in terms of amount of information that they can propagate, as well as for the E2E system. We have applied our methodology to two important microorganisms present in the human gut, whose behavior has been liked to obesity. Our results demonstrate that different chemical compounds present in the environment are associated with different amounts of information propagated through the aforementioned channels, and therefore their variations result in different consequences on cell behavior. In future work we plan to realize a complete communication model of these channels, by taking into account their dynamic behaviors and associated noise.

ACKNOWLEDGMENTS

This work was supported by the US NSF through grants MCB-1449014, EPSCoR EPS-1004094, and CCF-1161767, and the NIH through grant 1-P20-GM113126-01. We also thank Aditya Immaneni for providing constrictive feedback for the data analysis and coding.

REFERENCES

[1] I F Akyildiz, M Pierobon, S Balasubramaniam, and Y Koucheryavy. 2015. The internet of bio-nano things. *IEEE Communications Magazine* 53, 3 (2015), 32–40.

[2] U Alon. 2006. *An Introduction to Systems Biology - Design Principles of Biological Circuits.* Chapman & Hall/CRC.

[3] G JE Baart and D E Martens. 2012. Genome-scale metabolic models: reconstruction and analysis. *Neisseria meningitidis: Advanced Methods and Protocols* (2012), 107–126.

[4] M Cashman, J L Catlett, M B Cohen, N R Buan, Z Sakkaff, M Pierobon, and C A Kelley. 2017. BioSIMP: Using Software Testing Techniques for Sampling and Inference in Biological Organisms. *ICSE workshop on Software Engineering for Computational Science (SE4S)* (2017).

[5] M B Cohen, J Firestone, and M Pierobon. 2016. The Assurance Timeline: Building Assurance Cases for Synthetic Biology. In *In Proceedings of the 4th International Workshop on Assurance Cases for Software-Intensive Systems (ASSURE 2016).* 75–86.

[6] J Dobson. 2008. Remote control of cellular behaviour with magnetic nanoparticles. *Nature nanotechnology* 3, 3 (2008), 139–143.

[7] B Dridi, M Henry, A El Khechine, D Raoult, and M Drancourt. 2009. High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PloS one* 4, 9 (2009), e7063.

[8] E Gonçalves, J Bucher, A Ryll, J Niklas, K Mauch, S Klamt, M Rocha, and J Saez-Rodriguez. 2013. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Molecular BioSystems* 9, 7 (2013), 1576–1583.

[9] KBase. 2011. Department of Energy Systems Biology Knowledgebase (KBase). http://kbase.us. (2011). [Online;accessed 7-April-2017].

[10] R Krajmalnik-Brown, Z-E Ilhan, D-W Kang, and J K DiBaise. 2012. Effects of gut microbes on nutrient absorption and energy regulation. *Nutrition in Clinical Practice* 27, 2 (2012), 201–214.

[11] C M Metallo and M G Vander Heiden. 2013. Understanding Metabolic Regulation and Its Influence on Cell Physiology. *Molecular Cell* 49 (February 2013), 388–398.

[12] D Ndeh, A Rogowski, A Cartmell, A S Luis, A Baslé, J Gray, I Venditto, J Briggs, X Zhang, A Labourel, and et al. 2017. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544, 7648 (2017), 65–70.

[13] D L Nelson and M M Cox. 2005. *Lehninger Principles of Biochemistry.* W. H. Freeman, Chapter 12.2, 425–429.

[14] J D Orth, I Thiele, and B Ø Palsson. 2010. What is flux balance analysis? *Nature biotechnology* 28, 3 (2010), 245–248.

[15] M Pierobon, Z Sakkaff, J L Catlett, and N R Buan. 2016. Mutual information upper bound of molecular communication based on cell metabolism. In *Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on.* IEEE, 1–6.

[16] Z Sakkaff. 2016. *Characterization of Molecular Communication Based on Cell Metabolism Through Mutual Information and Flux Balance Analysis.* Master's thesis. University of Nebraska-Lincoln.

[17] A A Salyers. 1984. Bacteroides of the human lower intestinal tract. *Annual Reviews in Microbiology* 38, 1 (1984), 293–313.

[18] G Tkačik, C G Callan Jr., and W Bialek. 2008. Information capacity of genetic regulatory elements. *PHYSICAL REVIEW E* 78 (2008), 011910(17).

[19] J E Toettcher, C A Voigt, Orion D Weiner, and Wendell A Lim. 2011. The promise of optogenetics in cell biology: interrogating molecular circuits in space and time. *Nature methods* 8, 1 (2011), 35–38.

[20] J Xu, M K Bjursell, J Himrod, S Deng, L K Carmichael, H C Chiang, L V Hooper, and J I Gordon. 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299, 5615 (2003), 2074–2076.