

# Information- and Communication-Centric Approach in Cell Metabolism for Analyzing Behavior of Microbial Communities

Zahmeeth Sakkaff<sup>1,2,\*</sup> Andrew Freiburger<sup>1,3</sup>, Nidhi Gupta,<sup>1,4</sup> Massimiliano Pierobon,<sup>1,2,+</sup> and Christopher S. Henry<sup>1,3</sup>

<sup>1</sup> Argonne National Laboratory, Lemont, IL, USA

<sup>2</sup> Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE, USA

<sup>3</sup> University of Chicago, Chicago, IL, USA

<sup>4</sup>Corresponding authors: pierobon@cse.unl.edu and chenry@anl.gov

## Abstract

Microorganisms naturally form community ecosystems to improve fitness in diverse environments and conduct otherwise intractable processes. Microbial communities are therefore central to biogeochemical cycling, human health, agricultural productivity, and technologies as nuanced as nanotechnology-enabled devices; however, the combinatorial scaling of exchanges with the environment that predicate community functions are experimentally untenable. Several computational tools have been presented to capture these exchanges, yet, no attempt has been made to understand the total information flow to a community from its environment. We therefore adapted a recently developed model for singular organisms, which blends molecular communication and the Shannon Information theory to quantify information flow, to communities and exemplify this expanded model on idealized communities: one of *Escherichia coli* (*E. coli*) and *Pseudomonas fluorescens* to emulate an ecological community and the other of *Bacteroides thetaiotaomicron* (*B. theta*) and *Klebsiella* to emulate a human microbiome interaction. Each of these sample communities exhibit critical syntropy in certain environmental conditions, which should be evident through our community mutual information model. We further explored alternative frameworks for constructing community genome-scale metabolic models (GEMs) – mixed-bag and compartmentalized. Our study revealed that information flow is greater through communities than isolated models, and that the mixed-bag framework conducts greater information flow than the compartmentalized framework for community GEMs, presumably because the latter is encumbered with transport reactions that are absent in the former. This community Mutual Information model is furthermore wrapped as a KBase Application (*Run Flux Mutual Information Analysis, RFMIA*) for optimum accessibility to biological investigators. We anticipate that this unique quantitative approach to consider information flow through metabolic systems will accelerate both basic and applied discovery in diverse biological fields.

## Author Summary

### Author summary

Microorganisms frequently communicate information via information-bearing molecules, which must be fundamentally understood to engineer biological cells that properly engage with their environments, such as the envisioned Internet of Bio-NanoThings. The study of these molecular communications has employed information and communication theory to analyze the exchanged information via chemical reactions and molecular transport. We introduce an information- and communication-centric computational approach to estimate the information flow in biological cells and its impacts on the behavior of single organisms and communities. This study complements our previous work of cell metabolism by developing an end-to-end perspective of molecular communication based on enzyme-regulated reactions. We explore the mutual information using Shannon information theory, measured in bits, between influential nutrients and cellular growth rate. The developed *RFMIA* computational tool is deployed in the U.S. Department of Energy's Systems Biology Knowledgebase, where it quantitatively estimates information flow in both organism and community metabolic networks and extends recent developments in computer communications to explore and explain a new biology for the open-source community.

## Introduction

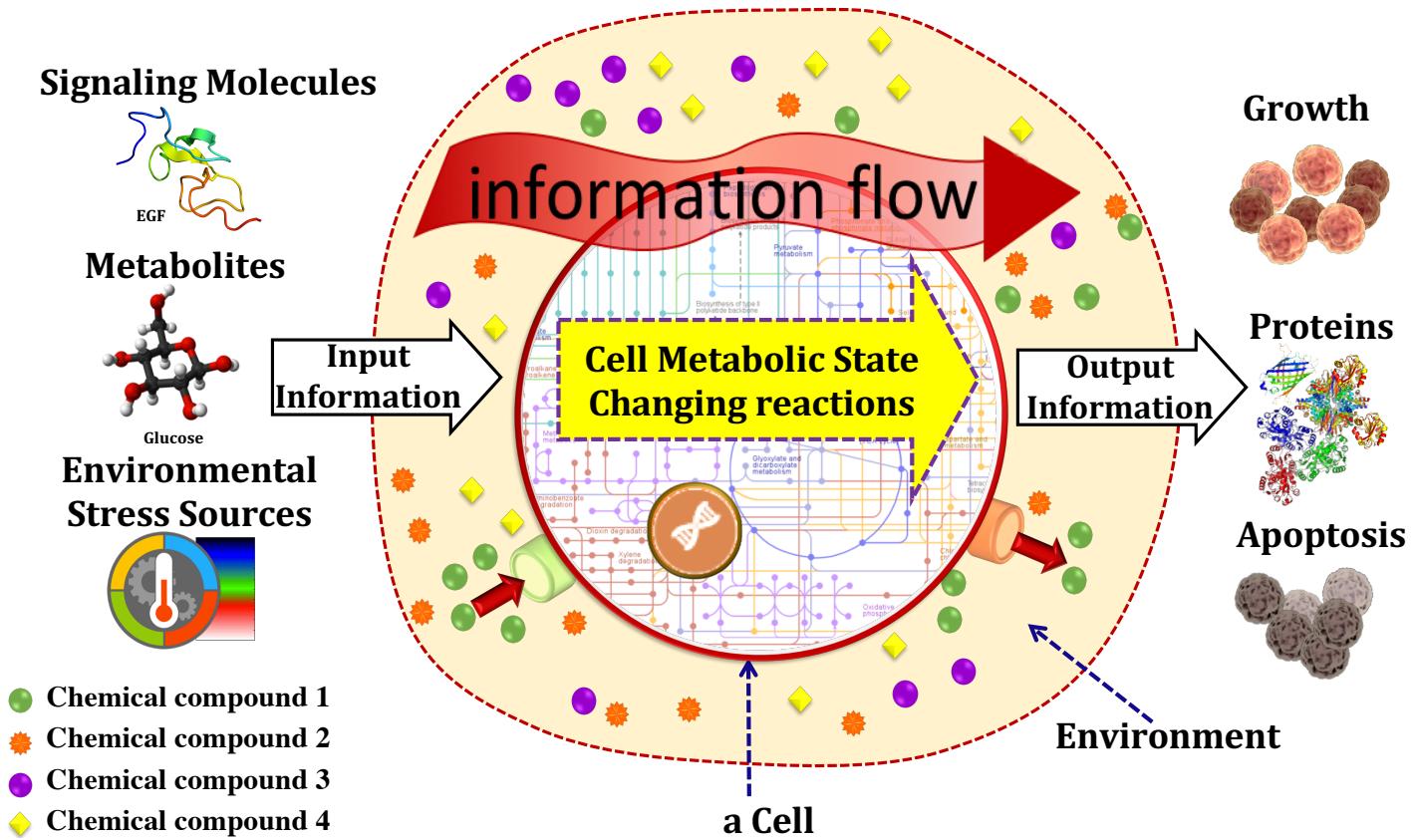
Microorganisms naturally assemble into communities to diversify strengths and weakness and to conduct complex functions such as biogeochemical cycling, intricate bioproduction, and digesting intractable nutrients. Cellular metabolism is the

foundation of these transformative processes, where chemical reactions convert environmental substrates into energy and biomass in a tightly regulated pipeline [1] (Fig. 1). These reactions are activated by the cell's environment to optimally survive, often meaning growth and reproduction [2]. Environmental perturbations may therefore be useful to control cellular behaviors, although the limits of this approach and the communication mechanisms by which cells and communities coordinate functions in response to environmental triggers remain opaque. The specific mediating molecules (e.g. quorum sensing) and triggers for these communication mechanisms represent a large knowledge gap [3, 4] around cellular communication. Community communications especially remain understudied [5], partly due to the difficulty in recovering representative samples [6]; hence computational biology is an attractive alternative domain to study communities without necessitating a complete experimental description of a system. Some computational models apply top-down ecological principles to represent microbial communities while other models represent communities from the bottom-up assembly of metabolic pathways and fundamental biochemistry [7]. Molecular communication (MC) in Fig. 1 is an ideal blend of both approaches – while also leveraging mathematics, computer science, and chemical engineering – to fully capture the multi-dimensional complexity of community interactions by deducing information flow from chemical exchanges. We previously derived an MC abstraction of a single communication channel of interdependent inputs and outputs [8], where information propagates from media into cellular metabolism and where we derived quantitative limits for this information flow [9], and for a two molecular communication channels for each community [10]: one where regulatory mechanisms depend upon combinations of extracellular compounds, and the other where cell metabolism depends on growth and extracellular exchange. The application of information theory in these papers (illustrated in Fig. 2) expresses the performance of both channels and examined the end-to-end (E2E) limits to molecular communication system, such as exchange flux and biomass growth.

Herein, we expand our previously defined molecular communication model into a Mutual Information score (MI) that is crafted to examine community systems. with different combinations of seven substrates for two human gut microbes – an obligate anaerobe bacterium *B. theta* and an Archaea *M. smithii*, where the substrates were found to propagate different amounts of information [?]. The MI score uniquely 1) characterizes cellular and inter-species molecule exchanges, 2) detects extracellular signatures – exchange fluxes and biomass growth – that indicate intra-cellular information flow, and 3) generates data that can reveal community interactions and thereby minimize resource-intensive experiments to elucidate community communications and their sensitivity to environmental perturbations. MI in Fig. 2 leverages Shannon information theory [11, 12] to quantify variability in metabolic information flow and steady-state FBA [13] to define an upper limit of information flow from the given environment based on metabolic flux, biomass growth, and E2E (uptake, secretion of metabolites and biomass). Shannon's theory defines information as the details that distinguish a state of a system from the space of possible system states, which is depicted in Fig. 3. The MI score is the uncertainty difference between the uncertainty contained in the input molecule(s) and the conditional uncertainty of the channel output, and thereby quantifies information change via each channel. We exemplify the MI score to understand an idealized microbiome community of *Escherichia coli* (*E. coli*) and *Bacteroides thetaiotaomicron* (*B. theta*) that has a prominent role on our adaptive immunity [14] and biotic health [15], and which exhibits a unique community synergy of *B. theta* only growing in aerobic environments when coupled with *E. coli* [16]. We moreover compared the information differences between two alternative frameworks for defining community GEMs: 1) the mixed-bag framework, where all member networks are pooled into a single network; and (2) the compartmentalized framework, where each member network is segregated into a designated compartment that can exchange with a common environment compartment. We consistently observe a higher MI score in communities, especially the mixed-bag abstraction, relative to isolates, which is qualitatively expected yet quantitatively invaluable when comparing different community designs or ecological factors that govern community formation. The MI score therefore has far-reaching implications in diverse biological fields and should accelerate engineering efforts of controlling microbiomes or synthesizing communities. The MI score is available as a point-and-click Application (*Run Flux Mutual Information Analysis, RFMIA* in the U.S. Department of Energy's Systems Biology Knowledgebase (KBase) [17] where genomes can be constructed into metabolic networks and ultimately merged community models [18] (See the example KBase Narrative: <https://kbase.us/n/40576/330/>).

## Materials and Methods

Mutual Information ( $I$ ) is defined as the input uncertainty minus the output uncertainty that depends on the inputs:  $I = H(inputs) - H(inputs, outputs)$ . The MI was applied to explore how much environmental perturbations ( $inputs = \{c_i\}_{i=1}^N$ , as the concentration of all  $N$  examined substrates) can control metabolic flux and biomass growth ( $outputs = (\{F_e\}_{e=1}^E, Gr)$ , as all exchange fluxes and the biomass growth flux [10], respectively). The MI for the overall E2E channel quantifies information flow that can be perceived from the environment through exchange fluxes and biomass growth



**Fig 1. Molecular communication systems send and receive information through chemical reactions.** Information flows from the cell's environment in the form of input signals such as signaling molecules, metabolites, environmental stress sources like temperature, and pH. Output information flows from the metabolic state and how information of the internal cell state can be perceived from the outside environment in the form of growth of a cell, proteins and apoptosis of a cell.

### Upper Bound to the Steady-State Mutual Information – UBS-sMI

FBA is employed to computationally evaluate (1) and (3), which yields the profile of activated reactions in a given the environment and enables estimating the state  $\{r_i^*\}_{i=1}^M$  that maximizes biomass production and the associated fluxes *i.e.*,  $\{U_i^*\}_{i=1}^t, \{S_i^*\}_{i=1}^j, (Gr)^*$ . Parsimonious FBA (pFBA) [19] is furthermore applied to reduce the space of optimal solutions to those with the lowest total flux, thereby imposing a secondary optimization of metabolic efficiency while improving reproducibility of the predicted flux profiles. The upper bound for *MI* of metabolic flux [9]:

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

where  $\{c_i\}_{i=1}^N$  is the initial concentrations. The relations between *MI* and its upper bound

$$I(\{c_i\}_{i=1}^N; \{r_i^*\}_{i=1}^M) \geq I(\{c_i\}_{i=1}^N; \{r_i\}_{i=1}^M). \quad (2)$$

and *MI* of biomass growth and its upper bound

$$I(\{r_i^*\}_{i=1}^M; \{U_t^*\}_{t=1}^T, \{S_j^*\}_{j=1}^J, Gr^*). \quad (3)$$

are analogously derived. The upper bound of the E2E steady-state mutual information  $I(\{c_i\}_{i=1}^N; \{U_t^*\}_{t=1}^T, \{S_j^*\}_{j=1}^J, Gr^*)$  is solved by substituting  $\{r_i\}_{i=1}^M$  with  $\{c_i\}_{i=1}^N$  and then integrating. The mathematical formulation is presented in supplementary S1.

## Reconstruction of Community Models

We sourced published GEMs for *E. coli*/**iML1515** [20] and *B. theta*/**iAH991** [21] that offer the most complete and experimentally-validated reconstructions to-date for these organisms. These member GEMs were assembled to a mixed-bag community model – the pooled amalgamation of all reactions – and as a compartmentalized community model – each GEM is assigned a unique compartment and all compartments exchange with a common extracellular compartment – which are contrasted in Fig. 4. These community model frameworks are functionally identical in FBA, with the difference that transport reactions are not present in mixed-bag community models since they lack compartments. The **Merge metabolic models into community model** tool in KBase constructed each version of community model for the two-member community.

## Mutual Information score

The MI is a metric in units of bits for the metabolic ability to process information from a given environment, where the activity of state-changing chemical reactions are either activated (1) or inactive (0) depending on the media. The MI score is proportional with information flow between the cell metabolism and its environment.

## Results

### Integration of the Mutual Information Approach in KBase

The *RFMIA* KBase Application in Fig. 6 executes the MI score, where the user-friendly workflow in Fig. 5 accepts inputs of the models, media components, a *mmol* of carbon consumption limit, and model objectives and returns tables and charts that succinctly communicate information flows as well as rank compounds based on their influence on information flow. Steps 1–5 in Fig. 5a,b explain the backend processes of the RFMIA KBase App.

### Integration of Published *E. coli* and *B. theta* Models and Combination into Community models

The FBA simulation results, GEM metadata, and growth rates for aerobic and anaerobic conditions among each of the seven input nutrients are tabulated in Table 1 (corollary content is available in **S3** and Narrative <https://kbase.us/n/81575/184/>). The published individual models were subtly curated with missing pathways that are necessary for our simulation: *E. coli* was curated to produce dextrin from starch and sucrose and *B. theta*/**iAH991V2** was curated by adding choline exchange and transporter reactions and removing sink-chols\_c0 and sink-hpyr\_c0 sink reactions (the final models are available in Narrative <https://kbase.us/n/40576/330/>). We constrain the total number of carbon atoms that a model can consume –  $60 \frac{\text{mmol C}}{\text{gDW * hour}}$  by default, which corresponds to  $10 \frac{\text{mmol glucose}}{\text{gDW * hour}}$  when glucose is the only carbon source – to limit simulated growth to a reasonable value even when several carbon sources are simultaneously utilized. The ecological role of *B. theta* as an aerotolerant – can survive but not grow in aerobic environments [22] – digester of polysaccharides [23] was captured by manually setting *MI* and biomass growth in its GEM to zero when the isolated model was simulated in an aerobic environment, while leaving the GEM unconstrained in a community.

Several noteworthy observations are available from the various models in the various media. First, the community models, especially the mixed-bag model, generally exhibited higher growth relative to the individual models, although the highest for growth was observed with *E. coli* in aerobic conditions. Second, the community models successfully grew on dextrin despite that *E. coli* failed to grow as an isolate, which may be explained by *E. coli*'s inability metabolize complex polysaccharides [24] but this is not completely understood [25]. Third, the community models grew on glutamine (a nitrogen source) equivalently to glucose and dextrin (carbon sources), however, *E. coli* nearly doubled and tripled the community growth rate in anaerobic and aerobic conditions on glutamine relative to the carbon sources, respectively. Finally, the community models grew equivalently on sulfate as glucose (the mixed-bag model grew slightly more, while *B. theta* failed to grow and *E. coli* grew slightly more than the community models. The first two observations demonstrate community synergy that may justify its stable formation among these members. The latter two observations reveal conditions where *E. coli* self-sacrifices to join the community, since it could grow more as an independent isolate, notwithstanding that sulfur metabolism in bacteria is poorly understood [26].

Model Statistics	Single		Community	
	<i>B. theta</i> / iAH991	<i>E.coli</i> / iML1515	MB	CCM
<b>Original Model Description</b>				
<b>Number reactions</b>	1246	2379	3166	3625
<b>Number compounds</b>	1179	1878	2398	2944
<b>Genes</b>	993	1515	2508	2508
<b>Number compartments</b>	2	3	3	5
	c0-Cytosol e0-Extracellular	c0-Cytosol e0-Extracellular p0-Periplasm	c0-Cytosol e0-Extracellular p1-Periplasm	c0-Cytosol c1- <i>E. coli</i> c2- <i>B. theta</i> e0-Extracellular p1-Periplasm
<b>Aerobic Glucose MM</b>	Glucose carbon source (presence of Oxygen)			
<b>Active reactions</b>	0	417	578	1116
<b>Objective value / Growth</b>	0	0.477218	0.428449	0.414962
<b>Anaerobic Glucose MM</b>	Glucose carbon source (absence of Oxygen)			
<b>Active reactions</b>	594	413	564	979
<b>Objective value / Growth</b>	0.162608	0.16799	0.428449	0.173746
<b>Aerobic Dextrin MM</b>	Glucose replace with Dextrin carbon source (presence of Oxygen)			
<b>Active reactions</b>	0	0	573	1042
<b>Objective value / Growth</b>	0	0	0.428449	0.428449
<b>Anaerobic Dextrin MM</b>	Glucose replace with Dextrin carbon source (absence of Oxygen)			
<b>Active reactions</b>	522	0	569	934
<b>Objective value / Growth</b>	0.214225	0	0.428449	0.428449
<b>Aerobic Glutamine MM</b>	Ammonia replace with Glutamine nitrogen source (presence of Oxygen)			
<b>Active reactions</b>	0	416	563	1034
<b>Objective value / Growth</b>	0	1.11357	0.428449	0.428449
<b>Anaerobic Glutamine MM</b>	Ammonia replace with Glutamine nitrogen source (absence of Oxygen)			
<b>Active reactions</b>	537	413	566	982
<b>Objective value / Growth</b>	0.214225	0.675723	0.428449	0.428449
<b>Aerobic Sulfate MM</b>	Hydrogen Sulfate removed (presence of Oxygen)			
<b>Active reactions</b>	0	417	580	1148
<b>Objective value / Growth</b>	0	0.477218	0.428449	0.411395
<b>Anaerobic Sulfate MM</b>	Hydrogen Sulfate removed (absence of Oxygen)			
<b>Active reactions</b>	0	413	576	987
<b>Objective value / Growth</b>	0	0.16799	0.428449	0.172423

CMM – Compartmentalized Community Model MB – Mixed-Bag MM – Minimal Media (Details in supplementary S3)

Table 1. Individual and community models statistics.

## Mutual Information Analysis of Isolate and Community models

The RFMIA App was applied to each model isolate and community model, with upper bounds for all combinations of the seven nutrients in EZE: (1) Glucose ( $C_6H_{12}O_6$ ), (2) Glutamine ( $C_5H_{10}N_2O_3$ ), (3) Dextrin ( $H(C_6H_{10}O_5)xOH$ ), (4)

112

113

114

Ammonia ( $NH_3$ ), (5) Sulfate ( $SO_4^{2-}$ ), (6) Hydrogen sulfate ( $H_2SO_4$ ), and (7) Oxygen ( $O_2$ ). Diverse responses were observed in Fig. 7 for aerobic and anaerobic conditions and between the isolates and community models. Compound combinations without information flow were removed.

Figure 7 reveals that both *E. coli* and *B. theta* consist of information flow between 0.544–2.122 bits and 0.669–1.936 bits, where {1,2,3,4,5,6} generates the maximum values for each member with containing two carbon sources, two nitrogen sources, and two sulfur sources and {1,4,5,6} generates the minimum values for each member with one carbon source, one nitrogen source, and two sulfur sources. *B. theta* exhibits greater information flow in anaerobic conditions with the presence of 1-2 sources of carbon, nitrogen, and sulfur: e.g. {1,2,3,4,6}, {1,2,4,6}, {2,3,4,6}, {3,4,5,6}, {1,4,6}, {2,3,6}, and {3,4,6}. *E. coli* generally exhibits greater information flow otherwise, particularly with glutamine, which is consistent with the biological role of the latter as an obligate anaerobe that has substantial limitations as an isolate. The presence of sulfate in combination with carbon and nitrogen sources augments information flow in *E. coli*, while the absence of a sulfur, carbon, or nitrogen source often results in zero information flow.

Almost every set of E2E inputs in Fig. 7 manifests in greater information flow in the community models compared with single species, which supports the ecological synergy of these members. Information flow is even greater in mixed-bag models than compartmentalized models, presumably because the latter is encumbered with transport exchanges between compartments. This trend was observed in both aerobic conditions in Figure 8 and anaerobic conditions in Fig. 9. *E. coli* further outperformed *B. theta* in all conditions except two anaerobic conditions {1,3,4,5,6} and {1,4,5,6}, which is consistent with the ecological roles of each member [14]. Information flow in the absence of a carbon source is zero for *B. theta*, low for *E. coli*, and highest for the community models, which further reveals circumstances that justify community formation. The presence of glucose and dextrin manifested in the greatest information flow, followed by glucose exceeding dextrin when deployed singularly, *ceteris paribus*. The presence of both glutamine and ammonia nitrogen sources similarly manifested in greater information flow than either source individually, followed by glutamine exceeding ammonia. *B. theta* does not exchange information in the absence of hydrogen sulfate, regardless of the presence of sulfate, which suggests that *B. theta* may require additional protons for survival, while both isolates exhibit zero information flow when hydrogen sulfate is present without sulfate. These results affirm the nutritional necessity of these mineral and energy sources, and indicate that information flow generally follows *B. theta* < *E. coli* < CCM < MB.

## Discussion

Biological cells leverage communication, both among themselves and with other cells, to conduct relatively complex cellular functions and govern interactions and community dynamics. Understanding these understudied communication mechanisms is therefore essential for understanding human health, ecological systems, and agricultural productivity, and ultimately for rational engineering a range of cellular behaviors or stable communities. We devised a Mutual Information score (MI) to quantify the information flow of a metabolic system and probe these communication mechanisms via a unique molecular communication model that applies Shannon information theory. We exemplify MI with an idealized 2-member microbiome community of *E. coli* and *B. theta*, where information flow was determined to be greater in community models than in isolated members, consistent with the Shannon Information theory's definition of information which scales proportionally with the number of possible system states. The mixed-bag community framework furthermore exhibited greater information flow than compartmentalized community models since this framework is not encumbered by transporters to exchange compounds between compartments and may therefore represent a theoretical limit of a frictionless community metabolism. The MI values of *E. coli* were also generally greater than those of *B. theta* – except for several anaerobic conditions – which is consistent with the broader ecological habitable zone of *E. coli*. We further ingrained MI into KBase as an Application (*Run Flux Mutual Information Analysis*) that conveniently visualizes information flows for concise interpretation and importantly streamlines usage of MI and broadens its accessibility to non-technical biologists. MI analyses identify the combinations of nutrients that optimally trigger information flow, and may therefore be a valuable tool to reduce the number of resource-intensive experiments to probe a given metabolic system. We envision that MI and the *RFMIA* App will accelerate basic and applied discoveries in myriad biological fields. Future work will characterize larger microbial communities and chart information flow as a function of member abundances to understand how competitive forces influence information flow in community systems.

## Supporting information

**S1 Mathematical Models. Mutual Information Mathematical Models** Details of the mathematical models are in the Mutual\_Information\_MathematicalModels.pdf file.

## S2 Figures 7, 8, and 9.

165

**S3 Table Minimal Media (MM)** The table contains the compounds the minimal necessities for growth of the both organisms. Details of the compound list are in file Btheta\_Ecoli\_minimal\_media.csv

166

167

**S4 Figures 7, 8, and 9 bar chart values in excel file** The values and the original formats for B\_theta\_E\_Coli\_BarCharts.xlsx are in the Excel file sheets 1, 2, and 3, respectively.

168

169

## Acknowledgments

170

We thank the KBase developers team in Chris Henry's lab at Argonne National Laboratory (ANL) for their active support throughout the development of this research. Also, special thanks to Dr. Gail W. Pieper at ANL for helpful discussions and comments on revisions. This work was financially supported by the Molecular and Biochemical Telecommunications (MBiTe) Lab at the Department of Computer Science and Engineering, University of Nebraska - Lincoln. This work was supported by the US National Science Foundation through grant MCB-1449014, EPSCoR EPS-1004094, NSF CCF-1816969, and the National Institutes of Health (NIH) through grant 1-P20-GM113126-01. The material was based in part on work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.

171

172

173

174

175

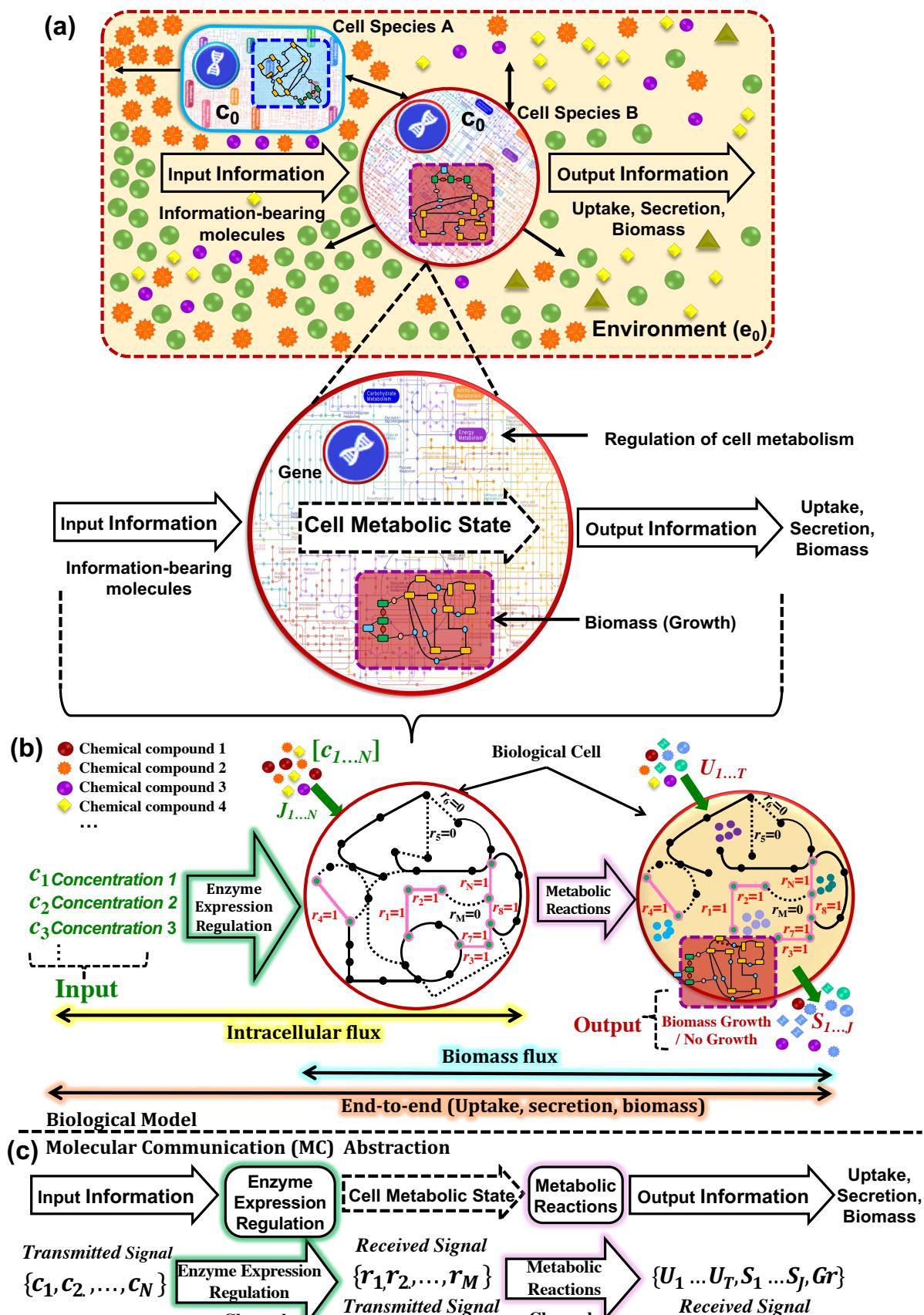
176

177

## References

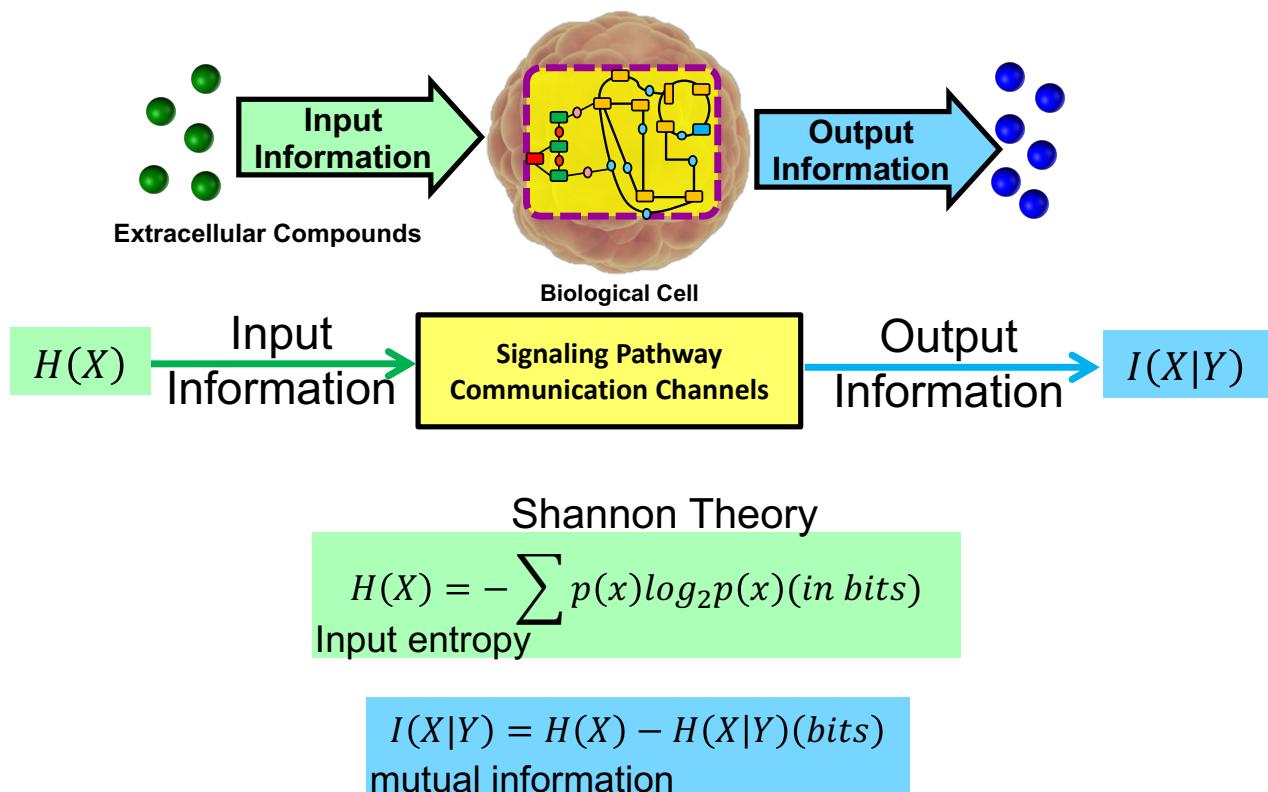
1. E. Gonçalves, J. Bucher, A. Ryll, J. Niklas, K. Mauch, S. Klamt, M. Rocha, and J. Saez-Rodriguez, "Bridging the layers: Towards integration of signal transduction, regulation and metabolism into mathematical models," 2013.
2. C. M. Metallo and M. G. Vander Heiden, "Understanding Metabolic Regulation and Its Influence on Cell Physiology," 2 2013.
3. R. E. Antwis, S. M. Griffiths, X. A. Harrison, P. Aranega-Bou, A. Arce, A. S. Bettridge, F. L. Brailsford, A. de Menezes, A. Devaynes, K. M. Forbes, E. L. Fry, I. Goodhead, E. Haskell, C. Heys, C. James, S. R. Johnston, G. R. Lewis, Z. Lewis, M. C. Macey, A. McCarthy, J. E. McDonald, N. L. Mejia-Florez, D. O'Brien, C. Orland, M. Pautasso, W. D. Reid, H. A. Robinson, K. Wilson, and W. J. Sutherland, "Fifty important research questions in microbial ecology," 5 2017.
4. M. Rosenberg, N. F. Azevedo, and A. Ivask, "Propidium iodide staining underestimates viability of adherent bacterial cells," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.
5. S. Blasche, Y. Kim, A. P. Oliveira, and K. R. Patil, "Model microbial communities for ecosystems biology," 12 2017.
6. M. C. Tsilimigras and A. A. Fodor, "Compositional data analysis of the microbiome: fundamentals, tools, and challenges," 5 2016.
7. J. P. Faria, T. Khazaei, J. Edirisinghe, P. B. Weisenhorn, S. Seaver, N. Conrad, N. Harris, M. DeJongh, and C. S. Henry, "Constructing and analyzing metabolic flux models of microbial communities," 2016.
8. Z. Sakkaff, A. P. Freiburger, J. L. Catlett, M. Cashman, A. Immaneni, N. R. Buan, M. B. Cohen, C. Henry, and M. Pierobon, "A Molecular Communication model for cellular metabolism," *bioRxiv*, 2023.
9. M. Pierobon, Z. Sakkaff, J. L. Catlett, and N. R. Buan, "Mutual information upper bound of molecular communication based on cell metabolism," in *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, vol. 2016-August, Institute of Electrical and Electronics Engineers Inc., 8 2016.
10. Z. Sakkaff, J. L. Catlett, M. Cashman, M. Pierobon, N. R. Buan, M. B. Cohen, and C. A. Kelley, "End-to-end molecular communication channels in cell metabolism: An information theoretic study," in *Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication, NanoCom 2017*, Association for Computing Machinery, Inc, 9 2017.
11. A. Rhee, R. Cheong, and A. Levchenko, "The application of information theory to biochemical signaling systems," *Physical Biology*, vol. 9, 8 2012.

12. T. D. Schneider, "The founder of information theory used biology to formulate the channel capacity," *IEEE ENGINEERING IN MEDICINE AND BIOLOGY MAGAZINE*, 2006.
13. J. D. Orth, I. Thiele, and B. O. Palsson, "What is flux balance analysis?," *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.
14. T. Conway and P. S. Cohen, "Commensal and Pathogenic Escherichia coli Metabolism in the Gut," *Microbiology Spectrum*, vol. 3, 6 2015.
15. R. Krajmalnik-Brown, Z. E. Ilhan, D. W. Kang, and J. K. DiBaise, "Effects of gut microbes on nutrient absorption and energy regulation," 4 2012.
16. T. Khazaei, R. L. Williams, S. R. Bogatyrev, J. C. Doyle, C. S. Henry, and R. F. Ismagilov, "Metabolic multistability and hysteresis in a model aerobe-anaerobe microbiome community," *Science Advances*, vol. 6, no. 33, 2020.
17. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. Dejongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, and D. Yu, "KBase: The United States department of energy systems biology knowledgebase," *Nature Biotechnology*, vol. 36, no. 7, pp. 566–569, 2018.
18. C. S. Henry, H. C. Bernstein, P. Weisenhorn, R. C. Taylor, J. Y. Lee, J. Zucker, and H. S. Song, "Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction," *Journal of Cellular Physiology*, vol. 231, no. 11, pp. 2339–2345, 2016.
19. N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, K. K. Weitz, R. Eils, R. König, R. D. Smith, and B. Palsson, "Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models," *Molecular Systems Biology*, vol. 6, 2010.
20. J. M. Monk, C. J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, A. M. Feist, and B. O. Palsson, "iML1515, a knowledgebase that computes *E. coli* traits," *Nature Biotechnology*, vol. 35, no. 10, pp. 8–12, 2017.
21. A. Heinken, S. Sahoo, R. M. T. Fleming, and I. Thiele, "Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut," *Gut Microbes*, vol. 4, no. 1, pp. 28–40, 2013.
22. H. M. Wexler, "Bacteroides: The good, the bad, and the nitty-gritty," 10 2007.
23. S. Mishra and J. A. Imlay, "An anaerobic bacterium, *Bacteroides thetaiotaomicron*, uses a consortium of enzymes to scavenge hydrogen peroxide," *Molecular Microbiology*, vol. 90, pp. 1356–1371, 12 2013.
24. L. C. Hoskins, M. Agustines, W. B. McKee, E. T. Boulding, M. Kriangs, and G. Niedermeyer, "Mucin Degradation in Human Colon Ecosystems Isolation and Properties of Fecal Strains that Degrade ABH Blood Group Antigens and Oligosaccharides from Mucin Glycoproteins," tech. rep., 1985.
25. S. Sabri, L. K. Nielsen, and C. E. Vickers, "Molecular control of sucrose utilization in *Escherichia coli* W, an efficient sucrose-utilizing strain," *Applied and Environmental Microbiology*, vol. 79, pp. 478–487, 1 2013.
26. A. Sekowska, H.-F. Kung, and A. Danchin, "Sulfur Metabolism in *Escherichia coli* and Related Bacteria: Facts and Fiction JMMB Review," Tech. Rep. 2, 2000.

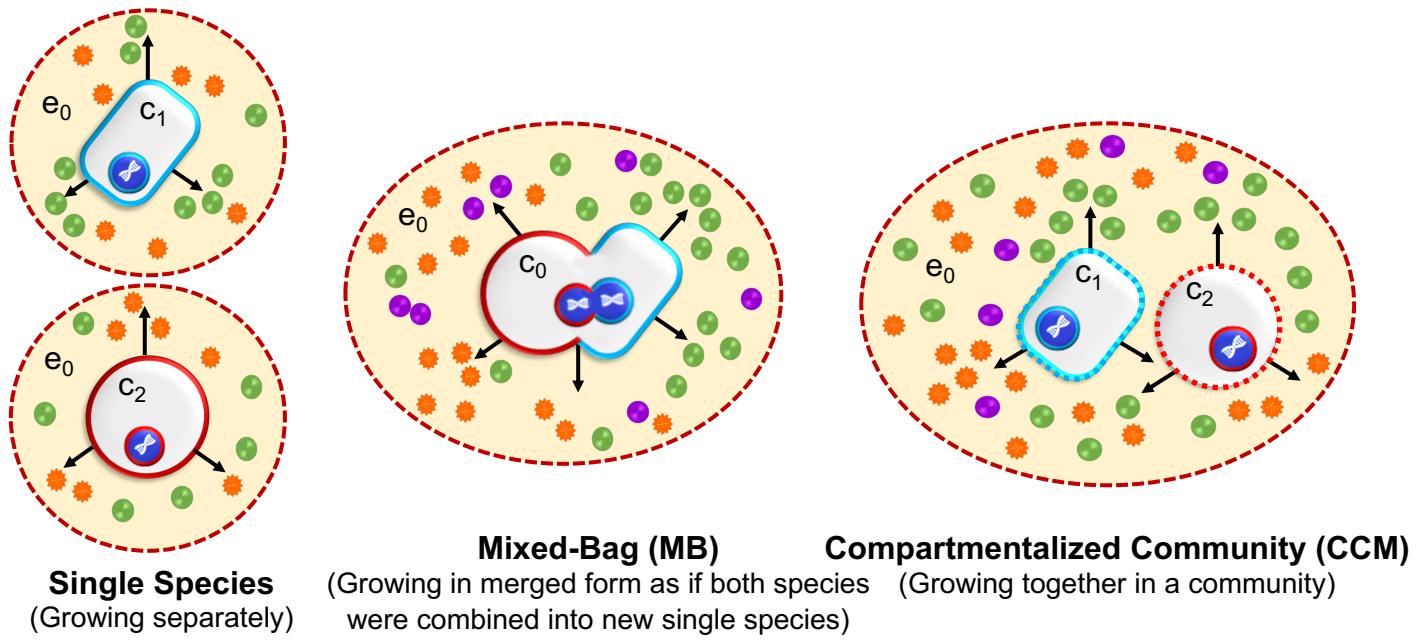


**Fig 2. Overall graphical representation of the molecular communication abstraction of cell metabolism.**

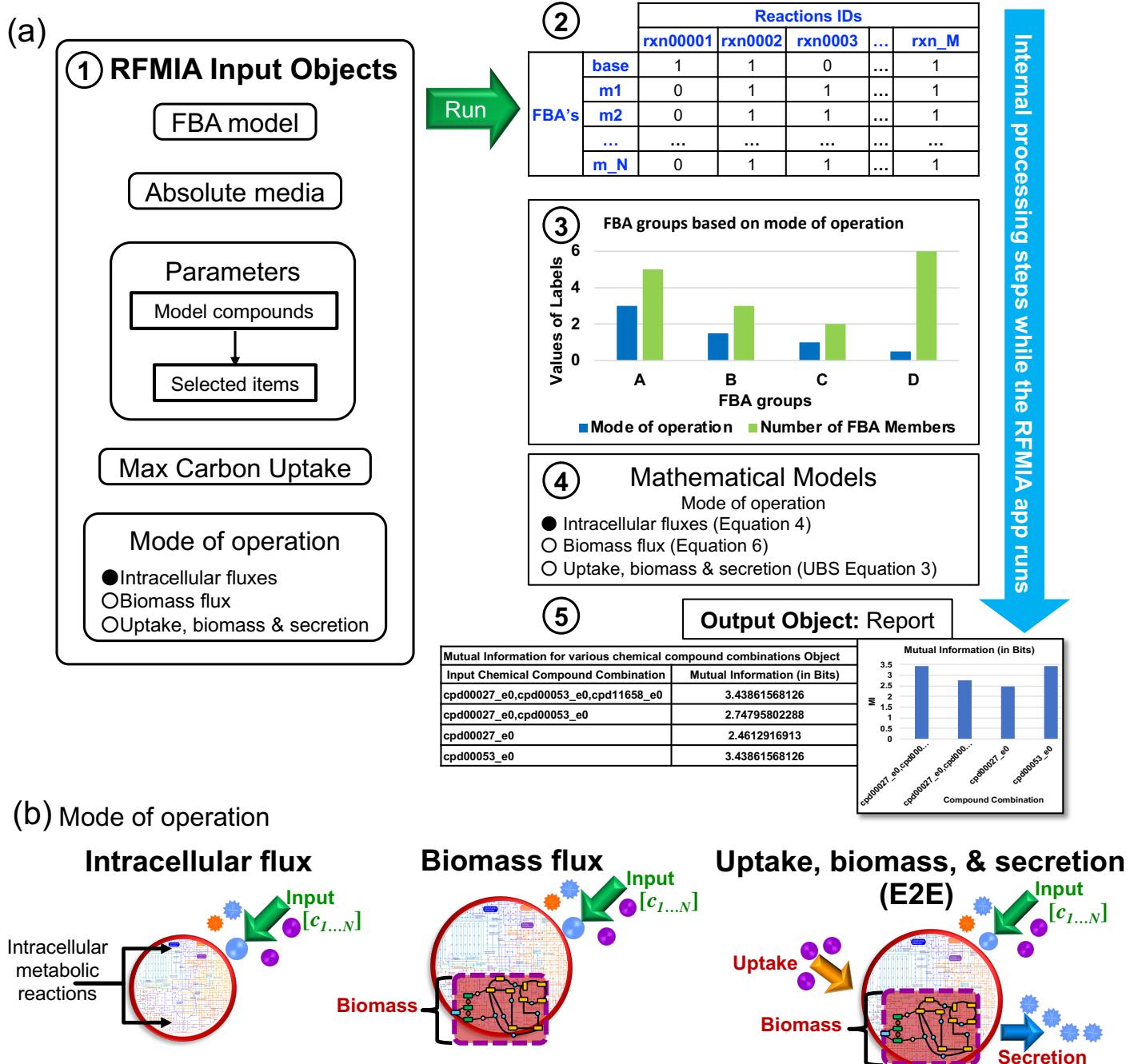
(a) Interaction of two different cell species A, and B with the presence of various compounds in the environment and the flow of information. (b) The cell takes certain concentrations of chemical compounds into the cell metabolism, where the variations in these concentrations cause state changes in the cell's metabolic network. Intracellular flux represents the cells response to the chemical composition of the cell's environment; Biomass flux captures variations in the growth rate of the cell in response to environmental changes; and End-to-End describes variations in the uptake and secretion of chemical compounds and the biomass (cell growth), which can be measured when we consider the cell metabolic processes from the perspective of the environment. (c) Mapping MC abstraction with biological model



**Fig 3. Fundamental mathematical framework of Shannon information theory in the biological system.**  
Formulation of pure uncertainty of the input signals presented as extracellular compounds and the mutual information for output signals presented as cellular responses to inputs.



**Fig 4. Details of interspecies models in this study.** Three different formalisms we use to study information flow through metabolic systems. Left-hand models show the environmental interaction of individual species. Middle MB shows individual species blended into a single new species. Right hand CCM, where individual species tightly interact with each other while interacting with the environment. Species 1 are represented in purple, species 2 in green, and anything common in both organisms in orange. Filled and open circles represent extracellular and intracellular metabolites, respectively; solid boxes mean that compartments outside of which the quasi-steady state assumption may not hold, and dashed boxes indicate compartments outside of which the steady-state assumption continues to hold. As stated, the communities did not have equal abundances, and that situation may lead to inaccurate conclusions about any interactions. Hence, the relative abundance values allow one to formulate the biomass composition for the entire community.



**Fig 5. General RFMIA tool pipeline.** (a) 1-5 steps involved: 1-input to the tool, 2-internal processing of reaction table with FBAs, 3-FBA groups with MI values, 4-MI math models and calculations, and 5-MI table and bar chart. (b) A pictorial description of the mode of operation.

**(a)**

**Toolbar**

**Narrative**

**Data**

**Apps**

**FBA model**

**Base media**

**Available Compounds**

**Selected Compounds**

**Result Report**

**(b)**

**Report**

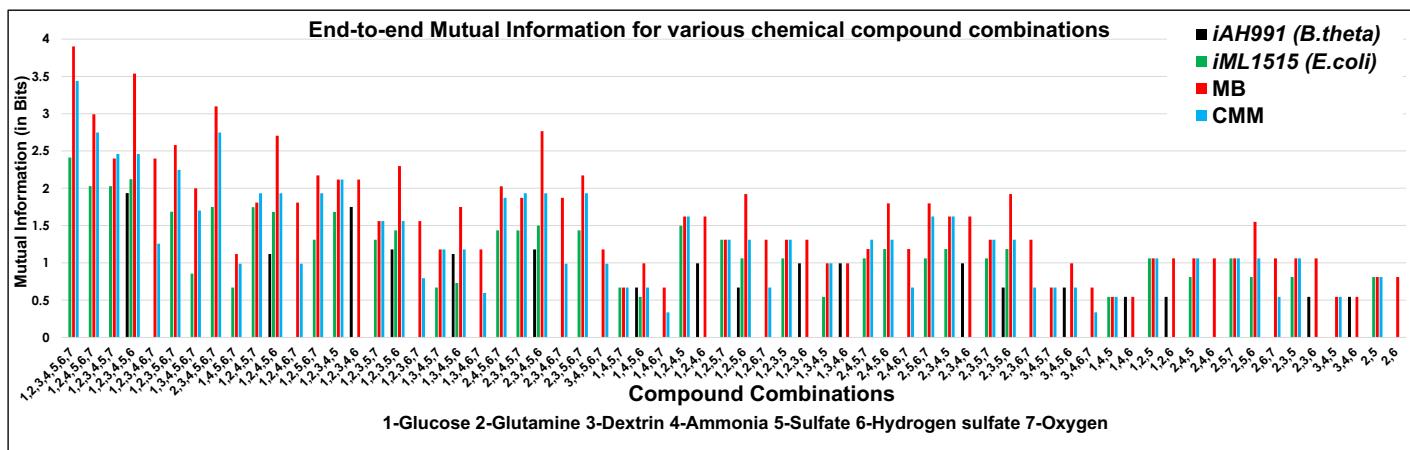
**Result Report**

**List of chemical compound combination**

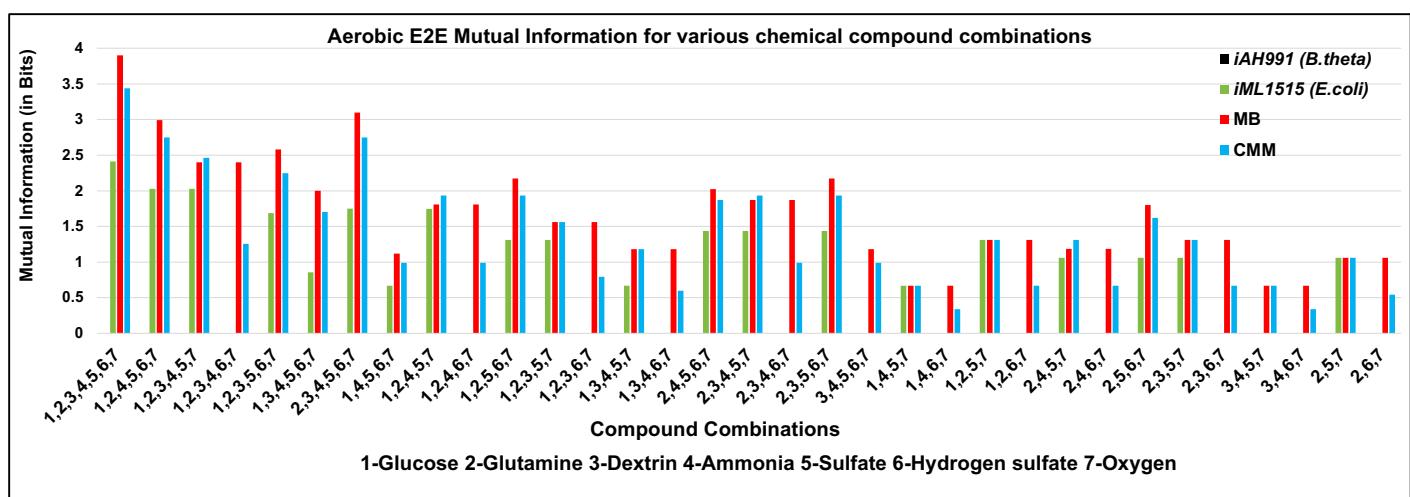
**Mutual Information (in Bits)**

**Bar Chart of Mutual Information (in Bits)**

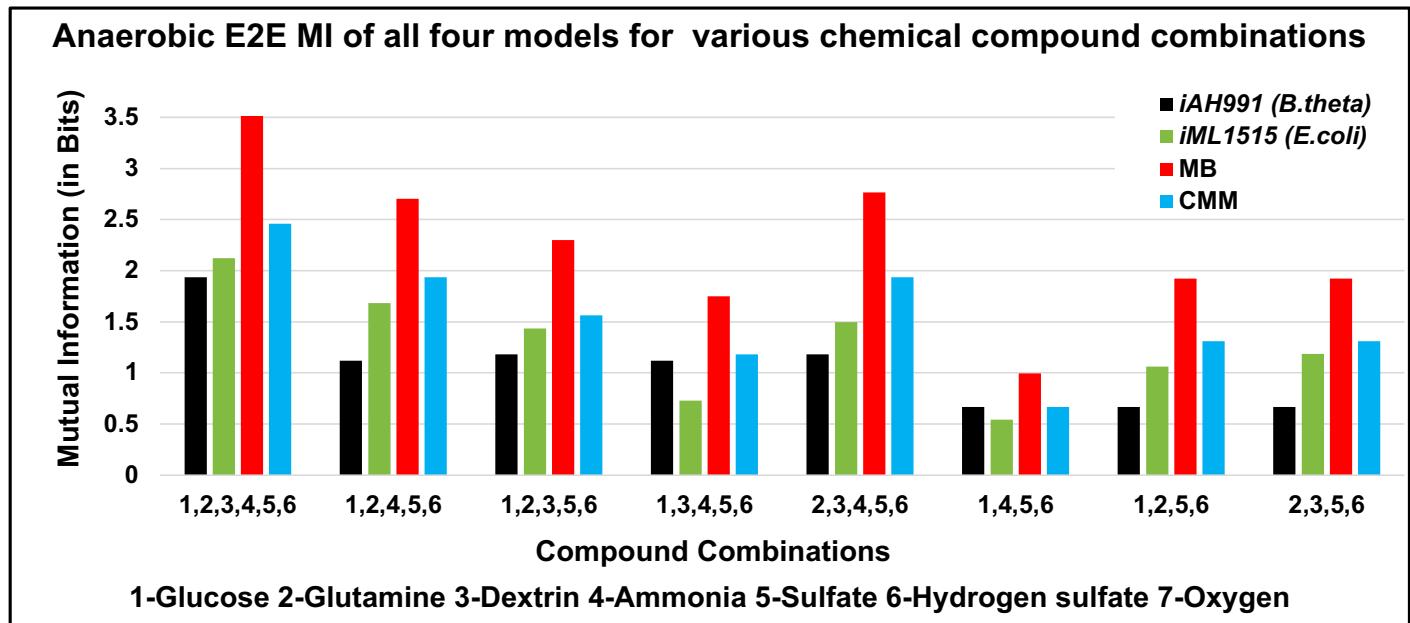
**Fig 6. RFMIA online user interface available in KBase [?] as point and click apps.** (a) Describes the user interface containing narrative description, RFMIA tool features, input data, and the list of applications. (b) RFMIA run results in tabular and bar chart formats.



**Fig 7. End-to-end (E2E) mutual information chart with respect to various input compound combinations.** E2E mutual information for the different combinations of seven compounds of individual *E. coli*, *B. theta*, mixed-bag(MB), and compartmentalized community model (CCM), with respect to flux of uptaken and secreted compounds and biomass. Compound combinations with zero bits of all four models are not presented. Corresponding bar chart values are available in the supporting document S4.



**Fig 8. Aerobic E2E mutual information chart for different combinations of seven compounds.** MI values of *E. coli*, *B. theta*/iAH991V2, mixed-bag (MB), and compartmentalized community model (CCM) with all the combinations of seven compounds, with respect to E2E MC systems. MI combinations with zero bits are not presented.



**Fig 9.** Anaerobic E2E mutual information values of all four models for seven compound combinations. A chart describes the anaerobic E2E MI values of *B. theta*/iAH991V2, *E. coli*, mixed-bag (MB), and compartmentalized community model (CCM) with all the combinations of seven compounds respectively, to E2E MC systems. MI combinations with zero bits are not presented.