

3652 (errors:4) words
File: main.tex
Encoding: ascii
Sum count: 3652
Words in text: 2268
Words in headers: 35
Words outside text (captions, etc.): 1258
Number of headers: 13
Number of floats/tables/figures: 18
Number of math inlines: 86
Number of math displayed: 5
Subcounts:
text+headers+captions (#headers/#floats/#inlines/#displayed)
263+8+0 (2/0/0/0) _top_
565+1+0 (1/0/10/1) Section: Introduction
129+1+163 (1/1/7/2) Section: Methods
234+2+0 (1/0/15/2) Subsection: Mutual Information} \label{sec:MIofMCA
263+4+193 (1/1/8/0) Subsection: Computational Tools and Workflow
53+2+0 (1/0/2/0) Section: Numerical Results} \label{sec:numerical_results
285+4+145 (1/2/32/0) Subsection: Results for each channel} \label{sec: MC_Stage I
149+3+86 (1/1/6/0) Subsection: Input Signal Constellation
262+4+0 (1/0/0/0) Section: Conclusions and Future Work} \label{sec:Conclusion
54+1+0 (1/0/0/0) Section: Acknowledgment
11+3+0 (1/0/0/0) Section: Conflict of interest
0+2+671 (1/13/6/0) Section: Supplementary Information

File: output.bbl
Encoding: utf8
Sum count: 0
Words in text: 0
Words in headers: 0
Words outside text (captions, etc.): 0
Number of headers: 0
Number of floats/tables/figures: 0
Number of math inlines: 0
Number of math displayed: 0

Total
Sum count: 3652
Words in text: 2268
Words in headers: 35
Words outside text (captions, etc.): 1258
Number of headers: 13
Number of floats/tables/figures: 18
Number of math inlines: 86
Number of math displayed: 5
Files: 2
Subcounts:
text+headers+captions (#headers/#floats/#inlines/#displayed)
2268+35+1258 (13/18/86/5) File(s) total: main.tex
0+0+0 (0/0/0/0) File(s) total: output.bbl

(errors:4)

A Molecular Communication model for cellular metabolism

Zahmeeth Sakkaff^{✉1} · Andrew P. Freiburger^{✉1} · Jennie L. Catlett² · Mikaela Cashman³ · Aditya Immaneni³ · Nicole R. Buan² · Myra B. Cohen³ · Christopher Henry^{✉1} · Massimiliano Pierobon^{3*}

the date of receipt and acceptance should be inserted later

Abstract Understanding cellular engagement with the environment is essential to control and monitor metabolism. Molecular Communication theory (MC) quantifies the amount of information about a molecular environment that is transmitted through a metabolic system, and thereby offers a computational means for identifying environmental perturbations that direct cellular behavior or signal metabolic changes. We developed a model that couples conventional flux balance analysis (FBA) metabolic modeling with MC and defined several input-output metabolic channels of communication that may intrigue investigators: particularly one that identifies reaction activity and one that identifies excreta and biomass growth from a consumption profile. We quantify the information flow (or mutual information) about an environment through cellular metabolism in bits. The application of FBA further identifies the upper limit of mutual information, and thus the maximal capacity of the environment to control cellular metabolism and behaviors. We exemplified our method on two intestinal symbionts *Bacteroides thetaiotaomicron* and *Methanobrevibacter smithii*, and developed visualization tools that construct hive plots from the FBA flux outputs and constellation diagrams from the MI results to facilitate biological interpretation. The unique confluence of metabolic modeling and information theory in this model advances basic understanding of cellular metabolism and has applied value

for the Internet of Bio-Nano Things, synthetic biology, microbial ecology, and autonomous laboratories.

Keywords Molecular communication · information theory · cell metabolism · regulation of gene expression · flux balance analysis · mutual information

1 Introduction

Engineering biological systems and bio-nanotechnologies is essential to realize the Internet of Bio-Nano Things (IoBNT) [1] where wearable bio-computers monitor and direct biological systems *in situ* [2]. Externally controlling and monitoring cellular functions is often pursued through biosensing [3], optogenetics [4], or magnetic nanoparticles [5]; however, these experimental approaches have notable limitations, such as off-target effects on the biological system and inherent biochemical noise [6, 7]. The ideal method would instead leverage native metabolism, and thereby minimally disturb the studied system: e.g. regulating a patient's microbiome through dietary intervention [8, 9]. The basic principles of information flow from an environment into complex metabolic systems [10, 11] and moreover how these factors influence behaviors – post-translation modification, signal transduction [12], gene regulation, metabolism, and biomass growth [6, 10, 13] – remain bottlenecks to realize this ideal control of cellular behavior [14].

Molecular communication (MC) theory is an emerging confluence of communication and information theories [7, 15–17] that can quantify the flow of information about an environment through a cellular metabolism to excretion fluxes and biomass growth, which is illustrated in Figure S3. MC black-boxes metabolism into channels of inputs and outputs [18]. The information

¹Argonne National Laboratory, Data Science and Learning Division, Lemont, IL 60148 USA

²Department of Biochemistry, University of Nebraska-Lincoln, Lincoln

³Department of Computer Science & Engineering, University of Nebraska-Lincoln, Lincoln

*✉ pierobon@cse.unl.edu

flow through each of these channels can then be quantified in binary digits, or bits (1 or 0), as uncertainty (mutual information) I [19]

$$I = H(inputs) - H(inputs, outputs) \quad (1)$$

as the excess of input uncertainty $H(inputs)$ from the output uncertainty $H(inputs, outputs)$, which depends on the input values. The *input* and *output* sets are characteristic for each channel. Flux Balance Analysis (FBA) [20], in contrast, is a metabolic modeling framework that mechanistically resolves metabolism through a linear optimization that determines the profile of active metabolic reactions and maximal biomass growth for an organism in a given environment, assuming that the cell can appropriately acclimates [21]. FBA represents a metabolic system as a matrix of stoichiometries for all metabolites in all reactions S and determines the reaction fluxes $\frac{mmol}{hr * g_{DW}}$ as a vector \mathbf{v} that represents metabolic activity at a steady-state ($S\mathbf{v} = 0$) [22] while optimizing a sub-set of fluxes (generally those contributing to biomass growth) and conforming to flux ranges for each reaction ($\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max}$) that can approximate the effects of diverse biological phenomena: e.g. reaction energetics [?] or regulatory processes [23]. This method is represented by Figure S1.

We define a unique coupling of MC and FBA that facilitates computing information flow through defined metabolic channels by MC while elucidating metabolic mechanisms by FBA. We define a comprehensive MC channel from an input of consumed substrates to excretions and bimoass growth, and define two channels as partitions of this comprehensive channel [24], which is illustrated in Figure S7. We exemplify our novel coupling of metabolic FBA and MC information theory with genome-scale models (GEM) [25] of two mutualistic [26] human symbionts that are associated with metabolic diseases [27, 28] – the bacterium *B. theta* and the archaeon *M. smithii* – and seven substrates that represent the major carbon, nitrogen, sulfur, and oxygen nutrient flows in chyme and have been previously studied with these organisms [26, 29]. We also present a signal constellation diagram – a useful visualization that can reliably transmit information across noisy channels [30] by mapping symbols 1-to-1 to their encoded counterpart with a unique phase and amplitude – that visualizes mutual information flow for all potential consumption combinations of these seven substrates to reveal which consumption profiles optimally perturbs cellular metabolism despite biochemical noise. We envision that our model expands basic knowledge of cellular behavior and can specifically address basic questions – 1) how much environmental information (quantified in bits) can a cell encode in its metabolism; and 2)

how much information about intra-cellular metabolism can be externally perceived? – as well as accelerating progress towards engineering and rationally designing microbiomes.

2 Methods

Figure 1 illustrates the general abstraction of cellular metabolism as a molecular communication model of information flow from substrate consumption through black-boxed intra-cellular metabolism and finally externally perceived behaviors of metabolite excretion and biomass growth. We partition this pipeline into several molecular communication channels: **Stage I** in Figure S2

$$\{U_1, U_2, \dots, U_N\} \xrightarrow[\text{Consumption}]{\text{Substrate}} \{r_1, r_2, \dots, r_M\} . \quad (2)$$

for the activation of intra-cellular metabolism via substrate consumption; **Stage II** in Figure S5 for the externally perceived exchanges and biomass growth

$$\{r_1, r_2, \dots, r_M\} \xrightarrow[\text{Reactions}]{\text{Metabolic}} \{S_1, S_2, \dots, S_J, Gr\} ; \quad (3)$$

and **end-to-end** as the combination of **Stage I** and **Stage II** to encompass the entire pipeline. The employed terms include: $U_i \forall i \in N$ as binary indications of consumption for the N compounds of interest (specified by the user); $r_j \forall j \in M$ are binary indications of reaction activity for all M metabolic reactions; $S_e \forall e \in E$ are binary indications of excretion for all E exchangeable compounds; and Gr is a binary indication of biomass growth.

2.1 Mutual Information

The input uncertainty $H(inputs)$ of Equation 1 is generally defined as

$$H(inputs) = - \sum_{i=1}^I (P(inputs)) * \log_2(P(inputs))d(inputs) , \quad (4)$$

where I is the number of unique inputs and $P()$ is a function that returns the probability of the provided argument, with all inputs and outputs being equally probable. The conditional output uncertainty $H(inputs)$ is generally defined as

$$H(inputs, outputs) = - \sum_{o=1}^O (P(o)) * \sum_{i=1}^I (P(inputs, outputs)) * \log_2(P(inputs, outputs))d(inputs) \quad (5)$$

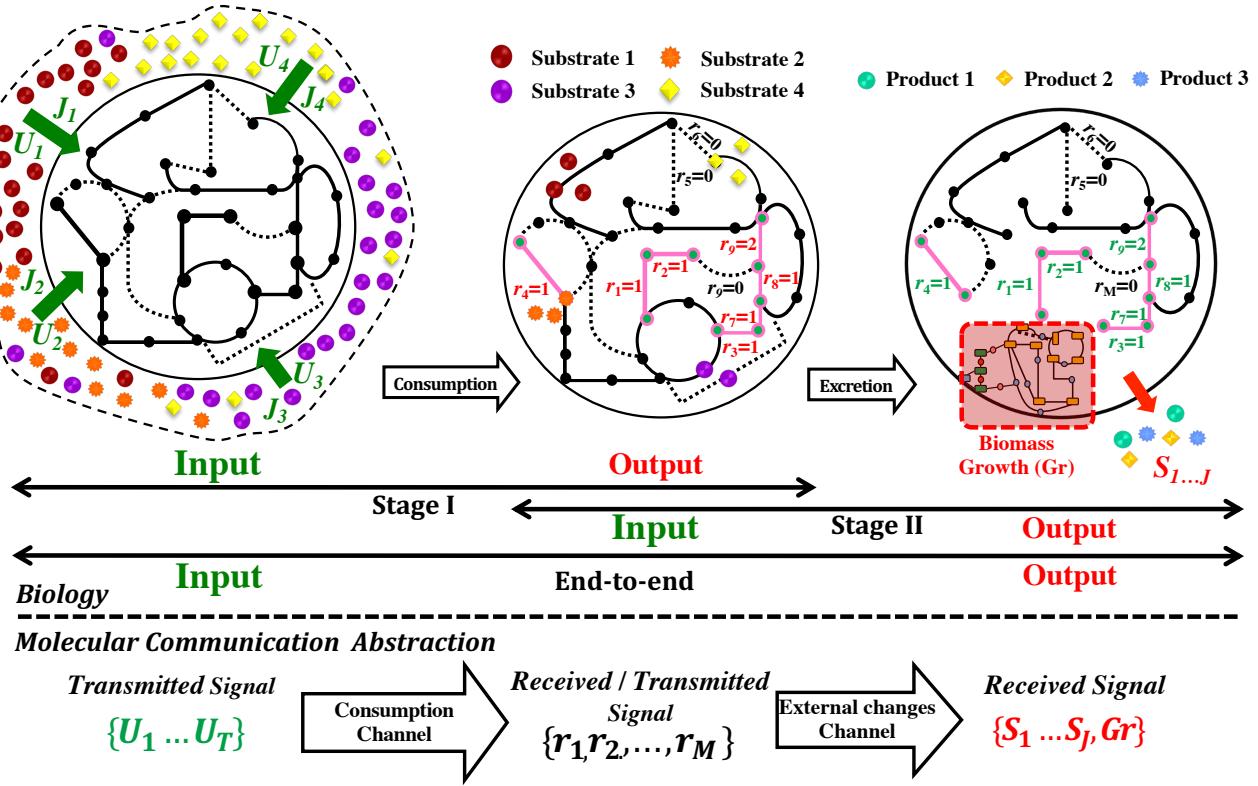


Fig. 1: The molecular communication abstraction of metabolism as a transceiver, in comparison to the biological processes. The top pipeline above the dashed line depicts the biological phenomena of a) substrates being absorbed into a biological cell; b) metabolic pathways being triggered by these nutrients; and c) output signals of metabolic activity, such as excreta or biomass growth. The bottom pipeline below the dashed line depicts the abstraction of phenomena in molecular communication where the substrates are reduced to a set of chemical concentrations, the triggered metabolic pathways are reduced to a set of activated reactions, and the perceived outputs are reduced to a set of reaction fluxes for specific biological processes. This abstraction is further discretized into three partitions: Stage I that captures triggering metabolic reactions based on consumed substrates; Stage II that captures perceived outputs based on activated reactions; and end-to-end that captures the full pipeline as the combination of Stages I & II. These partitions are individually examined in this study.

where O is the number of unique outputs. The $P(inputs, outputs)$ term specifically returns the probability of an output per a given input, while $P(outputs)$ returns the probability of an output in the collection of all outputs. The summations in these equations can be integrals when the inputs or outputs are continuous, however, our strict utilization of binary variables discretizes all domains and makes summation sufficient. Our implementation of mutual information defines variable domains for each channel: $inputs = U_i \forall i \in N$ and $outputs = \{r_j\}_{j=1}^M$ for the Stage I channel; $inputs = \{r_j\}_{j=1}^M$ and $outputs = \{S_e\}_{e=1}^E, Gr$ for the Stage II channel; and $inputs = \{U_i\}_{i=1}^N$ and $outputs = (\{S_e\}_{e=1}^E, Gr)$ for the end-to-end channel.

The aforementioned general case is tailored for FBA

simulations in several dimensions. First, the terms are demarcated with an asterisk (e.g. $\{r_j^*\}_{j=1}^M$) to denote that the activity corresponds with an optimum. Second, the total flux for each of the examined compounds is constrained to the total flux of all reactions that contain these compounds from the FBA optimization. The process of identifying activity of all consumptions, reaction activities, and excretions was facilitated by grouping FBA solutions according to their flux profiles in Figure S6. The mutual information from these optimized values I^* is considered to be the upper bound of information flow.

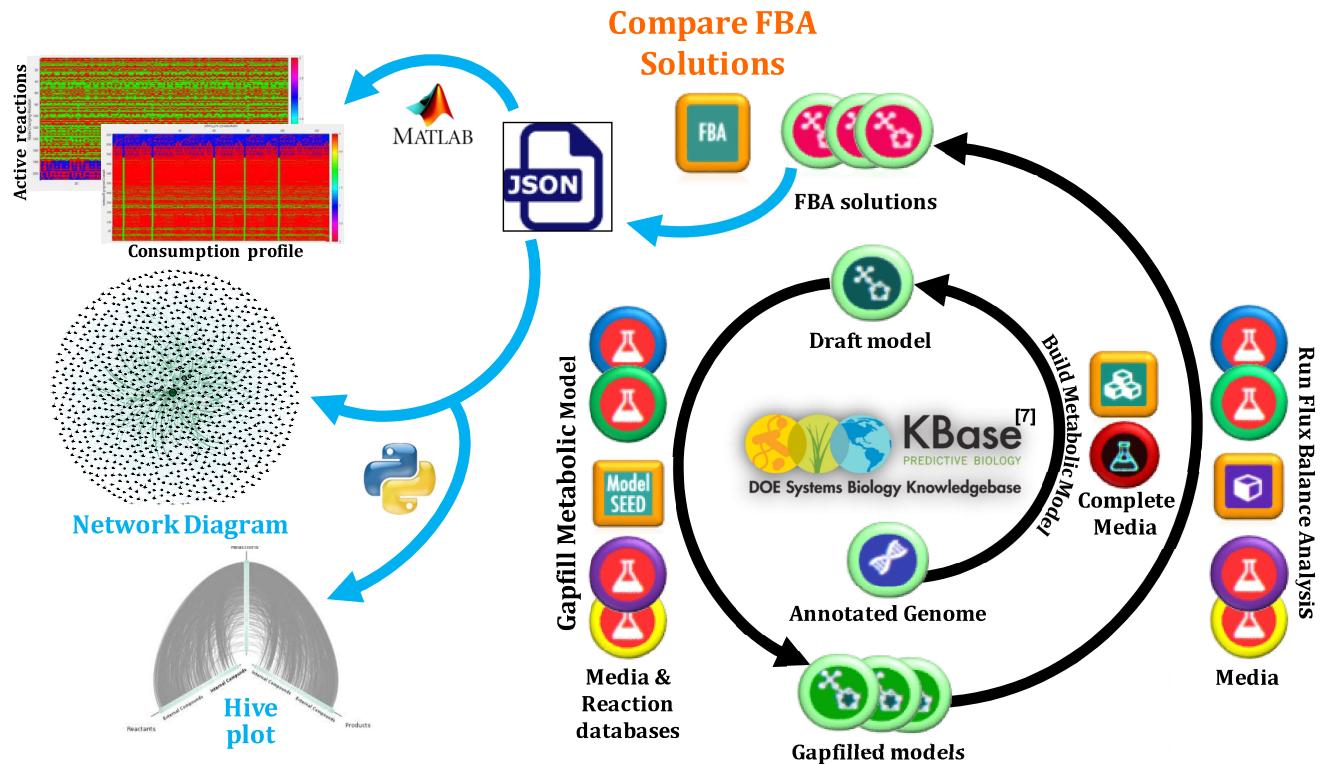


Fig. 2: Our modeling workflow, which leverages open-source KBase Applications and custom scripts. The steps that are conducted in KBase include: 1) annotating uploaded genomes, through algorithms such as RAST; 2) constructing draft genome-scale metabolic models (GEMs) from the annotated genomes, via the Build Metabolic Model application; 3) gap-filling the draft GEMs to create operational GEMs, via the Gapfill Metabolic Model application; 4) acquiring FBA flux profiles in the parameterized media, via the Run Flux Balance Analysis application; and then 5) exporting a JSON that compares FBA solutions of the different GEMs and conditions via the Compare FBA Solutions application. The exported JSON of compared FBA solutions is subsequently processed through custom MATLAB code into heatmaps that illustrate the sets of activated reactions in each FBA solution and through custom Python scripts that construct network diagrams (each node is a compound whose size is proportional the number of reactions (edges) that connect to it) and hive plots (a three-dimensional plot of reactions on the Z axis, reactants on the X axis, and the products on the Y axis, where substrates are higher in the axes than intra-cellular compounds) to illustrate relationships of the metabolic system.

2.2 Computational Tools and Workflow

The GEMs of *B. theta* and *M. smithii*, which are detailed in Figure S4, were constructed through the KBase pipeline [31] in Figure 2. First, the organism's genome sequences from the GenBank database [32, 33] were translated into protein sequences and then were mapped to functions via KEGG [34]. Second, these initial draft GEMs were gap-filled [35] to add the minimal number of biochemical reactions and compounds that enable growth in environments where the organisms experimentally grow to accommodate incorrect annotations or mappings. This process was repeated for each consumption profiles to generate an array of subtly different versions of each GEM. We examined seven compounds – glucose (*G*), hematin (*He*), for-

mate (*F*), H_2 , Vitamin (*B_{12}*), acetate (*A*), and Vitamin K (*K*) – similar to other studies with these organisms [26, 29], which yields $2^N = 2^7 = 128$ potential consumption profiles for which the mutual information was calculated [28], and used a standard base media for all simulations – calcium, chloride ion, carbon dioxide, cobalt, copper, ferrous ion, water, H^+ , potassium, L-cysteine, magnesium, manganese, sodium, ammonium, nickel, phosphate, sodium bicarbonate, sulfate, zinc, and ferric ion. All lower exchanges bounds were -100 while upper exchange bounds varied between 0.0000037 and 46166.8875 from FBA simulations in Table S1. Third, all of the GEMs were simulated through FBA and the flux profiles and objective values were compared and assembled into a single JSON file. Fi-

nally, the JSON FBA outputs were parsed into a matrix of reaction activation for each FBA solution for further processing and visualization via network diagrams and Hive plots [36], such as those in Figure 2 and Figure S8 respectively, or differential hive plots that contrasts activation profiles for simulations with the same or juxtaposed biomass growth fluxes such as Figure S9.

3 Numerical Results

The mutual information from various consumption profiles elucidated inter-dependent reactivity of the examined compounds, which would be important for experimentalists when attempting to environmentally control these organisms. Figure 3 depicts each possible consumption profile as a column and the activated reactions in ≥ 1 FBA simulations as M rows, which was reduced via grouping to foster Figure S6.

3.1 Results for each channel

The maximal mutual information for each channel was determined by solving each term in Equation 1. The presumed equal probability of any given consumption profile for our system of $N = 7$ examined substrates produces $P(\{U_G, U_{He}, U_F, U_{H_2}, U_{B_{12}}, U_A, U_K\}) = 1/128$. The input uncertainty for Stage I was computed by summing over the 128 inputs in Equation (4) to acquire $-1 * \log_2(128) = 7$ bits. The conditional output uncertainty for Stage I was calculated by substituting $M = 212$ for *B. theta* into Equation (5), which resulted in an output uncertainty 3.7 bits and a mutual information upper bound of $I^* = 3.3$ bits based on the 113 activated reactions and 114 unique FBA solutions, and $M = 556$ for *M. smithii*, which resulted in an output uncertainty of 2.5 bits and a mutual information upper bound of $I^* = 4.5$ bits based on the 136 activated reactions [37] and 97 unique FBA solutions. These results are depicted in Figure 3 and Figure S11, where *B. theta* reaches its maximum information flow occurs in the absence of formate while *M. smithii* reaches its maximum information flow in the presence of all seven examined substrates.

The plots of Figure S6 group all 128 FBA solutions for each species by activated reaction profiles, which resulted in 14 unique activation profiles for *B. theta* and 31 unique activation profiles for *M. smithii*. This simplifies the mutual information computation and can facilitate biological investigation in relating consumed substrates to metabolic activation and cellular behaviors. The matrix of the FBA solutions grouped by Gr

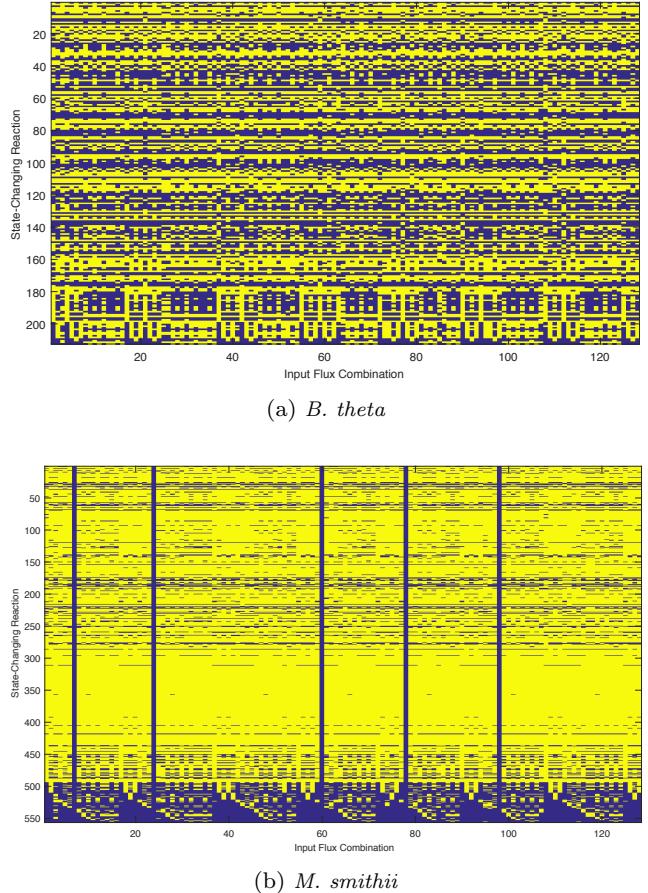
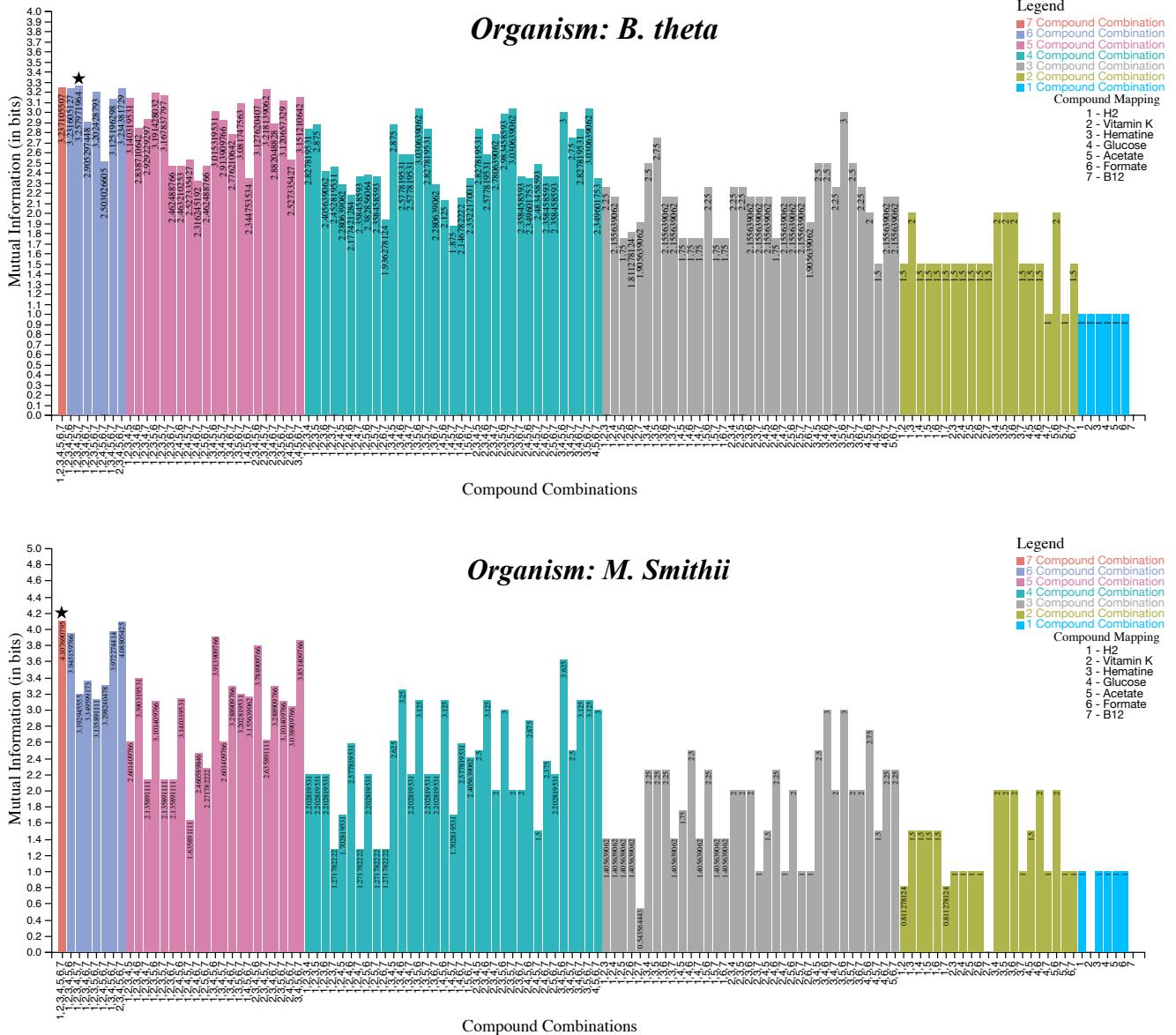


Fig. 3: Reaction activation from FBA simulations $\{r_i^*\}_{i=1}^M$ for each possible consumption profile, where yellow is activated and violet is inactive. *M. smithii* metabolism is evidently much more activated by the examined media, notwithstanding a handful of media profiles that did not support growth, than *B. theta*, although *B. theta* seems to have exhibited at least some metabolic activation in all of the examined media. This may be the consequence of *B. theta* utilizing anaerobic fermentation that expands its habitable zone beyond that of *M. smithii*.

was used to compute the maximal mutual information from $outputs = Gr$ as $I^* = 2.75$ bits in for *B. theta* and $I^* = 3.7264$ bits for *M. smithii* in Figure S12. A matrix of the FBA solution grouped by both $S_e \forall e \in E$ and Gr was analogously used for the end-to-end channel with $outputs = (S_e \forall e \in E, Gr)$ and obtained $I = 3.23$ bits for *B. theta* and $I^* = 4.11$ bits for *M. smithii* in Figure 4. All four of these maximal upper bounds for the Stage II and end-to-end channels cooresponded to consuming all seven of the examined substrates.



3.2 Input Signal Constellation

The constellation diagrams of Figures 4, S11, and S12 depict a symbol for each of the 128 possible consumption profiles, while Figure 5 revealed the unique consumption profiles that manifest in discrete cellular behaviors. These diagrams were constructed by first creating a separate table for each consumed compound and then: 1) grouping the tables by I^* from the end-to-end channel; 2) determining the minimum and maximum I^* for each consumption profile; and 3) representing the minimum, maximum, and average (excluding the minimum and maximum) I^* as separate symbols. These diagrams interestingly revealed that the maximum I^* begin to plateau after the consumption of ≥ 5 of the examined compounds, which suggests that there are diminishing returns to cellular control from environmental perturbations. This observation is echoed by the plateauing increase over the consumption profiles in Figure 4, and is moreover consistent with BioSIMP [29]

for each consumption profile; and 3) representing the minimum, maximum, and average (excluding the minimum and maximum) I^* as separate symbols. These diagrams interestingly revealed that the maximum I^* begin to plateau after the consumption of ≥ 5 of the examined compounds, which suggests that there are diminishing returns to cellular control from environmental perturbations. This observation is echoed by the plateauing increase over the consumption profiles in Figure 4, and is moreover consistent with BioSIMP [29]

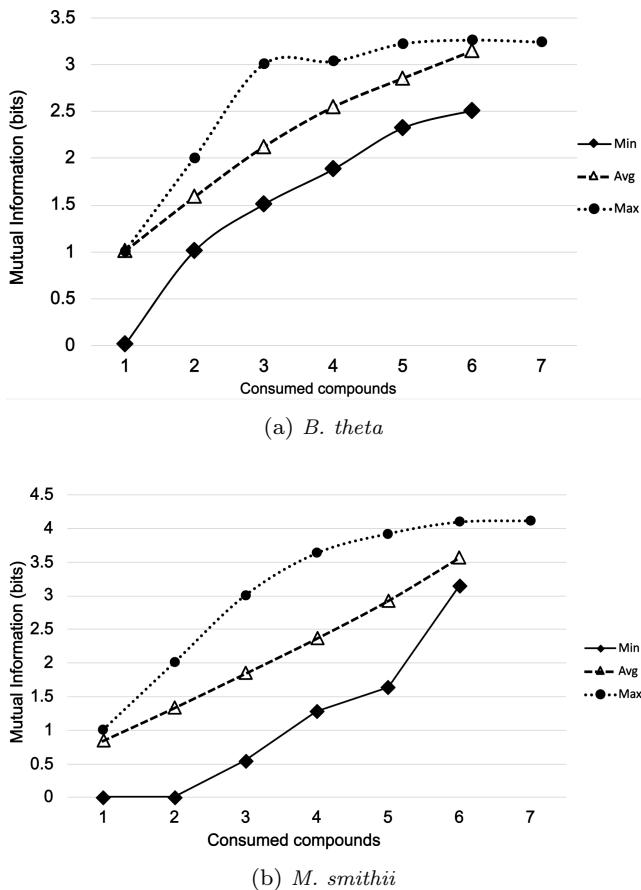


Fig. 5: Constellation diagrams for the mutual information from all consumption profiles in the total end-to-end abstraction. The minimum, maximum, and average values are plotted against the number of compounds in the consumption profiles, which reveals that the maximum information flow increases logarithmically while the average information flow increases linearly. This observation is intuitive where inherent noise or otherwise defines a fundamental limit to information flow but generally an increase in the considered dimensions (consumed compounds) increases the uncertainty (mutual information) of the system.

that was used to identify the environmental nutritional factors required for growth of *B. theta* and *M. smithii*.

4 Conclusions and Future Work

Molecular communication theory quantifies the flow of information (in bits) about the extra-cellular environment that channel through cellular metabolism into activated reactions, excreted metabolites, or biomass growth. We materialized this concept into a model that couples conventional FBA with MC theory and two

black-box communication channels of metabolism: one whose output is intra-cellular reaction activity and another whose output is exchange fluxes and biomass growth. The quantified mutual information flow from our model enabled ranking consumption profiles based on their capacity to transmit information and manifest cellular outcomes, which can potentially minimize the number of resource-intensive experiments that are necessary to find environmental substrates that desirably control cellular behaviors. The use of FBA further identifies the upper bound of information flow via environmental substrates, and thereby defines the limit of environmental perturbation to control behaviors. We further developed visualization scripts that generate succinct figures, such as hive plots and a novel use of constellation diagrams, to facilitate interpreting information flow by investigators. Code and additional details are available in our GitHub repository (https://github.com/zahmeeth/MetabolicMC_Supplementary_Git_v1/). This model advances basic biological knowledge of cellular engagement with its environment, and provides a tool to identify specific substrates that can most effectively engineer cellular behavior from media perturbations that has applications in medical or ecological contexts. Subsequent work will refine our model by 1) better compensating biochemical noise; 2) capturing dynamic information flows via the dynamic FBA algorithm, and thereby introducing continuous variables into our formulation; 3) validating predictions with experimental data; and 4) empowering autonomous laboratories to monitor and control cellular behaviors based on information transmission.

Acknowledgment

This work was funded by the US National Science Foundation through grant MCB-1449014, the National Institutes of Health 11-P20-GM113126-01, hi Sa the NSF EPSCoR EPS-1004094 and the CCF-1161767. Also the grant supported by NIH 11-P20-GM113126-01. The authors would also like to thank KBase developers for their active support throughout the development of this research.

Conflict of interest

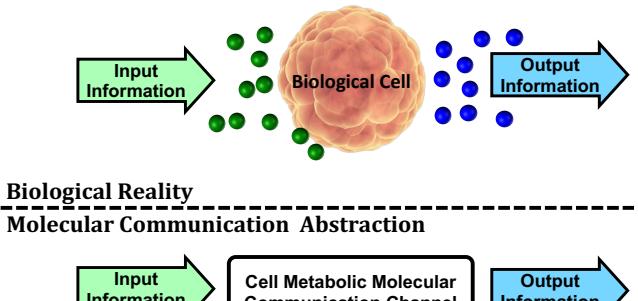
The authors declare no conflict of interest with the presented work.

References

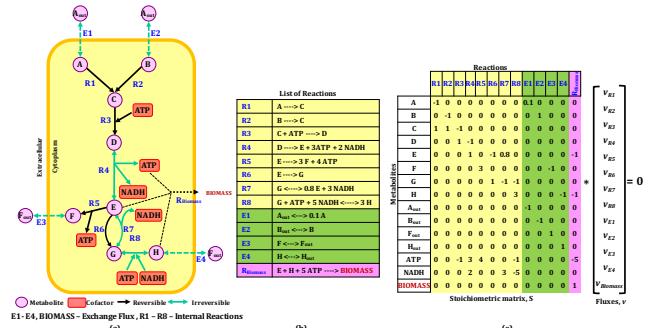
1. I. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Kouchevayev, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, 2015.
2. I. F. Akyildiz, J. M. Jornet, M. Pierobon, and S. Balasubramaniam, "Moving forward with molecular communication: From theory to human health applications," *Proceedings of the IEEE*, vol. 107, pp. 858–865, May 2019.
3. W. A. Lim, "The promise of optogenetics in cell biology: interrogating molecular circuits in space and time," *Nature Reviews*, vol. 11, pp. 393–403, 2010.
4. J. E. Toettcher, C. A. Voigt, O. D. Weiner, and W. A. Lim, "The promise of optogenetics in cell biology: interrogating molecular circuits in space and time," *Nature Methods*, vol. 8, pp. 35–38, 2011.
5. J. Dobson, "Remote control of cellular behaviour with magnetic nanoparticles," *Nature Nanotechnology*, vol. 3, pp. 139–143, 2008.
6. 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," *Molecular BioSystems*, vol. 9, no. 7, pp. 1576–1583, 2013.
7. A. F. Villaverde and J. R. Banga, "Reverse engineering and identification in systems biology: strategies, perspectives and challenges," *Journal of The Royal Society Interface*, vol. 11, no. 91, p. 20130505, 2014.
8. D. M. Klurfeld, C. D. Davis, R. W. Karp, E. Allen-Vercoe, E. B. Chang, B. Chassaing, G. C. Fahey, B. R. Hamaker, H. D. Holscher, J. W. Lampe, A. Marette, E. Martens, S. J. O. D. J. Rose, M. Saarela, B. O. Schneeman, J. L. Slavin, J. L. Sonnenburg, K. S. Swanson, G. D. Wu, C. J. Lynch, and F. G. Jr., "Considerations for best practices in studies of fiber or other dietary components and the intestinal microbiome," *Am J Physiol Endocrinol Metab*, vol. 315, pp. 1087–1097, 2018.
9. N. Fei, S. Miyoshi, J. B. Hermanson, J. Miyoshi, B. Xie, O. Deleon, M. Hawkins, W. Charlton, M. D. Souza, J. Hart, D. Sulakhe, K. B. Martinez-Guryn, E. B. Chang, M. R. C. Leone, and A. Affiliations, "Imbalanced gut microbiota after food diet mouse model," *bioRxiv*, 2023.
10. D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, ch. 12.2, pp. 425–429. W. H. Freeman, 2005.
11. O. S. Soyer, M. Salathe, and S. Bonhoeffer, "Signal transduction networks: topology, response and biochemical processes," *Journal of Theoretical Biology*, vol. 238, no. 2, pp. 416–425, 2006.
12. C. M. Metallo and M. G. V. Heiden, "Understanding metabolic regulation and its influence on cell physiology," *Molecular Cell*, vol. 49, pp. 388–398, February 2013.
13. N. Campbell and J. Reece, "Biology 7th edition, ap," 2005.
14. C. J. Myers, *Engineering genetic circuits*. Chapman & Hall/CRC, Mathematical and Computational Biology Series, 2009.
15. T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Edition*. Wiley, 2006.
16. F. Fabris, "Shannon information theory and molecular biology," *Journal of Interdisciplinary Mathematics*, vol. 12, no. 1, pp. 41–87, 2009.
17. I. F. Akyildiz, J. M. Jornet, M. Pierobon, and S. Balasubramaniam, "An information theoretic framework to analyze molecular communication systems based on statistical mechanics," *Proceedings of the IEEE*, vol. 107, pp. 1230–1255, July 2019.
18. A. Rhee, R. Cheong, and A. Levchenko, "The application of information theory to biochemical signaling systems," *Phys Biol.*, vol. 9, p. 045011, August 2012.
19. K. R. Gorla and M. Pierobon, "A mutual information estimate for a redox-based molecular-electrical communication channel," Association for Computing Machinery, Inc, 9 2021.
20. K. Höffner, S. Harwood, and P. Barton, "A reliable simulator for dynamic flux balance analysis," *Biotechnology and bioengineering*, vol. 110, no. 3, pp. 792–802, 2013.
21. E. H. Davidson, *The regulatory genome: gene regulatory networks in development and evolution*. Academic press, 2010.
22. J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?," *Nature biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.
23. M. W. Covert, C. H. Schilling, and B. Palsson, "Regulation of gene expression in flux balance models of metabolism," *Journal of Theoretical Biology*, vol. 213, pp. 73–88, 2001.
24. 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 (ACM NanoCom)*, September 2017.
25. G. J. Baart and D. E. Martens, "Genome-scale metabolic models: reconstruction and analysis," *Neisseria meningitidis: Advanced Methods and Protocols*, pp. 107–126, 2012.
26. J. L. Catlett, S. Carr, M. Cashman, M. D. Smith, M. Walter, Z. Sakkaff, C. Kelley, M. Pierobon, M. B. Cohen, and N. R. Buan, "Metabolic synergy between human symbionts bacteroides and methanobrevibacter," *Microbiology Spectrum*, vol. 10, 6 2022.
27. M. Cashman, J. L. Catlett, M. B. Cohen, N. Buan, Z. Sakkaff, M. Pierobon, and C. Kelley, "Sampling and Inference in Configurable Biological Systems: A Software Testing Perspective," Tech. Rep. TR-UNL-CSE-2016-0007, University of Nebraska-Lincoln, 2016.
28. M. Pierobon, M. B. Cohen, N. Buan, and C. Kelley, "SCIM: Sampling, Characterization, Inference and Modeling of Biological Consortia," Tech. Rep. TR-UNL-CSE-2015-0002, University of Nebraska-Lincoln, 2015.
29. M. Cashman, J. L. Catlett, M. B. Cohen, N. R. Buan, Z. Sakkaff, M. Pierobon, and C. A. Kelley, "Biosimp: using software testing techniques for sampling and inference in biological organisms," in *2017 IEEE/ACM 12th International Workshop on Software Engineering for Science (SE4Science)*, pp. 2–8, IEEE, 2017.
30. R. Tognari and J. Christopher, *Fundamentals of information theory and coding design*. CRC Press, 2003.
31. KBase, "Department of Energy Systems Biology Knowledgebase (KBase)." <http://kbase.us>. [Online; accessed 22-September-2016].
32. A. Sheydina, R. Y. Eberhardt, D. J. Rigden, Y. Chang, Z. Li, C. C. Zmasek, H. L. Axelrod, and A. Godzik, "Structural genomics analysis of uncharacterized protein families overrepresented in human gut bacteria identifies a novel glycoside hydrolase," *BMC bioinformatics*, vol. 15, no. 1, p. 1, 2014.
33. M. Zhou, Y.-H. Chung, K. Beauchemin, L. Holtshausen, M. Oba, T. McAllister, and L. Guan, "Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive," *Journal of applied microbiology*, vol. 111, no. 5, pp. 1148–1158, 2011.
34. KEGG, "Kyoto encyclopedia of genes and genomes (KEGG)." <http://www.genome.jp/kegg/>. [Online; accessed 22-September-2016].
35. C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, "High-throughput generation, optimization and analysis of genome-scale metabolic models," *Nature biotechnology*, vol. 28, no. 9, pp. 977–982, 2010.

36. M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra, "Hive plots—rational approach to visualizing networks," *Briefings in bioinformatics*, vol. 13, no. 5, pp. 627–644, 2012.
37. Z. Sakkaff, "Characterization of Molecular Communication Based on Cell Metabolism Through Mutual Information and Flux Balance Analysis," Master's thesis, University of Nebraska-Lincoln, December 2016.

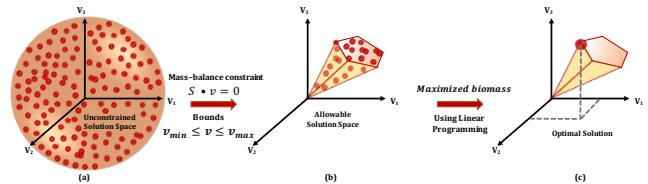
Supplementary Information



(a) Overview of biological realm and molecular communication abstraction.



(b) From metabolic network to stoichiometric matrix. (a) shows a conceptual model of a metabolic network where the internal chemical reactions are represented by R_i . (b) shows the exchange fluxes with the environment labeled as E_j . (c) shows the stoichiometric matrix S where each row corresponds to a chemical compound in the metabolic network, while each column corresponds to the reaction or flux exchanged with the environment. Each entry of the stoichiometric matrix S is the stoichiometric coefficient that indicates how many molecules of a chemical compound, represented by row entry, are consumed (coefficient < 0) or produced (coefficient > 0) in one of the possible reactions.



(c) Conceptual model of the FBA linear program formulation for finding the optimal solution. Optimization of the objective function $a'v$ identifies a flux distribution on the edge of the allowable solution space.

Fig. S1: Background information on a) the molecular communication abstraction of cellular systems, where cellular metabolism is black-boxed; b) the FBA algorithm and its use of matrix algebra; and c) the geometric solution space that linear optimization in FBA navigates to find the optimal solution.

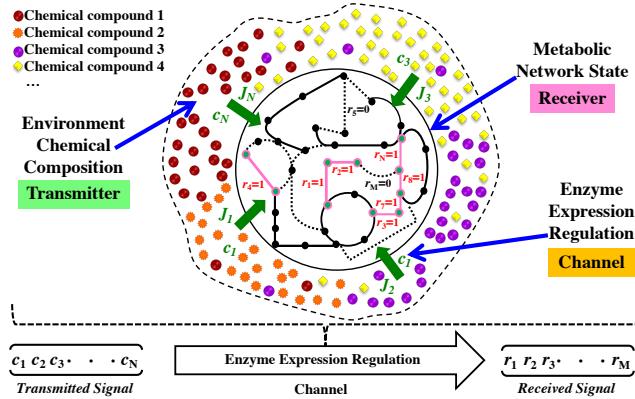


Fig. S2: Sketch of the proposed molecular communication system based on cell metabolism.

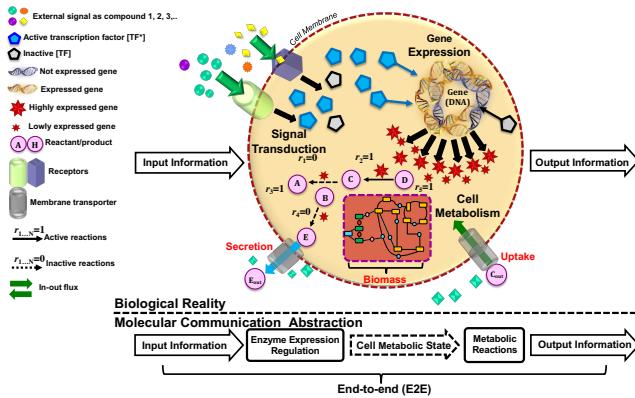


Fig. S3: Biological reality of Molecular communication in cell metabolism.

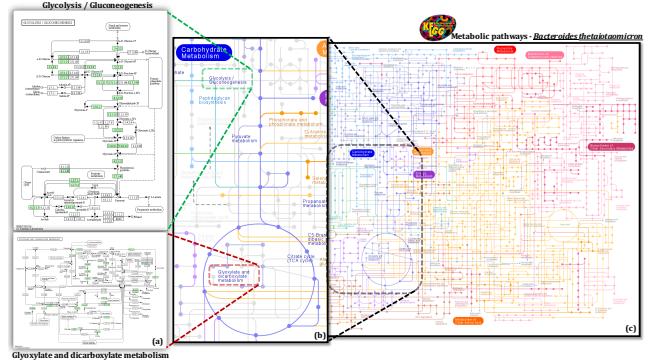


Fig. S4: KEGG module of *Bacteroides thetaiotaomicron* metabolic pathways. (a) Summary of the biological processes shown in the pathway map of Glycolysis / Gluconeogenesis and Glyoxylate and dicarboxylate metabolism. (b) Enlarged fine details of a section of a complete metabolic model. (c) Part of the complete KEGG database pathway maps of *Bacteroides thetaiotaomicron*. visualized parts of a GEM (on the right) for the organisms *B. theta*, used in our study, which is obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [34]. The nodes represent compounds that are inputs/outputs to the reactions, and edges represent the chemical reactions. Inputs from the environment taken by the organism are involved in the reactions of metabolic pathways, resulting in the exchange of fluxes with the environment (uptake and secretion) or in the production of biomass (growth).

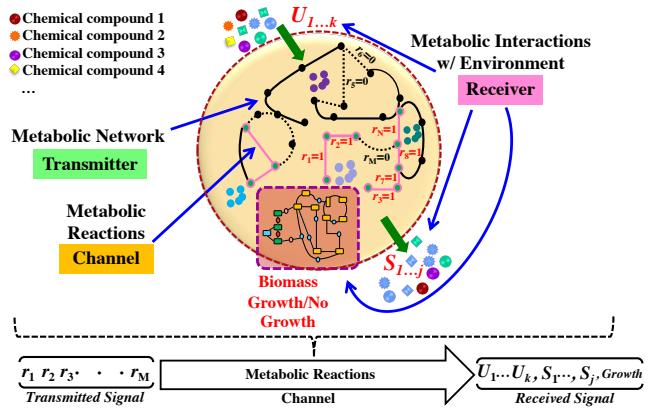


Fig. S5: Sketch of the proposed metabolic reaction abstraction.

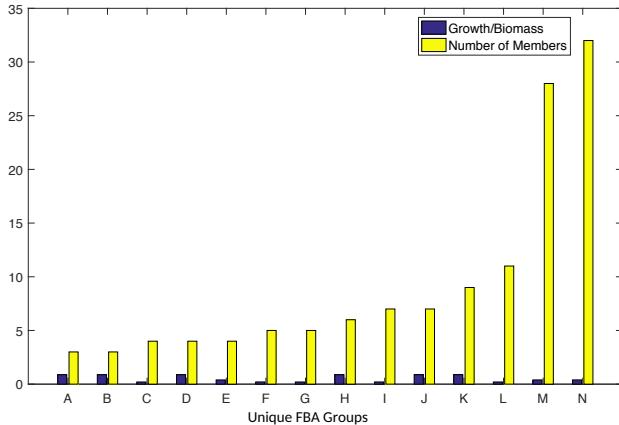
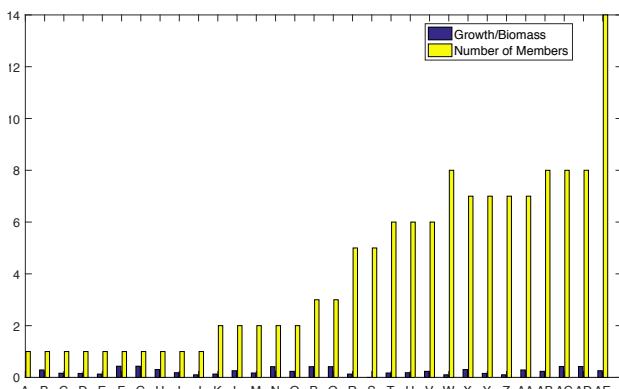
(a) 14 FBA groups of *B. theta*(b) 31 FBA groups of *M. smithii*

Fig. S6: The number of FBA solutions per group of unique activation profiles, over all consumption profiles. The *M. smithii* metabolism exhibited more than twice as many unique metabolic responses to the consumption profiles than *B. theta*, which suggests that it is more dynamic to its environment. This is corroborated by the much higher proportion of active reactions in Figure 3 relative to *B. theta*.

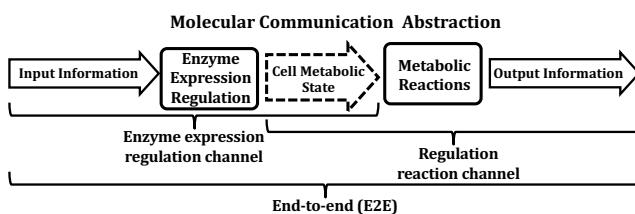


Fig. S7: Molecular communication channel abstraction of cell metabolism.

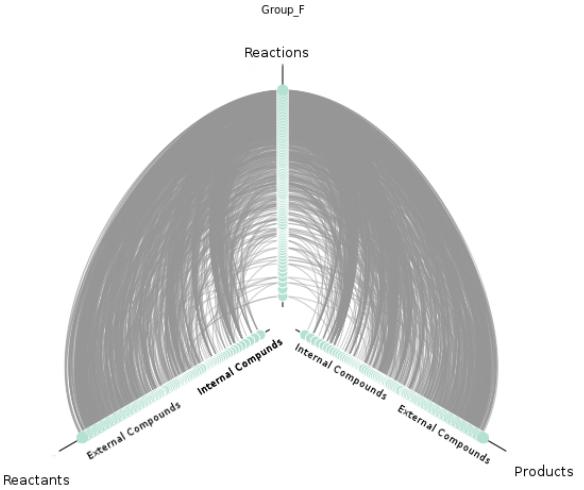


Fig. S8: A hive plot for a configuration F is shown in the figure. The reactions are placed on the Z axis, the reactants on the X axis and the products on the Y axis. Further the External compounds are placed higher on the X and Y axes than the Internal compounds.

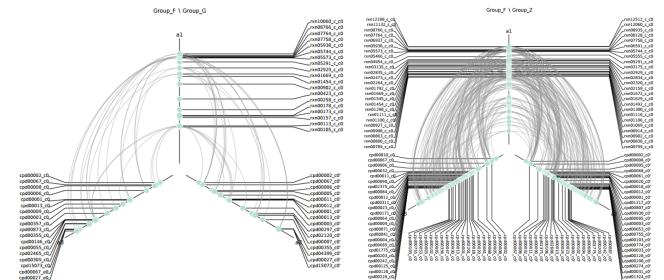


Fig. S9: In this figure we take 3 different configuration of state changing reactions, labelled F, G and Z. It shows differential hive plots of F vs G and F vs Z. The groups F and G in F vs G hive plot has the same biomass whereas, the groups F and Z in F vs Z hive plot have the least and highest biomass respectively. When a reaction is present in F and absent in G or Z the reaction is represented along with its links to the compounds. When a reaction is present in the other groups but absent in group F the reaction is shown as a node not connected to any other compounds.

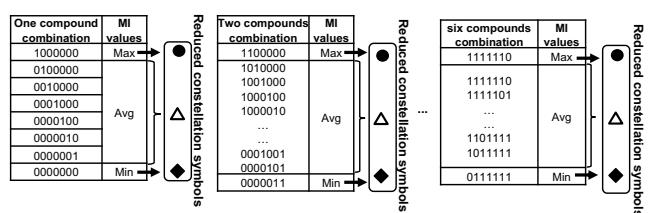


Fig. S10: Steps involved in constructing constellation diagram.

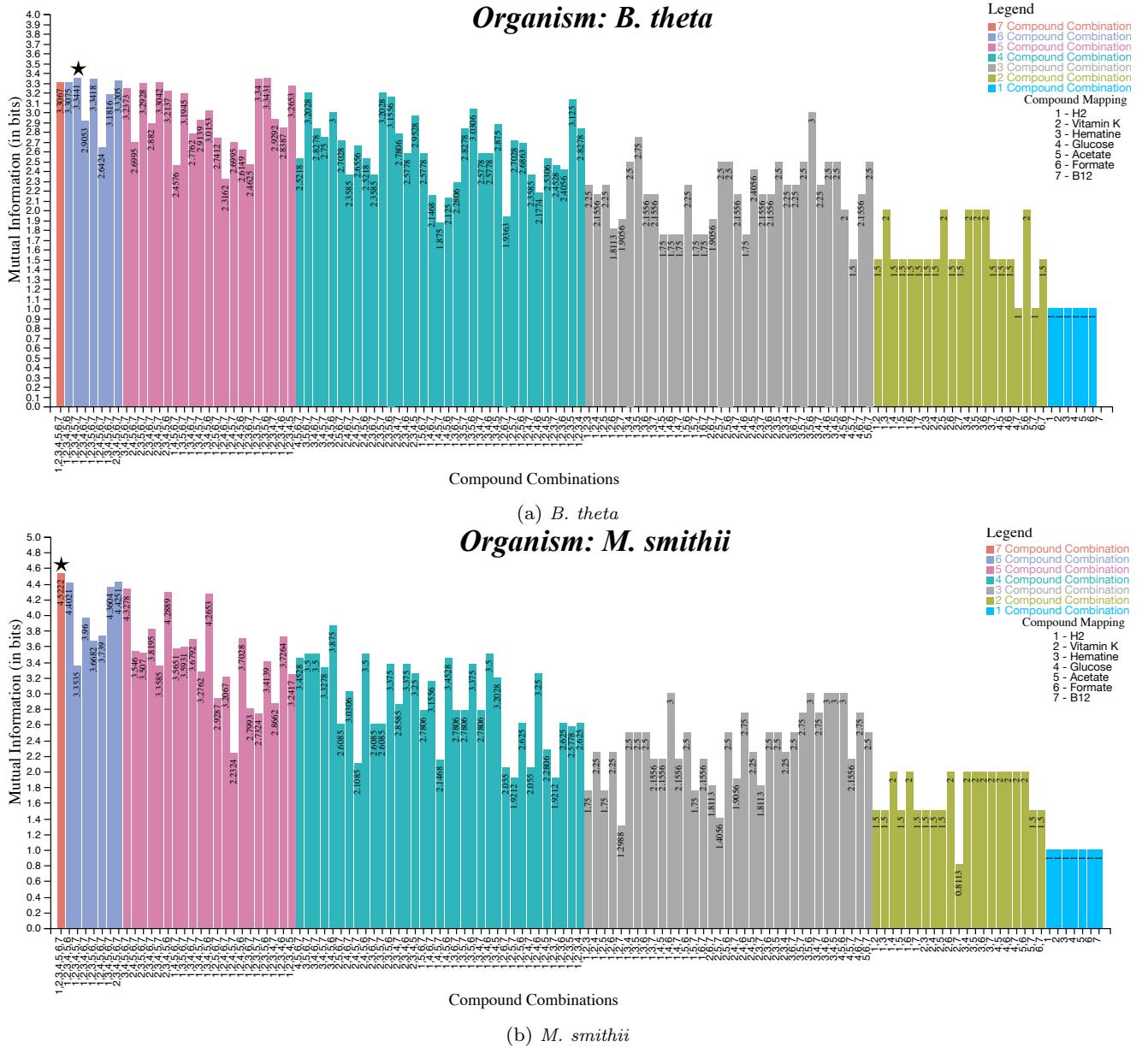


Fig. S11: Upper bounds of the steady-state mutual information for all the different combinations of seven compounds in Enzyme Expression Regulation channel in . The best estimation of the maximum information is starred.

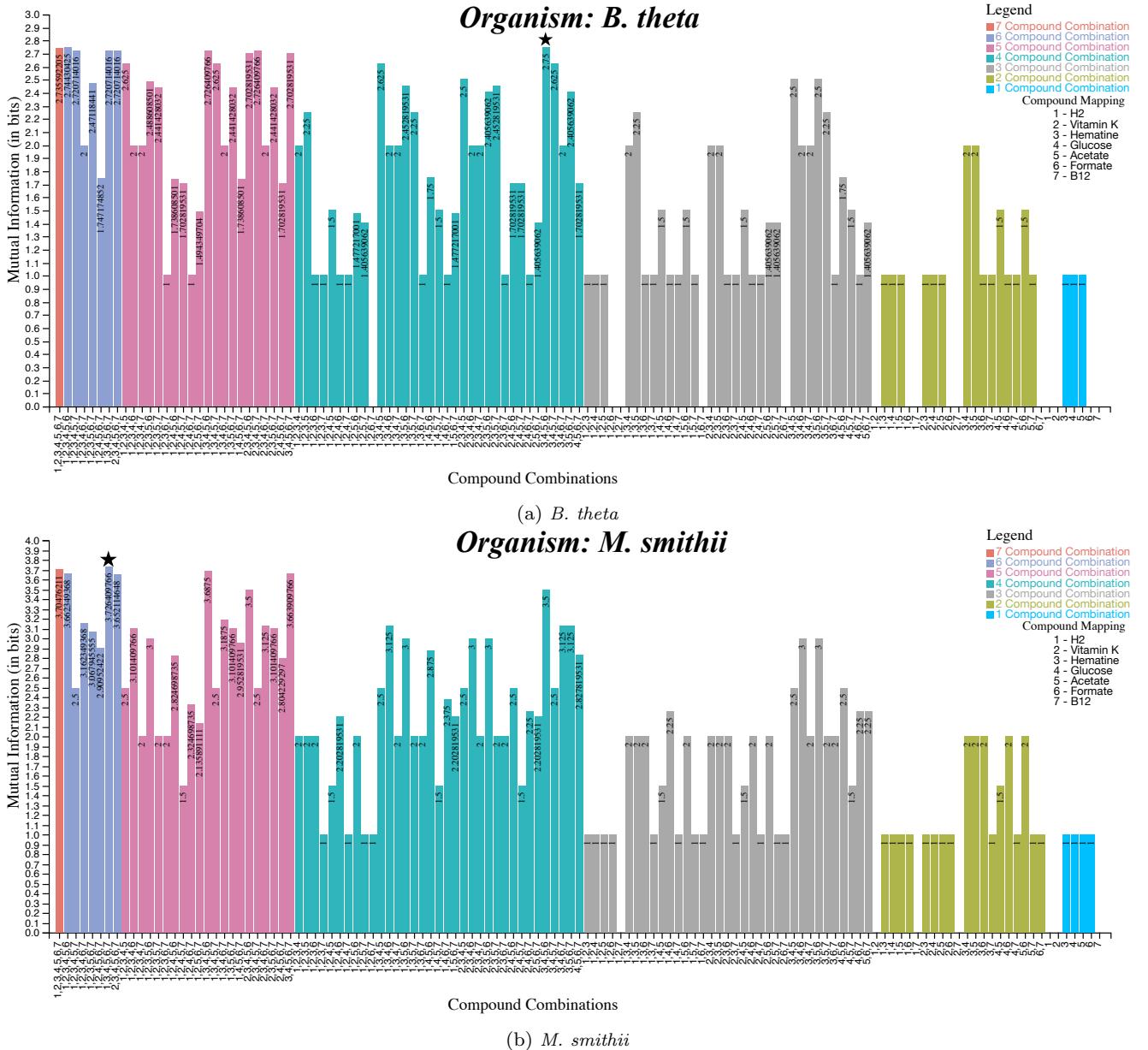


Fig. S12: Upper bounds of the steady-state mutual information for all the different combinations of seven compounds in E2E in with respect to **Biomass only**. The best estimation of the maximum information is starred.

Table S1: v_{max} of minimal media compounds and important 7 compounds.

Compound Name	MaxFlux (mmol/g CDW/hr)
Calcium (Ca^{2+})	0.180254
Chloride (Cl^-)	16.058
Carbon dioxide (CO_2)	34.00204
Cobalt (Co^{2+})	0.042029
Copper (Cu^{2+})	1
Ferrous (Fe^{2+})	0.014
H^+	1
Water (H_2O)	46166.89
Potassium (K^+)	100
L – Cysteine	2.8
Magnesium (Mg)	0.098375
Manganese (Mn^{2+})	0.050529
Sodium (Na^+)	17.564
Ammonium (NH_3)	7.5
Nickel (Ni^{2+})	1
Phosphate	100
Sodium bicarbonate	11.9
Sulfate	7.5
Zinc (Zn^{2+})	1
Ferric (Fe^{3+})	0.014
D-Glucose	2.78
Hematin	0.02
Formate	10
H_2	0.78000078
Vitamin B12	0.0000037
Acetate	10
Vitamin K	5.8