A Molecular Communication model for cellular metabolism

Zahmeeth Sakkaff[1], Andrew P. Freiburger[1], Jennie L. Catlett[2], Mikaela Cashman[3], Aditya Immaneni[3], Nicole R. Buan[2], Myra B. Cohen[4], Christopher Henry[1*], and Massimiliano Pierobon[3]

[1]Argonne National Laboratory, Data Science and Learning Division, Lemont, IL 60148 USA
[2]Department of Biochemistry, University of Nebraska, Lincoln, NE
[3]Department of Computer Science & Engineering, University of Nebraska, Lincoln, NE
[4]Department of Computer Science, Iowa State University, Ames, IA
[*]e-mail: cshenry@anl.gov

July 26, 2023

Abstract

Understanding cellular engagement with its environment is essential to control and monitor metabolism. Molecular Communication theory (MC) offers a computational means to identify environmental perturbations that direct or signify cellular behaviors by quantifying the information about a molecular environment that is transmitted through a metabolic system. We developed an model that integrates conventional flux balance analysis metabolic modeling (FBA) and MC to mechanistically expand the scope of MC, and thereby uniquely blends mechanistic biology and information theory to understand how substrate consumption is captured reaction activity, metabolite excretion, and biomass growth. This is enabled by defining several channels through which environmental information transmits in a metabolic network. The information flow in bits that is calculated through this workflow further determines the maximal metabolic effect of environmental perturbations on cellular metabolism and behaviors, since FBA simulates maximal efficiency of the metabolic system. We exemplify this method on two intestinal symbionts – *Bacteroides thetaiotaomicron* and *Methanobrevibacter smithii* – and visually consolidated the results into constellation diagrams that facilitate interpretation of information flow from given environments and thereby cultivate the design of controllable biological systems. 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 • Metabolism • Flux Balance Analysis • Information Flow

1 Introduction

Engineering biological systems at the cellular-level is essential to realize the Internet of Bio-Nano Things (IoBNT) [1], where wearable bio-computers externally monitor and direct biological systems *in situ* [2]. These external perceptions utilize biosensing [3], optogenetics [4], or magnetic nanoparticles [5], but limitations such as off-target effects on the biological system and inherent biochemical noise compromise accuracy and liability of these technologies [6, 7]. The ideal method would instead leverage native metabolism to minimally disturb the studied system: e.g. ameliorating dysbiosis by changing the chyme nutrient flows via dietary intervention [8, 9]. The lack of basic knowledge of information flow, however, from an environment to behaviors in a metabolic network [10, 11] – post-translation modification, signal transduction [12], gene regulation, and biomass growth [6,10,13] – remains a bottleneck to realize this vision of environmentally controlling native metabolism [14].

Molecular communication (MC) theory in Figure S1 is an emerging confluence of communication and information theories [7, 15–17] that quantifies information flow (or mutual information) of an environment through black-boxed channels of chemical inputs to observed outputs [18]. The information flow IF [19] through each of these channels is quantified in binary digits, or bits (1 or 0),

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

as the excess of input uncertainty $H(inputs)$ from the output uncertainty $H(inputs, outputs)$ from the given inputs. Flux Balance Analysis (FBA) in Figure S2 [20], by contrast, is a metabolic modeling framework that mechanistically predicts reactions fluxes and cellular behaviors of a metabolic system in a given environment [21] through linear optimization. FBA represents a metabolic system as a matrix S of stoichiometries for all metabolites in all reactions and determines the profile of reaction fluxes ($\frac{mmol}{hr*g_{DW}}$) as a vector $\mathbf{v}$ that represents steady-state metabolic activ-

ity ($S\mathbf{v} = 0$) [22]. FBA optimizes for a sub-set of reaction fluxes (conventionally those contributing to biomass growth) and can be tailored through constraints on flux ranges ($\mathbf{v_{min}} \leq \mathbf{v} \leq \mathbf{v_{max}}$) that can represent diverse chemical phenomena: e.g. reaction energetics [23] or regulatory processes [24].

MC and FBA methods offer complementary value for simulating the metabolic effects of an environment, however, these methods have never before been integrated into a single framework that quantifies information flow with mechanistic fortification. We therefore melded MC and FBA into a unique model that computes information flow through defined mechanically-resolved channels of metabolism. We defined a comprehensive MC channel from an input of consumed substrates to externally perceived behaviors such as excretions and biomass growth, and further partitioned this comprehensive channel into two intermediary channels [25]: one whose outputs are cytoplasmic reaction fluxes and one whose outputs are the externally perceived behaviors. These channels are illustrated in Figure S3. We exemplify our model on two [26] human symbionts that are associated with preventing metabolic diseases [27, 28]: the bacterium *B. theta* and the archaeon *M. smithii*. Seven substrates were selected for investigation on their propagation through the metabolism, which represent carbon, nitrogen, sulfur, and oxygen nutrient flows in chyme and have been previously studied with these organisms [26, 29]. The information flow of these substrates through these organisms was visualized in constellation diagrams [30] that elucidate the substrates that optimally transmits information through the metabolic networks. This model expands basic knowledge of cellular behavior as the causal consequence of environmental substrates, and can specifically address basic questions of molecular biology: 1) how much environmental information can a cell encode in its metabolism? and 2) how much information about intra-cellular metabolism can be externally perceived? These basic advances will accelerate biological engineering for myriad applications from healthcare to space colonization [31, 32].

2 Methods

Figure 1 compares each of the MC channels to the biological processes that they represent. The **Stage I** channel

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

defines the activation of all $j \in M$ cytoplasmic reactions $r_j \in [0, 1]$ from the consumption of each substrate $U_i \in [0, 1]$ for all $i \in N$ compounds of interest, over a total of 2^N consumption profiles. The **Stage II** channel

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

defines external excretions $S_e \in [0, 1]$ for all $e \in E$ exchangeable compounds and/or biomass growth $Gr \in [0, 1]$ from the set of active reactions $r_j \in [0, 1]\ \forall\ j \in M$. The **end-to-end** channel is the entire pipeline that joins the **Stage I** and **Stage II** channels. The short-hand for each channel is therefore $inputs = \{U_i\}_{i=1}^{N}$ and $outputs = \{r_j\}_{j=1}^{M}$ for **Stage I**; $inputs = \{r_j\}_{j=1}^{M}$ and $outputs = \{S_e\}_{e=1}^{E}, Gr$ for **Stage II**; and $inputs = \{U_i\}_{i=1}^{N}$ and $outputs = (\{S_e\}_{e=1}^{E}, Gr)$ for **end-to-end**.

2.1 Information flow

The input uncertainty $H(inputs)$ from eq. (1) is generally defined as

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

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

$$\begin{aligned} H(inputs, outputs) = &-\sum_{o=1}^{O} (P(o)) \\ &* \sum (P(inputs, outputs) \\ &* \log_2(P(inputs, outputs)) \end{aligned} \qquad (5)$$

where O is the number of unique outputs. The $P(o)$ term returns the probability of an output among all outputs, while $P(inputs, outputs)$ returns the probability of an output per a given input. The summations in these equations become integrals when the inputs or outputs are continuous, respectively; however, our strict utilization of binary variables discretizes all domains and allows summations.

The presented general expressions for $H(inputs)$ and $H(inputs, outputs)$ are tailored for FBA simulations in several manners. First, the terms are demarcated with an asterisk $\left(\{r_j^*\}_{j=1}^{M}\right)$ to denote that the value corresponds with the maximal information. Second, identical FBA solutions from different consumption profiles, based on their raw (continuous) fluxes, were grouped in Figure S4 to avoid redundant computations of IF^*. Concentration profiles were likewise grouped in Figure S5, where identical substrate profiles across multiple FBA simulations were consolidated.

2.2 Computational Tools and Workflow

The genome-scale models (GEMs) [33] of *B. theta* and *M. smithii*, which are detailed in Figure S6, were constructed through the KBase pipeline [34] in Figure 2. First, the organism's genome sequences from the GenBank database [35, 36] were translated and mapped to functions via KEGG [37]. Second, these initial draft GEMs were gap-filled [38] in a media for each consumption profile (2^N differently gap-filled GEMs), where the minimal number of re-

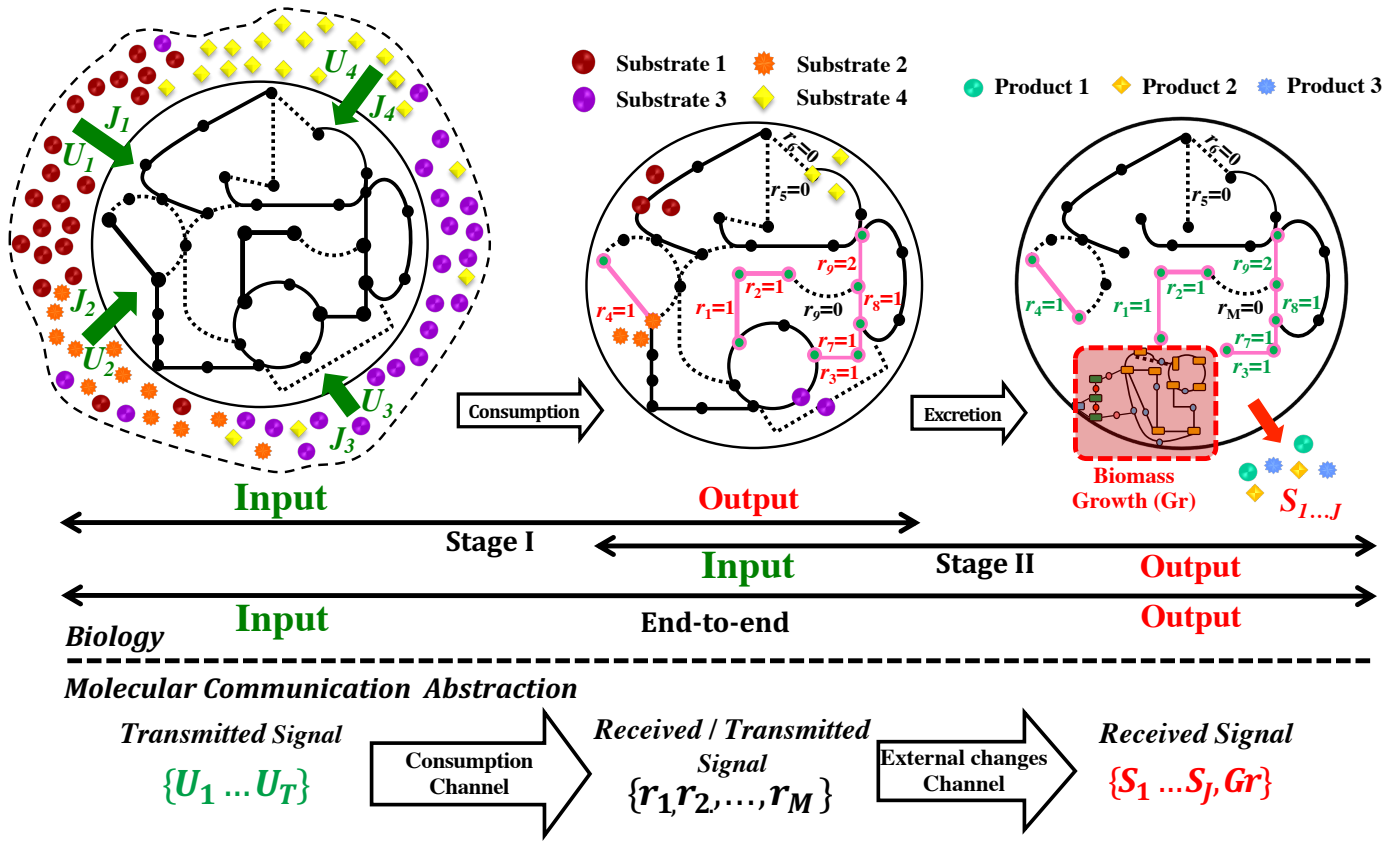


Figure 1: A comparison of biological metabolism and its molecular communication (MC) abstraction, above and below the dashed line, respectively. Biological pipeline depicts a) substrates consumption; b) the activation of metabolic pathways; and c) cellular outcomes such as excreta or biomass growth. The MC pipeline depicts these same phenomena as pairs of inputs and outputs through black-boxed channels: **Stage I** describes the activation of pathways from substrate consumption; **Stage II** describes the manifestation of cellular outcomes from the set of activated pathways; and the **end-to-end** channel describes the combination of **Stage I** and **Stage II** as the full pipeline.

actions that enables growth in the parameterized environment are added to the model. The seven examined compounds ($N = 7$) in our study, similar to other studies with these organisms [26, 29] – glucose (G), hematin (He), formate (F), H_2, Vitamin (B_{12}), acetate (A), and Vitamin K (K) – yielded $2^7 = 128$ consumption profiles for which IF^* will be computed. The 128 unique consumption profiles were added to a standard base media that contained: Ca^{+2}, Mg^{+2}, Cl^-, Na^+, K^+, CO_2, Co^{+2}, Cu^+, Fe^{+2}, Fe^{+3}, H_2O, H^+, L-cysteine, Mn^{+2}, NH_4^+, Ni^{+2}, PO_4^{-3}, SO_4^{-2}, $NaCO_3$, and Zn^{+2}. All lower exchanges bounds of the substrates were -100 while the upper exchange bounds were undefined, which resulted in the substrate consumptions of Table S1. Third, the flux profiles and objective values from simulating these GEMs through FBA were parsed to visualize relationships such as the network comparisons of Hive plots [39] (http://www.hiveplot.com/) in Figures 2 and S7 that better resolve complex relationships than traditional scatter network plots. Hive plots can moreover compare networks by plotting only the node-edge connections that are common between the networks, such as Figure S8 of common metabolite-reaction connections between two FBA solution groupings.

3 Numerical Results

The maximal information flow IF^* for each channel was determined by computing each term in eq. (1) according to the Stage channels that are defined in eqs. (2) and (3). The assumed equal probability of the $N = 7$ examined substrates is $P(\{U_G, U_{He}, U_F, U_{H_2}, U_{B_{12}}, U_A, U_K\}) = \frac{1}{128}$. The input uncertainty in eq. (4) for Stage I and the 128 possible inputs was computed: $H(inputs) = -1 * \log_2(128) = 7\ bits$. The conditional output uncertainty for Stage I of *B. theta* was calculated, from 113 unique reaction activation profiles ($M = 113$) over 14 unique FBA solutions,

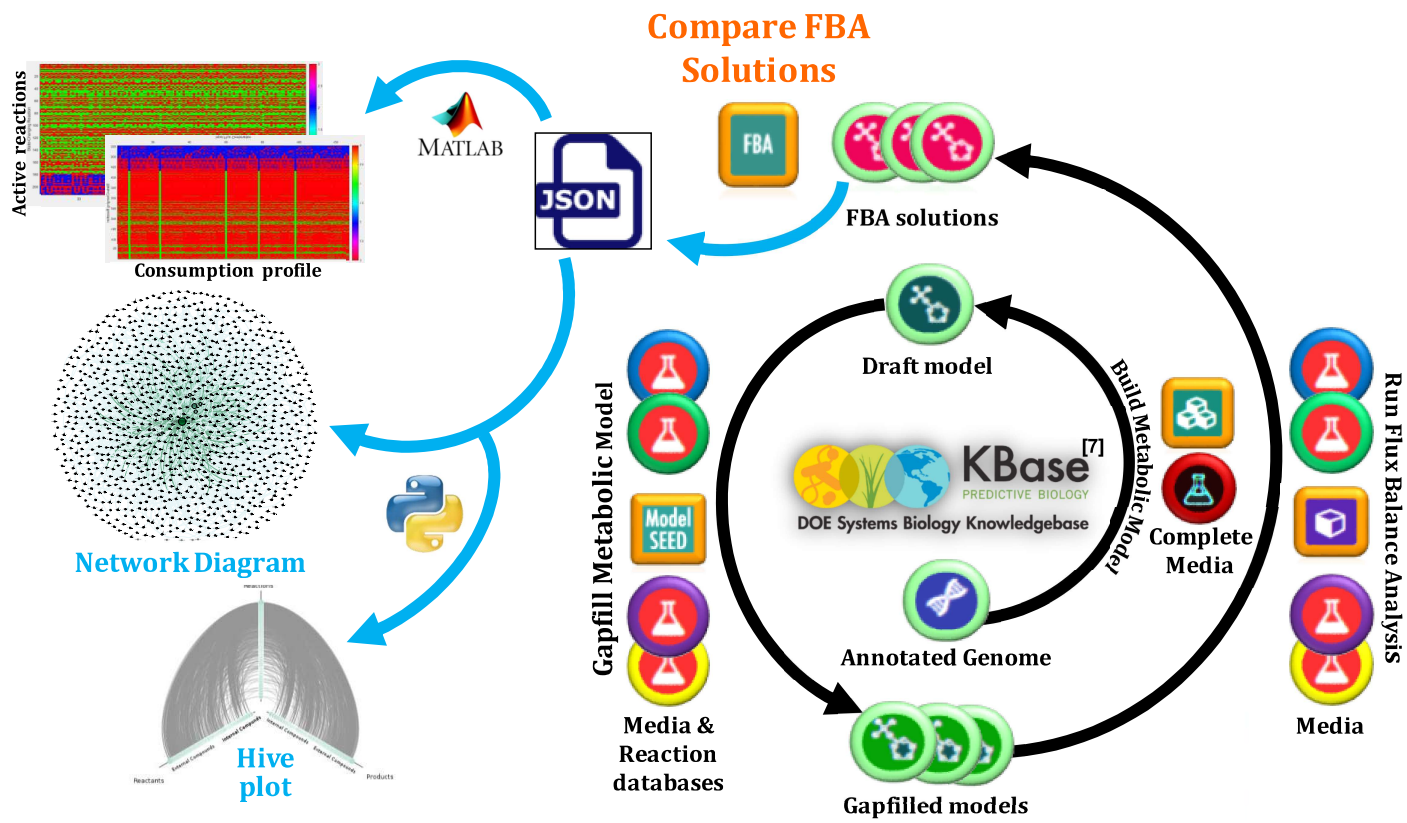


Figure 2: A summary of our modeling workflow. The initial steps leverage KBase as an open-source platform for conducting simulations and hosting our Application. Our KBase workflow involves: 1) annotating uploaded genomes through RAST or DRAMM; 2) reconstructing draft genome-scale metabolic models (GEMs) 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. Custom MATLAB and Python scripts then process the exported JSON of compared FBA solutions into heatmaps that illustrate activated pathways and into network diagrams and hive plots to illustrate input-output relationships in the metabolism, respectively.

$$
\begin{aligned}
H(inputs, outputs) &= \sum (P(inputs, outputs) * \sum (P(inputs, outputs) * \log_2(P(inputs, outputs))) \\
&= -\sum_{e}^{T} (P(e) * \sum_{g,e}^{G,E} (P(e \in g) * \log_2(P(e \in g))) \\
&= -\Bigg[P(FBA_1) * \sum_{e}^{E} (P(FBA_{1e}) log_2(P(FBA_{1e}))) + P(FBA_2) * \sum_{e}^{E} (P(FBA_{2e}) log_2(P(FBA_{2e}))) \\
&\quad + \cdots + P(FBA_{14}) * \sum_{e}^{E} (P(FBA_{14e}) log_2(P(FBA_{14e}))) \Bigg] \\
&= -\Bigg[\frac{6}{128} * (6 * \frac{1}{6} log_2(\frac{1}{6})) + \frac{28}{128} * (28 * \frac{1}{28} log_2(\frac{1}{28})) + 2 * \frac{5}{128} * (5 * \frac{1}{5} log_2(\frac{1}{5})) \\
&\quad + 2 * \frac{7}{128} * (7 * \frac{1}{7} log_2(\frac{1}{7})) + \frac{32}{128} * (32 * \frac{1}{32} log_2(\frac{1}{32})) + \frac{11}{128} * (11 * \frac{1}{11} log_2(\frac{1}{11})) \\
&\quad + 2 * \frac{3}{128} * (3 * \frac{1}{3} log_2(\frac{1}{3})) + \frac{9}{128} * (9 * \frac{1}{9} log_2(\frac{1}{9})) + 3 * \frac{4}{128} * (4 * \frac{1}{4} log_2(\frac{1}{4})) \Bigg]
\end{aligned}
$$

$$= -\left[-0.1212 + -1.0516 + -0.1814 + -0.3071 + -1.25 + -0.2973 + -0.0743 + -0.2229 + -0.1875\right] = 3.693 \approx 3.7\ bits\ ,$$

where the uncertainty of each FBA group is separately calculated and then aggregated into the total output uncertainty. The grouping of the 128 FBA solutions based on reaction fluxes in Figure S4 importantly simplifies the above computation and can facilitate biological investigation in relating consumed substrates to metabolic activation and cellular behaviors. The equivalent probability of a FBA group consumption profile (element e) among all other consumption profiles in the total set of consumption profiles (T) or in the same FBA group (g) – $\frac{1}{T\ or\ g}$ – allows the terms for all elements to be condensed with a coefficient of the number of elements in the group: e.g. the total uncertainty for the 6 elements of FBA group 1 (FBA_1) is $\frac{6}{128} * (\mathbf{6} * \frac{1}{6} log_2(\frac{1}{6}))$, where the coefficient **6** collapses the six element terms into a single term. Several FBA groups also possessed the same number of elements, which permitted further simplification with a coefficient of the number of identically sized FBA groups, where all FBA groups are themselves equally probable: e.g. FBA groups 3 and 4 each possess five elements and can therefore be jointly defined as $\mathbf{2} * \frac{5}{128} * (5 * \frac{1}{5} log_2(\frac{1}{5}))$, with the coefficient **2** denoting the two FBA groups.

The IF^* for Stage I of *B. theta* then finally equates $H(inputs) - H(inputs, outputs) = 7 - 3.7 = 3.3\ bits$. The conditional output uncertainty for Stage I of *M. smithii* was computed to be 2.5 *bits*, analogously to the above derivation for *B. theta* except with 136 unique reaction activation profiles ($M = 136$) over 31 unique FBA solutions. The IF^* for Stage I of *M. smithii* then equates $7 - 2.5 = 4.5\ bits$. These results are depicted Figure S9, where *B. theta* reaches its maximum information flow in the absence of formate while *M. smithii* reaches its maximum information flow with all seven examined substrates.

The FBA groupings that were defined for the Stage I calculations served as the inputs for Stage II and were used to compute the input uncertainty: $-1 * \log_2(14) = 3.8\ bits$ for *B. theta* and $-1 * \log_2(31) = 5.0\ bits$ for *M. smithii*. The FBA solutions were then regrouped according to the respective output – Gr ($outputs = Gr$) or both $S_e\ \forall\ e \in E$ and Gr ($outputs = (S_e\ \forall\ e \in E, Gr)$) – to calculate the conditional output uncertainty of the Stage II channel. Grouping by biomass growth ($outputs = Gr$) in Figure 3a separately evaluated the probability of each group g in the set of all FBA groups G. The Stage II IF^* for biomass growth is therefore $3.8 - 1 = 2.8\ bits$, and analogously $IF^* = 5 - 1.1 = 3.9\ bits$ for *M. smithii*. Grouping by both excretions and biomass growth manifested in $IF^* = 3.8 - .1 = 3.7$ for *B. theta* and $IF^* = 5.0 - 1.1 = 3.9$ for *M. smithii*. The modest increase in IF^* for *B. theta* when including metabolic excretion to the outputs, and the unchanged IF^* of *M. smithii*, is counter-intuitive, since more outcomes would seem to capture more inputs; evidently, biomass growth captures most of the information flow during substrate consumption through metabolism.

$$\begin{aligned}
H(inputs, outputs) &= \sum(P(inputs, outputs) * \sum(P(inputs, outputs) * \log_2(P(inputs, outputs))) \\
&= -\sum_g^G (P(g) * \sum_e^E (P(e \in g) * \log_2(P(e \in g))) \\
&= -\left[4 * \frac{1}{14} * (1 * \frac{1}{1} log_2(\frac{1}{1})) + 3 * \frac{2}{14} * (2 * \frac{1}{2} log_2(\frac{1}{2})) + \frac{4}{14} * (4 * \frac{1}{4} log_2(\frac{1}{4}))\right] \\
&= -\left[0 + -0.4286 + -0.5714 = 1.000 \approx 1\ bit\ ,\right.
\end{aligned}$$

(a) Stage II computation for *B. theta* and $outputs = Gr$

$$\begin{aligned}
H(inputs, outputs) &= \sum(P(inputs, outputs) * \sum(P(inputs, outputs) * \log_2(P(inputs, outputs))) \\
&= -\sum_g^G (P(g) * \sum_e^E (P(e \in g) * \log_2(P(e \in g))) \\
&= -\left[\frac{1}{128} * (1 * \frac{1}{1} log_2(\frac{1}{1})) + \frac{2}{128} * (2 * \frac{1}{2} log_2(\frac{1}{2})) + \frac{5}{128} * (5 * \frac{1}{5} log_2(\frac{1}{5})) \right. \\
&\quad + \frac{7}{128} * (7 * \frac{1}{7} log_2(\frac{1}{7})) + 7 * \frac{8}{128} * (8 * \frac{1}{8} log_2(\frac{1}{8})) + \frac{9}{128} * (9 * \frac{1}{9} log_2(\frac{1}{9})) \\
&\quad \left. + 3 * \frac{16}{128} * (16 * \frac{1}{16} log_2(\frac{1}{16}))\right] \\
&= -\left[0 + -0.0156 + -0.0907 + -0.1535 + -1.3125 + -0.2229 + -2.585\right] = 3.2952 \approx 3.3\ bits
\end{aligned}$$

(b) End-to-end computation for *M. smithii* and $outputs = Gr$

Figure 3: IF^* computations for Stage II and end-to-end channels for the two examined organisms.

The end-to-end channel – $inputs = 128\ consumption\ profiles$, $outputs = Gr$, and FBA groups according to Gr – was additionally computed to be $IF^* = 7 - 4.3 = 2.7\ bits$ for *B. theta* and $IF^* = 7 - 3.3 = 3.7\ bits$ for *M. smithii* in Figure S10, where the conditional output was computed in fig. 3b for *M. smithii*. The subtly lower IF^* values are expected, since the larger scope of this channel introduces more noise and loss that hinder reception of the signal in the final outputs.

3.1 Visualizations

The constellation diagrams of Figure 5 statistically compress the IF^* from all consumption profiles for rapid interpretation. These diagrams were constructed by first: 1) grouping consumption profiles by IF^* from the end-to-end channel; 2) determining the minimum and maximum IF^* for each consumption profile; and 3) graphing each of the minimum, maximum, and average IF^* (excluding the minimum and maximum) by the size of the consumption profile. These diagrams interestingly revealed that the maximum IF^* plateaus with ≥ 5 of the examined substrates, which suggests that there are diminishing returns to cellular control from environmental perturbations. This trend is consistent with simulations from the BioSIMP software that identified substrates for growth of *B. theta* and *M. smithii* [29]. This trend is moreover echoed by Figures 4, S9, and S10 that reveal relatively small consumption profiles with remarkably high IF^*: e.g. consuming F, B_{12}, and A for *B. theta* and *M. smithii*; or consuming F, H_2, A for *M. smithii*. These

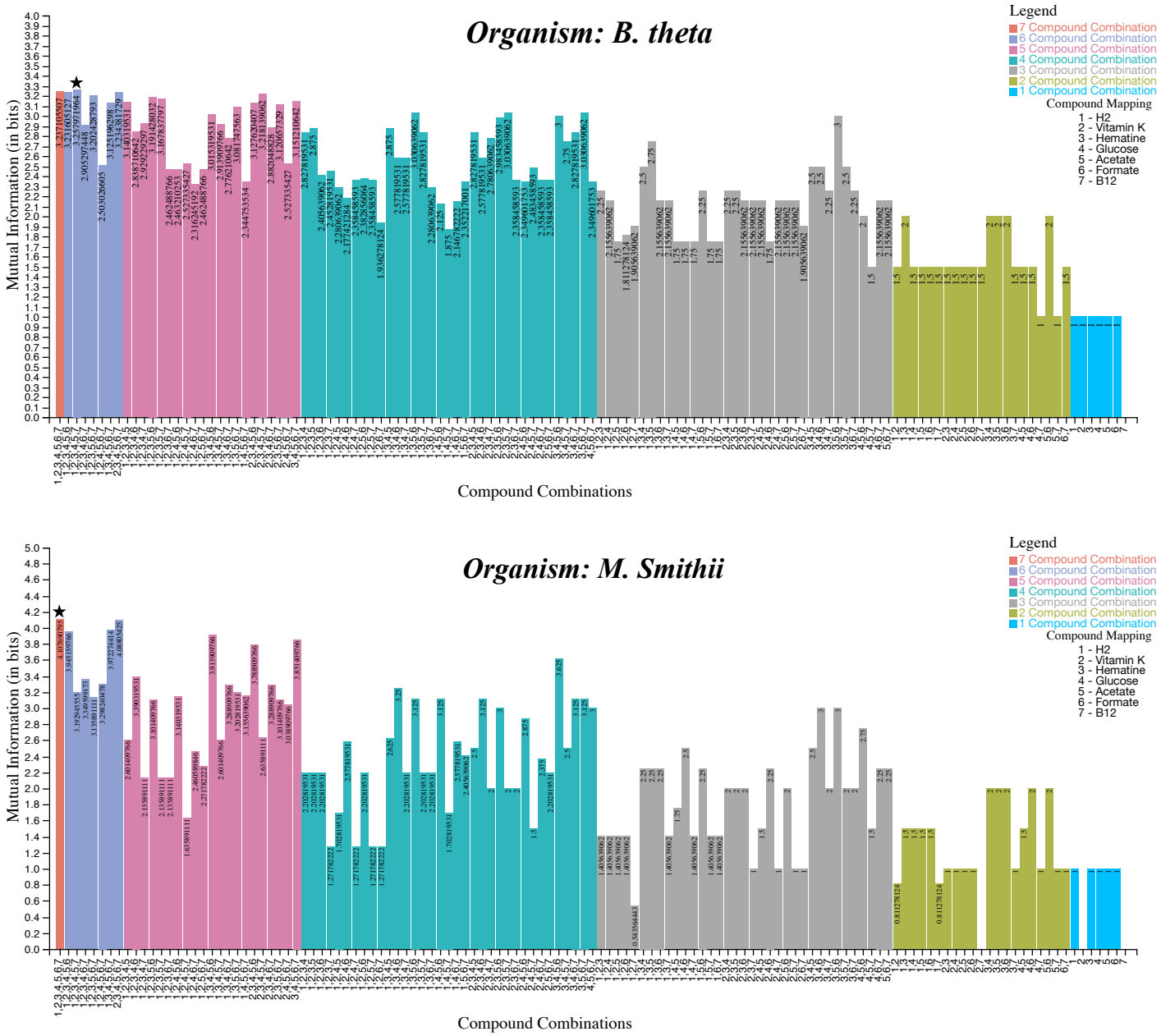
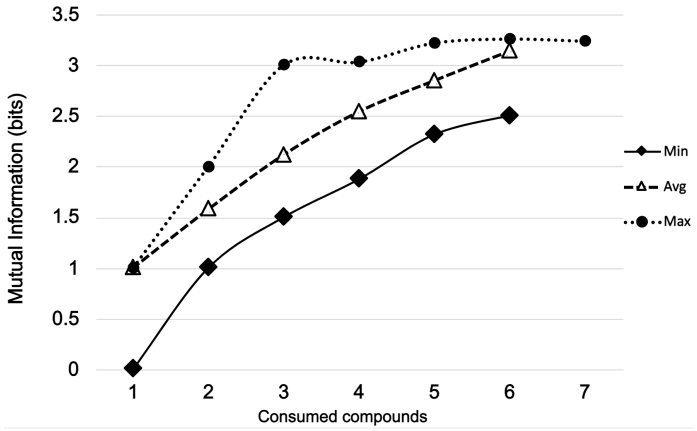


Figure 4: The end-to-end upper bounds for each consumption profile. The maximal information flow is starred. The consumption profiles with more compounds predictably manifested in greater potential information flow. The consumption profiles are condensed on the x-axis by numbering each substrate: 1 = glucose (G), 2 = hematin (He), 3 = formate (F), 4 = H_2, 5 = Vitamin (B_{12}), 6 = acetate (A), and 7 = Vitamin K (K).

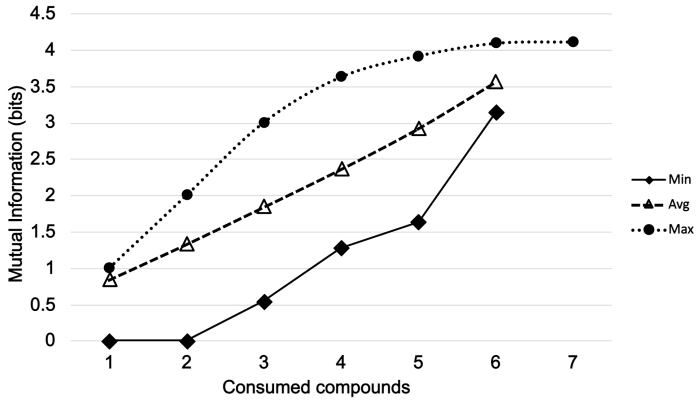
basic insights from various substrate conditions would be necessary for biomedical monitors and other systems that use molecular transport to categorize metabolic states.

4 Conclusions and Future Work

Molecular communication theory quantifies the flow of information from extracellular environments into metabolic activity, substrate excretions, or biomass growth. We merge this concept with mechanistic FBA, and defined several metabolic channels through which environmental information likely flows in a cell, to compose a unique model for quantifying optimal information flow through metabolic systems. Our model and constellation diagrams of the results identified consumption profiles that optimally transmitted information biochemically to cellular outcomes, such as biomass growth and chemical excretions, which can minimize the number of resource-intensive experiments that are needed to find substrate profiles that desirably control cellular behaviors. Hive plot visualizations of our results further conveyed connectivity between inputs and out-



(a) *B. theta*



(b) *M. smithii*

Figure 5: Constellation diagrams for the information flow from all consumption profiles in the end-to-end channel. The minimum, maximum, and average values are plotted against the number of compounds in the consumption profiles, which reveals that the maximum IF^* increases logarithmically while the average IF^* increases linearly. This observation is intuitive where biochemical noise over the channel defines a fundamental limit to IF^* but generally an increase in the degrees of freedom (consumed substrates) increases the uncertainty (IF^*) of a system.

puts, which empowers more detailed analyses. The Code, KBase Narratives, and additional details are available in our GitHub repository (https://github.com/freiburgermsu/MetabolicMC_Supplementary_Git_v1).

This model advances basic biological knowledge of cellular engagement with its environment, and provides a tool to identify environmental substrate combinations that can most effectively engineering and monitor cellular behavior. This tool has diverse applications in biomedical wearable technologies and ecological control through benign substrate perturbations. Subsequent work will refine our model by 1) better compensating biochemical noise; 2) capturing dynamic information flows and continuous variables via the dynamic FBA algorithm; 3) validating predictions with experimental data; and 4) exploring applications in autonomous laboratories that predictably monitor and control cellular behaviors through environmental perturbations.

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. Koucheryavy, "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. Charlton, and V. A. Leone, "Imbalanced gut microbiota predicts and drives the progression of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in a fast-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] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophysical Journal*, vol. 92, no. 5, pp. 1792–1805, 2007.

[24] 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.

[25] 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.

[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. Togneri and J. Christopher, *Fundamentals of information theory and coding design*. CRC Press, 2003.

[31] Lehner, "To new frontiers, microbiology for nanotechnology and space exploration," 2019.

[32] A. J. Berliner, J. M. Hilzinger, A. J. Abel, M. J. McNulty, G. Makrygiorgos, N. J. Averesch, S. S. Gupta, A. Benvenuti, D. F. Caddell, S. Cestellos-Blanco, A. Doloman, S. Friedline, D. Ho, W. Gu, A. Hill, P. Kusuma, I. Lipsky, M. Mirkovic, J. L. Meraz, V. Pane, K. B. Sander, F. Shi, J. M. Skerker, A. Styer, K. Valgardson, K. Wetmore, S. G. Woo, Y. Xiong, K. Yates, C. Zhang, S. Zhen, B. Bugbee, D. S. Clark, D. Coleman-Derr, A. Mesbah, S. Nandi, R. M. Waymouth, P. Yang, C. S. Criddle, K. A. McDonald, L. C. Seefeldt, A. A. Menezes, and A. P. Arkin, "Towards a biomanufactory on mars," *Frontiers in Astronomy and Space Sciences*, vol. 8, 7 2021.

[33] G. J. Baart and D. E. Martens, "Genome-scale metabolic models: reconstruction and analysis," *Neisseria meningitidis: Advanced Methods and Protocols*, pp. 107–126, 2012.

[34] KBase, "Department of Energy Systems Biology Knowledgebase (KBase)." http://kbase.us. [Online; accessed 22-September-2016].

[35] 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.

[36] 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.

[37] KEGG, "Kyoto encyclopedia of genes and genomes (KEGG)." http://www.genome.jp/kegg/. [Online; accessed 22-September-2016].

[38] 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.

[39] 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.

Supplementary Information

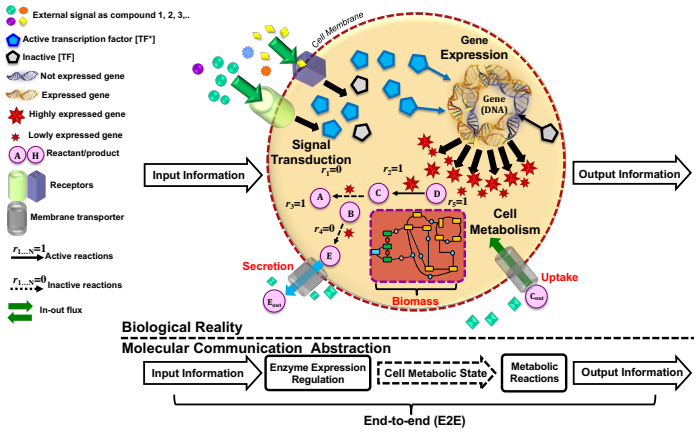
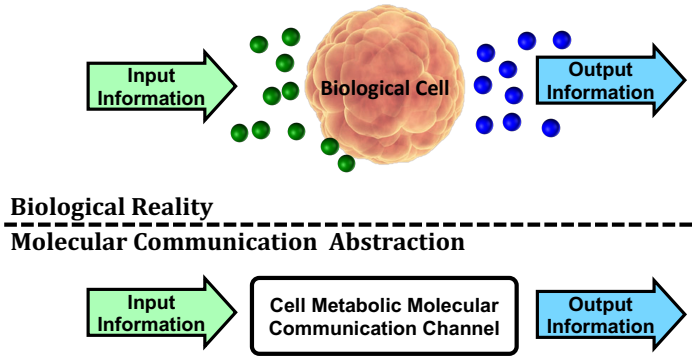
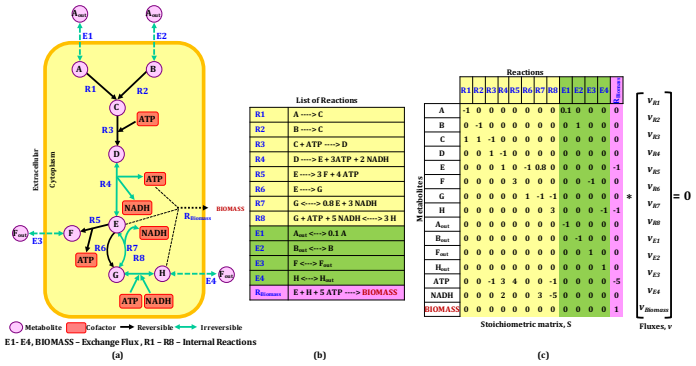


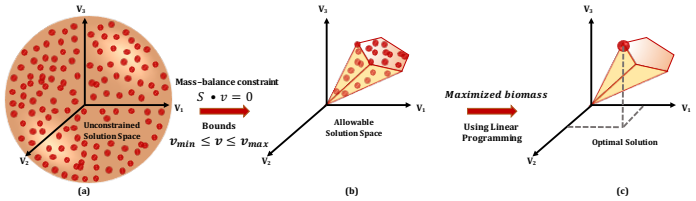
Figure S1: Biological reality of Molecular communication in cell metabolism.



(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 $a'\mathbf{v}$ identifies a flux distribution on the edge of the allowable solution space.

Figure S2: 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.

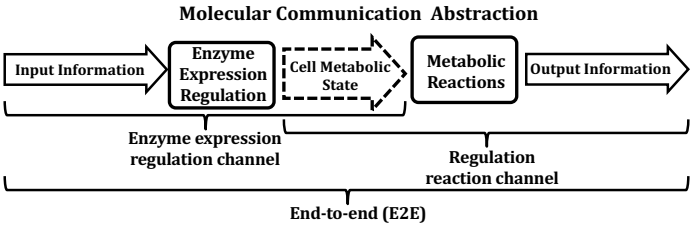
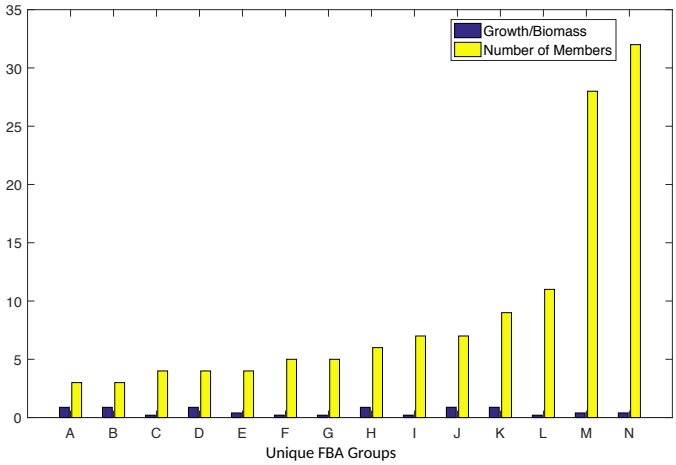
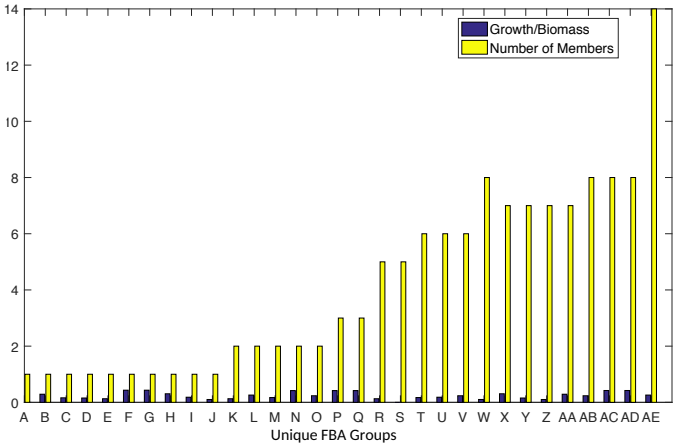


Figure S3: Molecular communication channel abstraction of cell metabolism.

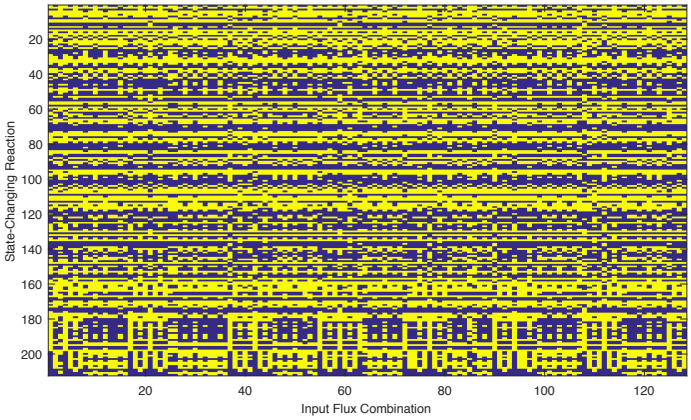


(a) 14 unique FBA groups for *B. theta*

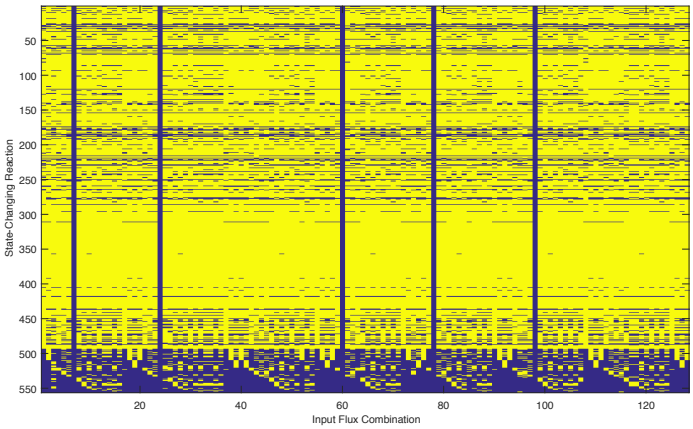


(b) 31 unique FBA groups for *M. smithii*

Figure S4: 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 S5 relative to *B. theta*.



(a) *B. theta*



(b) *M. smithii*

Figure S5: Reaction activation from FBA simulations $\{r_i^*\}_{i=1}^M$ for each possible consumption profile, where yellow is active and violet is inactive. *M. smithii* metabolism is evidently much more activated than *B. theta* by the examined media, notwithstanding a handful of media profiles that did not support growth (illustrated as vertical lines), 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*.

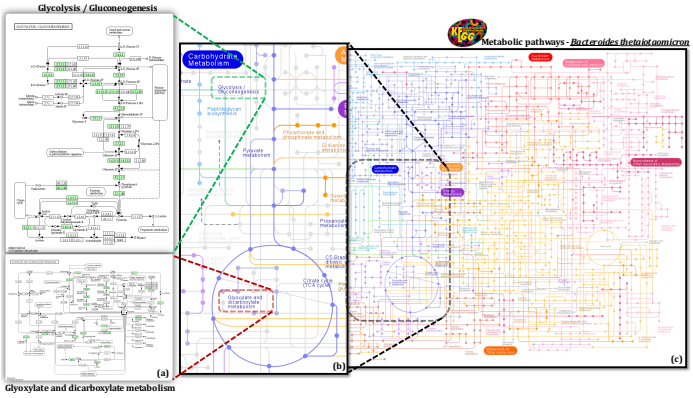


Figure S6: KEGG module of *Bacteroides thetaiotaomicron* metabolic pathways. **(a)** Summary of the biological processes shown in the pathway map of Glycolysis / Gluconeogenesis and Glyoxylate and dicarboxylate metabolism. **(b)** Enlarged fine details of a section of a complete metabolic model. **(c)** 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 [37]. 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).

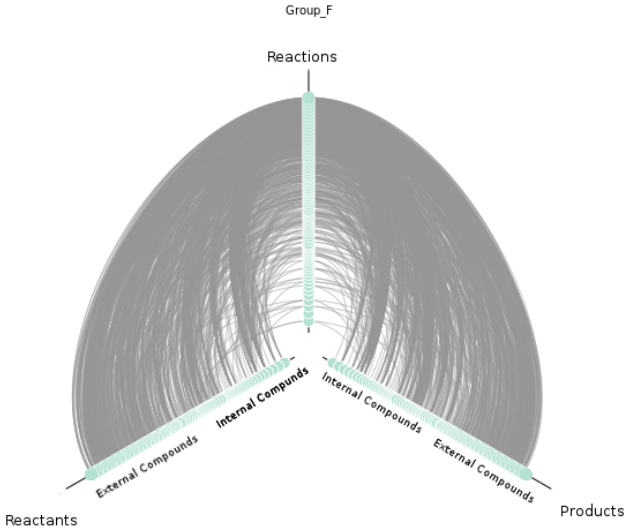


Figure S7: A hive plot for a configuration F is shown in the figure. The reactions are placed on the Z axis, the reactants on the X axis and the products on the Y axis. Further the External compounds are placed higher on the X and Y axes than the Internal compounds.

Table S1: $\mathbf{v_{max}}$ 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 ($C0_2$)	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

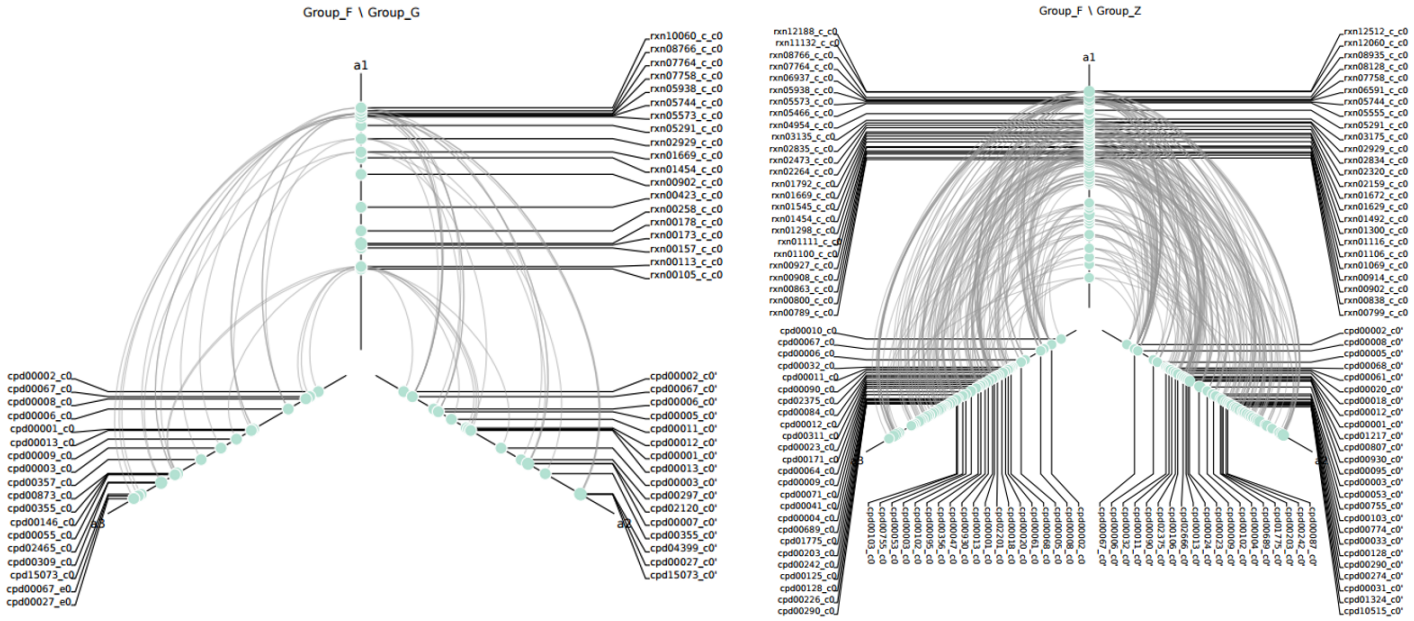
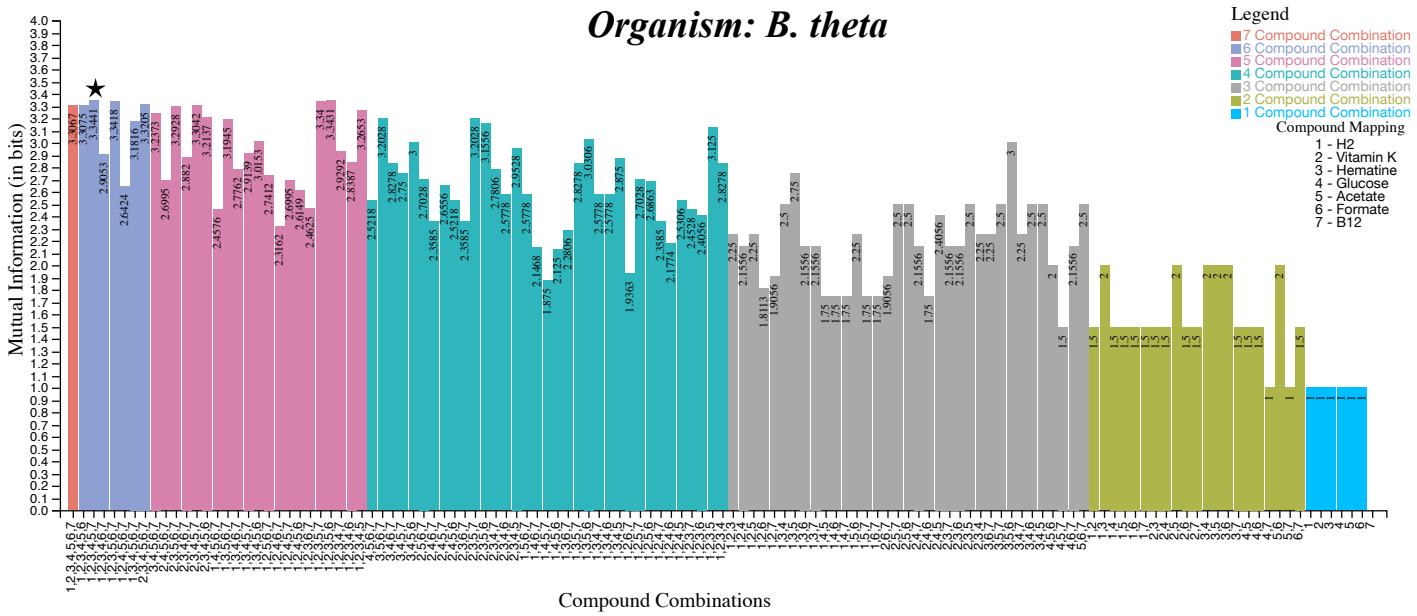


Figure S8: In this figure we take 3 different configuration of state changing reactions, labelled F, G and Z. It shows differential hive plots of F vs G and F vs Z. The groups F and G in F vs G hive plot has the same biomass whereas, the groups F and Z in F vz Z hive plot have the least and highest biomass 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.



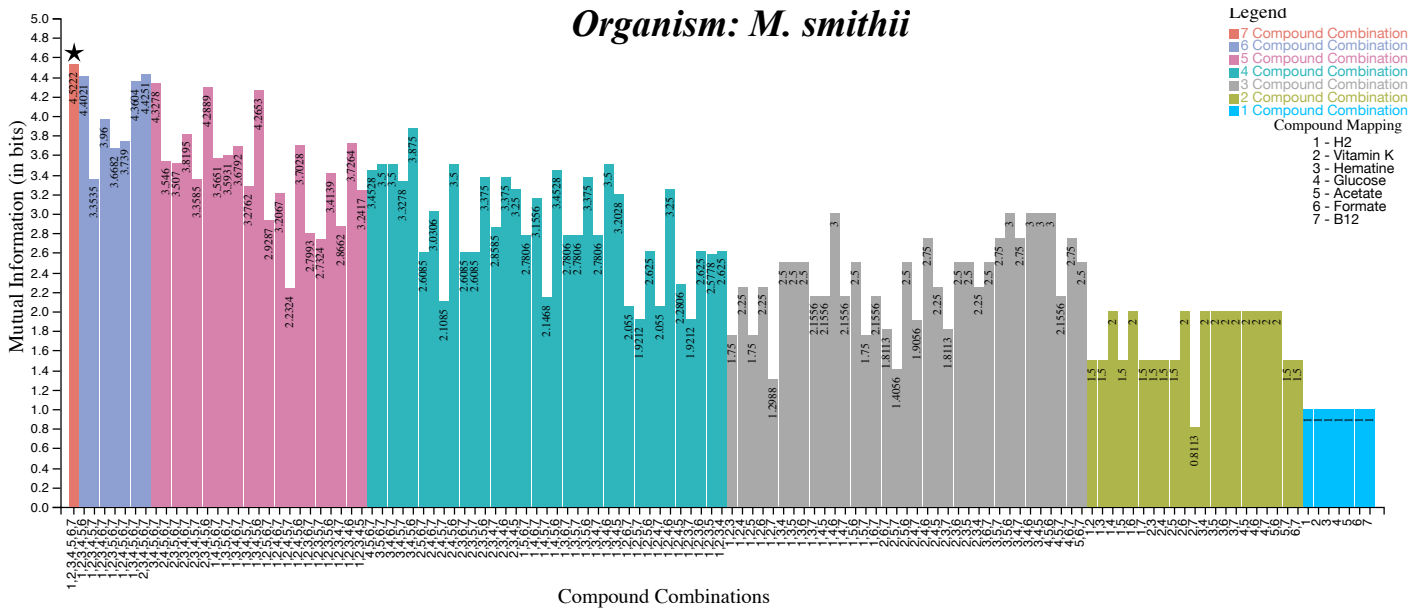
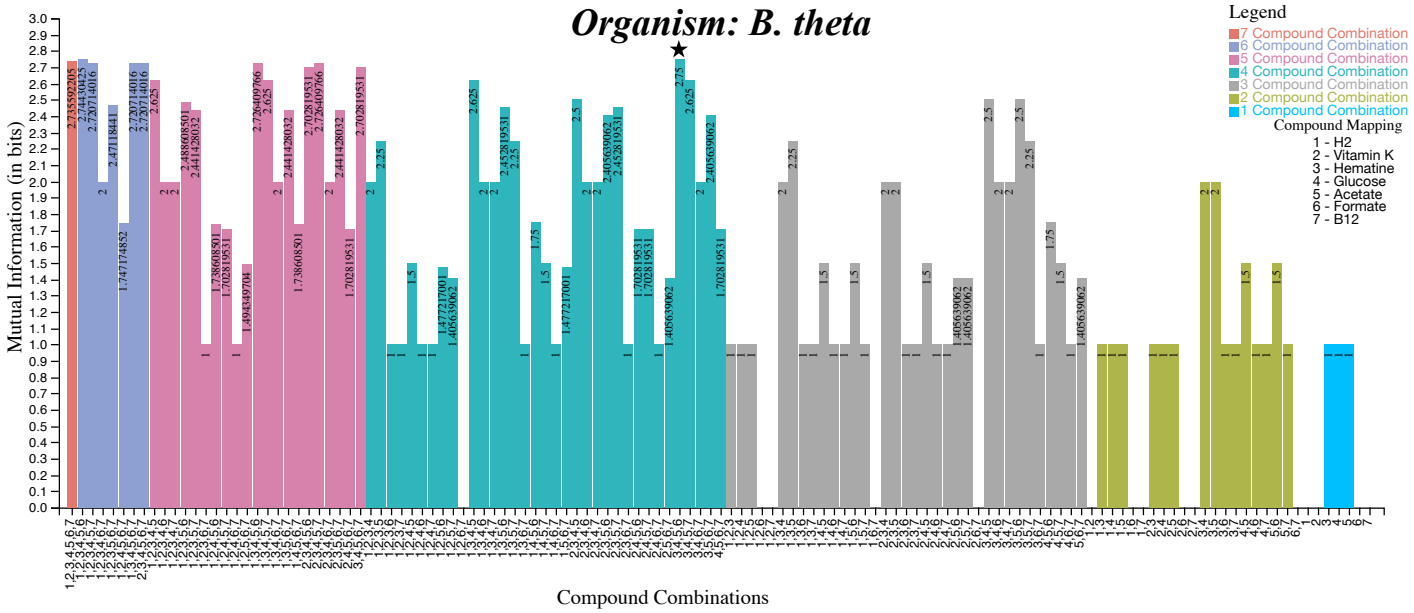


Figure S9: Upper bounds of the steady-state mutual information for all the different combinations of seven compounds in Enzyme Expression Regulation channel in . The best estimation of the maximum information is starred.



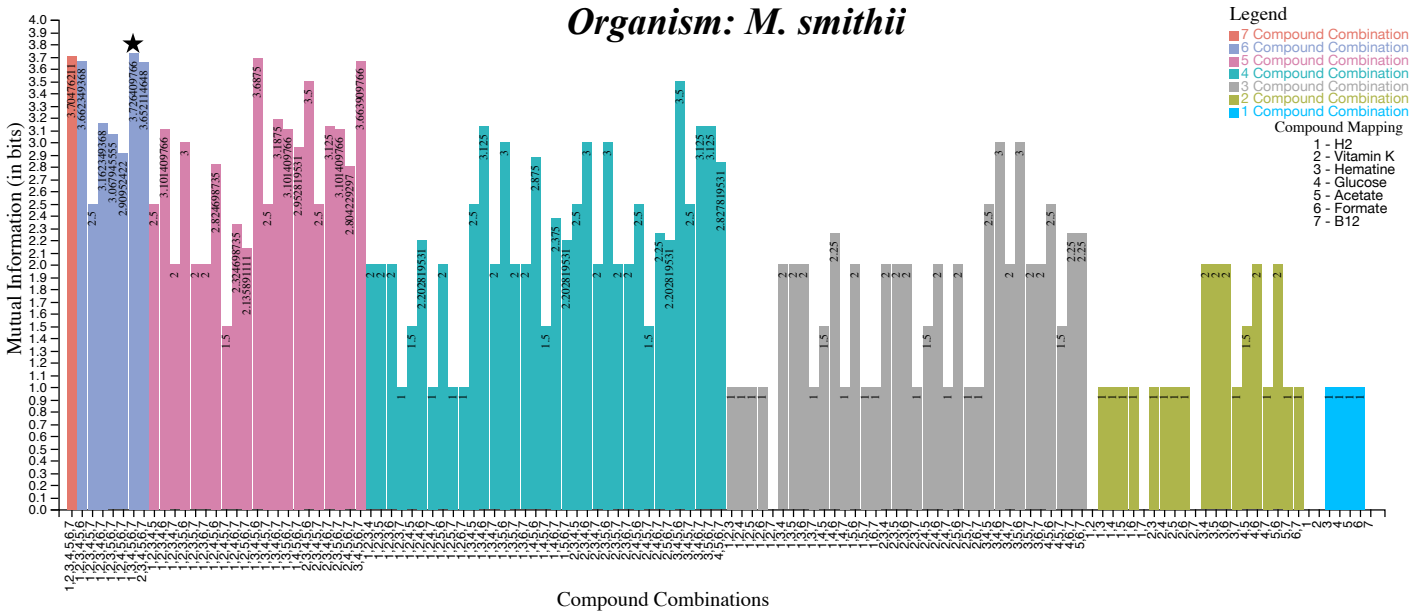


Figure S10: Upper bounds of the steady-state mutual information for all the different combinations of seven compounds in E2E in with respect to **Biomass only**. The best estimation of the maximum information is starred.

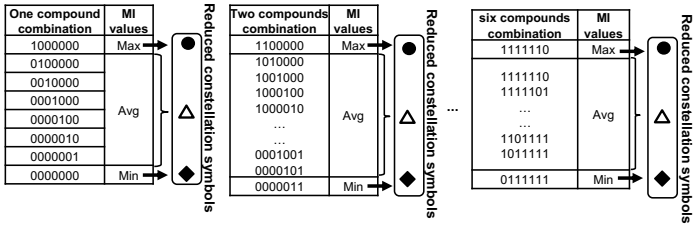


Figure S11: Steps involved in constructing constellation diagram.