Systems Biology

A hierarchy of metabolite exchanges in metabolic models of microbial species and communities

Ylva Katarina Wedmark [1,2], Jon Olav Vik [1,2] and Ove Øyås [1,2,*]

[1] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway.
[2] Faculty of Biosciences, Norwegian University of Life Sciences (NMBU), Ås, Norway. *Contact: ove.oyas@nmbu.no.

Abstract

Motivation: The metabolic network of an organism can be analyzed as a constraint-based model (CBM). CBM analyses can be biased, optimizing an objective such as growth rate, or unbiased, attempting to describe the feasible flux space through pathway analysis or random flux sampling. Biased methods scale to genome-scale models (GEMs) and beyond, but unbiased methods suffer from combinatorial explosion. One promising approach to improve the scalability of unbiased analyses is to focus on cells' metabolite exchanges with each other and the environment rather than the internal details of their metabolic networks. However, methods for unbiased analysis of metabolite exchanges have not been evaluated or compared.
Results: Using CBMs ranging in size from core models to GEMs, we applied pathway enumeration and random flux sampling to metabolite exchanges in microbial species and a microbial community. We focused on pathway definitions that allow direct targeting of metabolite exchanges: elementary conversion modes (ECMs), elementary flux patterns (EFPs), and minimal pathways (MPs). Enumerating growth-supporting ECMs, EFPs, and MPs for the same models, we found that the MPs were always a subset of the ECMs and EFPs, specifically the subset of support-minimal pathways. The EFPs were always a subset of the ECMs, which in turn were always a subset of flux patterns from elementary flux modes (EFMs). We argue that this hierarchical relationship between pathway definitions holds more generally and that it helps explain the relative scalability of ECM, EFP, and MP enumeration. Metabolite exchange frequencies in enumerated pathways were similar across methods with a few specific exchanges explaining large differences in pathway counts. Thus, biological interpretation was largely robust to the choice of pathway definition. Random flux sampling scaled to GEMs and complemented pathway enumeration.
Availability: Data and code are available at https://gitlab.com/YlvaKaW/exchange-enumeration.

1 Introduction

Metabolic pathways are combinations of biochemical reactions that occur in a cell or organism, and the interplay between pathways forms the cell or organism's metabolic network (Nielsen and Keasling, 2016). The growing availability of genomes and other omics data has enabled metabolic network reconstruction *in silico*, giving rise to genome-scale metabolic models (GEMs) that are usually formulated as constraint-based models (CBMs) to allow scaling (Bordbar *et al.*, 2014). A CBM describes a metabolic network with m metabolites and n reactions as an $m \times n$ stoichiometric matrix in which each element is the stoichiometric coefficient of a metabolite in a reaction. By making the quasi-steady-state assumption, justified by the fact that metabolism is very fast compared to other biological processes (Reimers and Reimers, 2016), the mass balances of the metabolites can be written as

$$\mathbf{N}\mathbf{r} = 0, \tag{1}$$

where $\mathbf{N}$ is the stoichiometric matrix, and $\mathbf{r}$ is the vector of fluxes (reaction rates). Solving for $\mathbf{r}$ yields a flux distribution that satisfies this linear system of equations, i.e., a feasible combination of reaction rates.

Without any other limitations on fluxes, there are infinitely many flux distributions that satisfy (1) and thus form the null space of $\mathbf{N}$. However, some of these solutions are not realistic because of physical, chemical, or environmental limitations. This can be taken into account by additional linear constraints, most commonly lower and upper bounds on fluxes:

$$r_i^{\text{lb}} \leq r_i \leq r_i^{\text{ub}}, \tag{2}$$

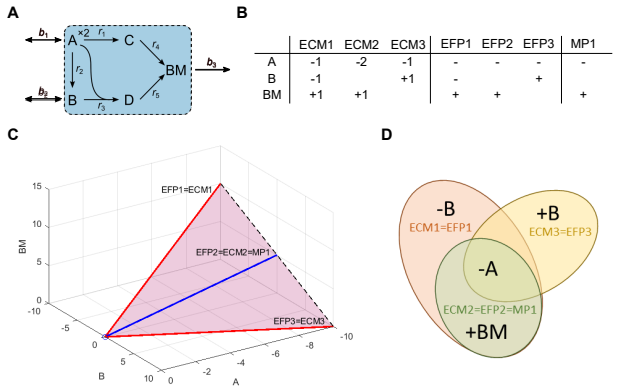
where r_i^{lb} and r_i^{ub} are the lower and upper flux bounds of reaction i, respectively. Geometrically, these bounds are hyperplanes that eliminate infeasible solutions by slicing the feasible flux space. In the simplest case, bounds are only applied to irreversible reactions to ensure flux in one direction, i.e., $r_i^{\text{lb}} = 0$ or $r_i^{\text{ub}} = 0$. These constraints are homogeneous, meaning that the right-hand-side is always zero. They slice the null space to a cone that can be further sliced to a more general polyhedron by adding non-zero bounds, i.e., $r_i^{\text{lb}} \leq r_i^{\text{ub}} \neq 0$ or $r_i^{\text{ub}} \geq r_i^{\text{lb}} \neq 0$, or other inhomogeneous constraints (Klamt *et al.*, 2017).

The feasible flux space of a CBM can be analyzed by biased or unbiased methods (Lewis *et al.*, 2012). Biased methods such as flux balance analysis (FBA) focus on specific flux distributions that optimize

an assumed cellular objective (Orth *et al.*, 2010). For microbes, it can be reasonable to assume that the objective of a cell is to grow as fast as possible, i.e., to maximize growth rate as modeled by a biomass reaction, but this assumption does not hold for all conditions or cells (Simensen *et al.*, 2022). Unbiased methods circumvent this by characterizing the entire solution space through random flux sampling (Schellenberger and Palsson, 2009) or pathway analysis (Schilling *et al.*, 1999). In flux sampling, Monte Carlo (MC) methods are used to draw random solutions uniformly from the feasible flux space, and the resulting samples provide flux probability distributions (Fallahi *et al.*, 2020). Pathway analysis generally involves enumeration of formally defined metabolic pathways, e.g., elementary flux modes (EFMs) or elementary flux vectors (EFVs), which are the minimal conformal generators of the flux cone and flux polyhedron, respectively (Klamt *et al.*, 2017). In other words, EFMs and EFVs are sets of vectors that can combine into all possible flux routes through a network without cancellation of fluxes. Biased methods scale to GEMs and combinations thereof, but unbiased approaches are often limited to smaller CBMs due to combinatorial explosion. Specifically, both the number of samples needed to uniformly cover the flux space and the number of pathways increase combinatorially with network size (Klamt and Stelling, 2002).

Uniform random flux sampling has become feasible for GEMs in recent years (Haraldsdóttir *et al.*, 2017), but to scale pathway analysis it is necessary to focus on a subset of pathways or a subnetwork of the full metabolic network (Zanghellini *et al.*, 2016). For example, elementary flux patterns (EFPs) have been defined as unique sign patterns (+/−) of EFMs from a full network that pass through a subnetwork, specifically the patterns that cannot be generated from other patterns without cancellations (Kaleta *et al.*, 2009). Many biological questions are primarily concerned with organisms' interactions with each other or their environment (Correia-Melo *et al.*, 2023), and in these cases one can ignore internal reactions entirely to focus on the subnetwork of boundary reactions that allow metabolite exchange. This is the motivation for elementary conversion modes (ECMs): net metabolite conversions that, analogously to EFVs, are the minimal set of conformal generators of a conversion cone (Urbanczik and Wagner, 2005; Clement *et al.*, 2021). Like EFPs, ECMs can be generalized to arbitrary subnetworks (Clement *et al.*, 2021), but both EFPs and ECMs are currently limited to homogeneous constraints. Minimal pathways (MPs), defined as minimal sets of reactions in a subnetwork that must have non-zero flux to satisfy all constraints on the full network, allow arbitrary subnetworks and constraints (Øyås and Stelling, 2020). ECMs, EFPs, and MPs can be compared as flux patterns, which, when applied to boundary reactions, represent net metabolite conversions (**Fig. 1**).

Although several unbiased methods suitable for analysis of metabolite exchanges are available, it remains unclear how they relate to each other in terms of prediction and biological interpretation of metabolite exchanges. To address this, we enumerated ECMs, EFPs, and MPs in metabolic models ranging in scope from cells to communities and in size from core models to GEMs. Focusing on flux patterns that support growth, we found that the MPs were always a subset of the EFPs, which in turn were always a subset of the ECMs. We also enumerated EFMs for all core models and found that the ECMs were always a subset of the subnetwork flux patterns from EFMs. Moreover, metabolite exchange frequencies were mostly stable across pathway definitions, showing that the same biological conclusions can be drawn from different methods. Flux sampling scaled to all analyzed models and complemented pathways with flux probabilities, particularly by sampling individually for each enumerated pathway. Overall, our results allow unbiased methods to be understood in conjunction with each other, which should enable users to choose the most appropriate approach for their research questions.



	ECM1	ECM2	ECM3	EFP1	EFP2	EFP3	MP1
A	-1	-2	-1	-	-	-	-
B	-1		+1	-		+	
BM	+1	+1		+	+		+

Fig. 1. Example metabolic network with comparison of ECMs, EFPs, and MPs. (**A**) Network with five internal reactions (r_1–r_5), three boundary reactions (b_1–b_3), and five metabolites (A–D and BM). Two A are needed to produce one C in r_1. Reaction reversibilities are indicated by arrows. The external metabolites A, B, and BM are imported or exported by their respective boundary reactions. BM represents biomass and the flux of its boundary reaction is the growth rate. (**B**) The ECMs, EFPs, and MPs of the network. ECMs include stoichiometry, while EFPs and MPs are given as flux patterns (− and + representing import and export, respectively). (**C**) The conversion cone of the network with ECMs, EFPs, and MPs represented as rays. The ECMs generate the cone without cancellations and the EFPs and MPs correspond to the flux patterns of the ECMs. The cone is unbounded in the direction of increasing import of A and bounded by the extreme rays, which are also the ECMs and EFPs, ECM1=EFP1 and ECM3=EFP3 (red). The last ray, ECM2=EFP2=MP1 (blue), is the only one that corresponds to an MP because it corresponds to the only minimal set of metabolite exchanges required to produce biomass. (**D**) Venn diagram comparing flux pattern representations of ECMs, EFPs, and MPs. The ECMs and EFPs overlap completely in this example, but this is not true in general.

2 Methods

2.1 Metabolic models

We obtained and analyzed an *Escherichia coli* core model (e_coli_core; Orth *et al.*, 2010), a *Helicobacter pylori* GEM (iIT341; Thiele *et al.*, 2005); and core models of a filamentous anoxygenic phototroph (fap), a sulfate-reducing bacterium (srb), and *Synechococcus* spp. (syn), both individually and interacting with each other in a microbial community model (Taffs *et al.*, 2009). The models e_coli_core and iIT341 were downloaded from the *ecmtool* repository (https://github.com/SystemsBioinformatics/ecmtool). We set the bounds of iIT341 to reflect the minII medium (Thiele *et al.*, 2005) and kept the minimal glucose medium for e_coli_core. We used COBRApy (Ebrahim *et al.*, 2013) to build the fap, srb, and syn models by extracting reaction and compartment information from the supplementary information of Taffs *et al.* (2009). We connected the models to each other, allowing the intermicrobial and environmental exchanges described in Fig. 1A of Taffs *et al.* (2009), to model a phototrophic microbial mat community in a daylight scenario in which fap and syn perform photosynthesis. Biomass produced by individual microbes were added as substrates to a community biomass reaction, thus requiring balanced growth of all three microbes.

2.2 EFM enumeration

We used the Python interface of *efmtool* (Terzer and Stelling, 2008) to enumerate EFMs for e_coli_core, fap, srb, syn, and the microbial community model. Irreversible reactions defined with negative flux, i.e., from right to left, were reversed before enumeration. We extracted unique exchange patterns from EFMs by removing internal reactions, taking the sign of the reduced EFM matrix, and removing duplicate patterns. For the community model, we included both environmental and intermicrobial exchanges. EFM enumerations were performed on a Lenovo ThinkPad laptop with an Intel Core i7-8665U (1.90GHz) processor and 32 GB RAM.

2.3 ECM enumeration

We used the Python package *ecmtool* to enumerate ECMs (Clement *et al.*, 2021). To reproduce the results of Clement *et al.* (2021), we ran ECM enumerations for e_coli_core and iIT341 using the command-line interface of *ecmtool* and scripts based on their supplementary information. These ECM enumerations were performed on the Orion cluster at the Norwegian University of Life Sciences (NMBU), using one core for e_coli_core and four cores for iIT341. We enumerated ECMs for fap, srb, syn, and the microbial community model using the Python interface of *ecmtool* with default settings. For the community model, we included both environmental and intermicrobial exchanges by using the "tag" method of *ecmtool*. We converted ECMs to unique exchange patterns by taking the sign of the ECM matrix and removing duplicate patterns. These ECM enumerations were performed on a Lenovo ThinkPad laptop with an Intel Core i7-8665U (1.90GHz) processor and 32 GB RAM.

2.4 EFP enumeration

We enumerated EFPs from the unique exchange patterns of EFMs or ECMs. Specifically, we identified EFPs as the patterns that could not be constructed from other patterns without cancellations. For each pattern, we took the union of all other patterns of which it was a superset and checked whether this union was equal to the pattern itself.

2.5 MP enumeration

We used the Python package *mptool* (Øyås and Stelling, 2020) to enumerate MPs. Specifically, we used the iterative graph method and default settings. We added a minimal growth rate requirement by setting the lower bound of the biomass reaction to a small positive value (10^{-4} h^{-1}). All reversible boundary reactions were split into two irreversible reactions to distinguish between production and consumption of metabolites. The subset of irreversible boundary reactions was then chosen as the subnetwork for MP enumeration. For the microbial community model, we also included intermicrobial metabolite exchanges in the subnetwork. All MP enumerations were performed on a Lenovo ThinkPad laptop with an Intel Core i7-8665U (1.90GHz) processor and 32 GB RAM.

2.6 Comparing pathways

EFMs and ECMs are enumerated as vectors, in which each element is a stoichiometric coefficient or flux, while EFPs and MPs are enumerated as unique patterns that only include the signs of these elements. MPs must also satisfy a functional requirement, in our case a minimal growth rate. To make all pathways comparable, we therefore extracted unique exchange patterns from EFMs and ECMs and further extracted growth-supporting patterns for EFMs, ECMs, and EFPs. We converted the patterns to sets, distinguishing between import and export, and intersected all pairs of pathways to count subsets and supersets (**Fig. 1**).

2.7 Metabolite exchange frequencies

We computed metabolite exchange frequencies separately for EFMs, ECMs, EFPs, and MPs by counting the number of pathways in which each metabolite was exchanged and dividing by the total number of pathways. Pairwise metabolite exchange frequencies were computed in the same way, i.e., by counting the number of pathways in which each pair of metabolites was exchanged together and dividing by the total number of pathways. For the microbial community model, we also computed the frequency of metabolic interactions, i.e., intermicrobial metabolite exchanges. We counted the number of pathways in which each metabolite was produced by each microbe and consumed by each other microbe and divided by the total number of pathways, separately for EFMs, ECMs, EFPs, and MPs.

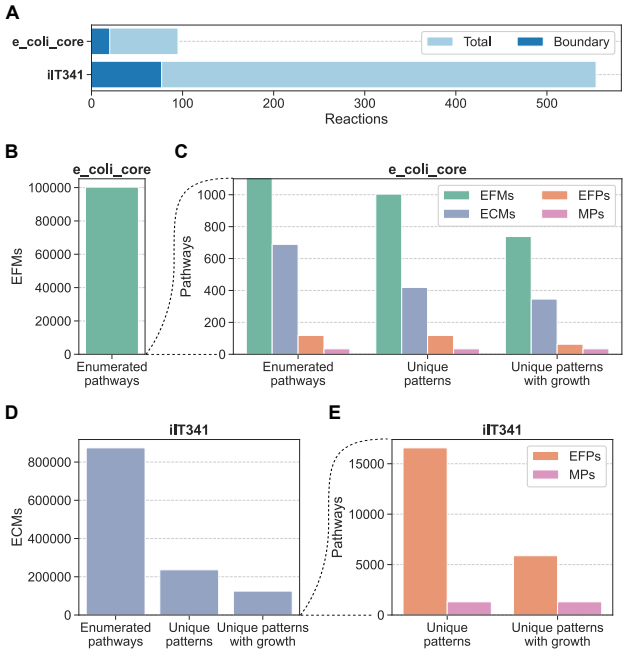


Fig. 2. Number of reactions and enumerated pathways for microbial species. (**A**) Number of reactions and boundary reactions in models. (**B**) Number of enumerated EFMs in e_coli_core (total). (**C**) Number of enumerated EFMs, ECMs, EFPs, and MPs in e_coli_core (total, unique patterns, and unique patterns with growth). (**D**) Number of enumerated ECMs in iIT341 (total, unique patterns, and unique patterns with growth). (**E**) Number of enumerated EFPs and MPs in iIT341 (unique patterns and unique patterns with growth).

2.8 Random flux sampling

Uniform random flux sampling of e_coli_core and iIT341 was performed using the PTA flux sampler (without thermodynamics) from the Python package *pta* (Gollub *et al.*, 2021). For each model, we set the lower bound of the biomass reaction to 0.1 h^{-1} and sampled 100,000 flux distributions from the resulting solution space. We ensured the space was sufficiently sampled by checking the convergence of the sampler with the "check_convergence" method of *pta*. These samplings were performed on a Lenovo ThinkPad laptop with an Intel Core i7-1165G7 (2.80GHz) prosessor and 16 GB RAM. We also applied the OptGP sampler (Megchelenbrink *et al.*, 2014) from COBRApy (Ebrahim *et al.*, 2013), to the same models with very similar results. Subsequently, we obtained 1,000 samples for each enumerated MP from e_coli_core and 100,000 samples each for fap, srb, syn, and the microbial community model using OptGP. These samplings were performed on a Lenovo ThinkPad laptop with an Intel Core i7-8665U (1.90GHz) processor and 32 GB RAM. We used flux variability analysis (FVA) to ensure that samples were within the feasible flux ranges (Mahadevan and Schilling, 2003).

3 Results

3.1 Metabolite exchanges in microbial species

We first applied pathway enumeration to an *E. coli* core model (e_coli_core) with 95 (20 boundary) reactions and an *H. pylori* GEM (iIT341) with 554 (77 boundary) reactions (**Fig. 2**). For e_coli_core, we enumerated 100,274 EFMs for the full network, from which we obtained 1,004 unique flux patterns in the subnetwork of boundary reactions. From these patterns, 738 of which were growth-supporting, we extracted 118 EFPs, 63 of which were growth-supporting. We also enumerated 689 ECMs, corresponding to 346 growth-supporting patterns, and 34 MPs, all of which were required to be growth-supporting patterns. Interestingly,

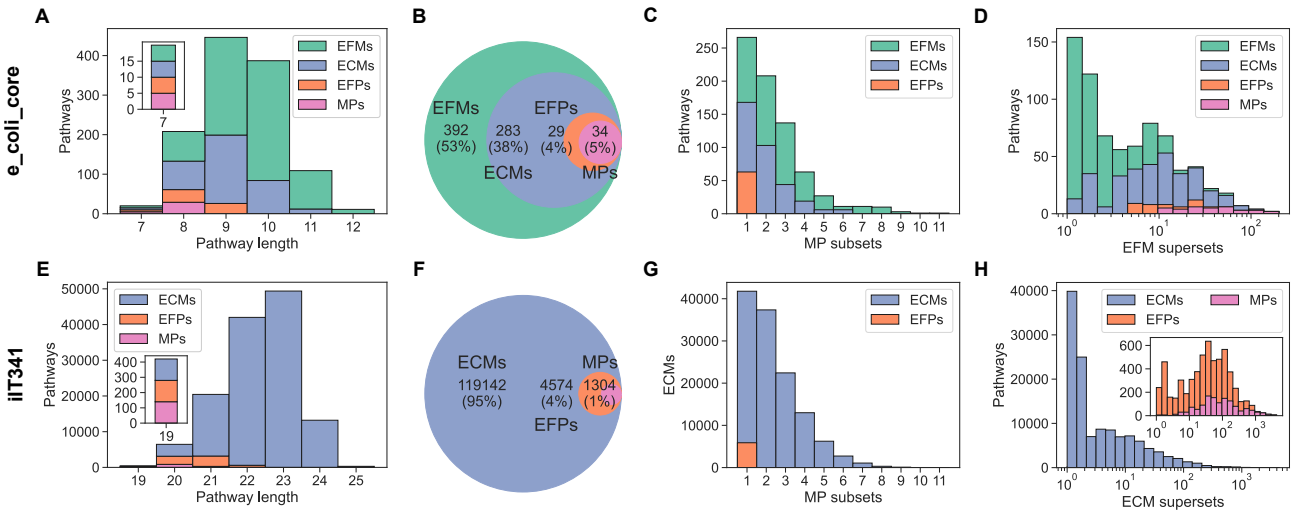


Fig. 3. Comparison of pathways for microbial species. (**A**) Distribution of EFM, ECM, EFP, and MP lengths for e_coli_core. The inset shows the shortest pathways. (**B**) Venn diagram of EFMs, ECMs, EFPs, and MPs for e_coli_core. Number of growth-supporting patterns and percent of total are shown. (**C**) Number of MPs that are subsets of EFMs, ECMs, and EFPs for e_coli_core. (**D**) Number of EFMs that are supersets of EFMs, ECMs, EFPs, and MPs for e_coli_core. (**E**) Distribution of ECM, EFP, and MP lengths for iIT341. The inset shows the shortest pathways. (**F**) Venn diagram of ECMs, EFPs, and MPs for iIT341. Number of growth-supporting patterns and percent of total are shown. (**G**) Number of MPs that are subsets of ECMs and EFPs for iIT341. (**H**) Number of ECMs that are supersets of ECMs, EFPs, and MPs for iIT341. Bars are stacked (on top of each other) in (**A**) and (**E**) and layered (behind each other) in (**B**)–(**D**) and (**F**)–(**H**).

we found that it was possible to extract the same 118 EFPs from the ECMs as from the EFMs. EFM enumeration was infeasible for iIT341 with the computational power available to us, but we enumerated 874,236 ECMs, 125,020 of which were unique and growth-supporting, as well as 1,304 MPs. Based on our observations from e_coli_core, we used the ECMs to extract 16,573 EFPs, 5,878 of which were growth-supporting. Thus, we consistently found more ECMs than EFPs and more EFPs than MPs. The ECMs, EFPs, and MPs all included the same metabolites (16 for e_coli_core and 45 for iIT341) in uptake and secretion combinations that were generally as expected for growing bacteria (**Fig. S1**).

Comparing the lengths of unique growth-supporting patterns for EFMs, ECMs, EFPs, and MPs, we found that the number of shortest pathways was always the same for all pathway definitions, which led us to compare all pairs of patterns by intersecting them with each other (**Fig. 3**). Indeed, the shortest EFMs, ECMs, EFPs, and MPs were identical to each other. More surprisingly, we found that the ECMs were a superset of the EFPs, which in turn were a superset of the MPs. Specifically, the MPs were equal to the EFPs and ECMs that were support-minimal, i.e., that only included metabolite exchanges required to support growth. Counting the number of MPs that were a subset of each of the other pathways, we found a decreasing trend from one to 11 MP subsets per EFM and ECM. For ECMs, this trend was remarkably consistent between e_coli_core and iIT341 despite the latter model being nearly six times larger than the former with four times as many boundary reactions. Also, all e_coli_core and iIT341 EFPs had exactly one MP subset each. In contrast to this, the number of EFMs and ECMs that were supersets of other pathways ranged from one to about 200 EFMs for e_coli_core and from one to more than 4,300 ECMs for iIT341. Again, we found a trend: for both models, the average number of EFM and ECM supersets decreased from EFMs to ECMs to EFPs to MPs. All subset counts as well as EFM and ECM superset counts correlated positively with each other and with length, while EFP, ECM and EFP superset counts correlated negatively with length (**Figs. S2** and **S3**).

For each exchanged metabolite, we computed the frequency of uptake and secretion in EFMs, ECMs, EFPs, and MPs as well as differences in frequency between pathway definitions (**Fig. 4**). In both e_coli_core and iIT341, a core set of metabolite exchanges were included

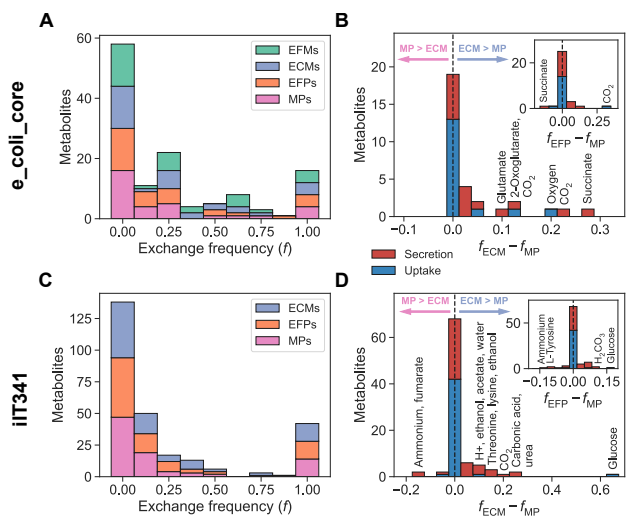


Fig. 4. Metabolite exchange frequencies for microbial species. (**A**) Metabolite exchange frequency (f) in EFMs, ECMs, EFPs, and MPs for e_coli_core. (**B**) Differences in metabolite exchange frequencies between ECMs and MPs (f_{ECM} - f_{MP}) and between EFPs and MPs (f_{EFP} - f_{ECM}) for e_coli_core. (**C**) Metabolite exchange frequency (f) in ECMs, EFPs, and MPs for iIT341. (**D**) Differences in metabolite exchange frequencies between ECMs and MPs (f_{ECM} - f_{MP}) and between EFPs and MPs (f_{EFP} - f_{ECM}) for iIT341.

in all pathways, and thus essential, while most metabolites were only exchanged in one or a few pathways. In general, both individual and pairwise exchange frequencies were remarkably consistent across pathway definitions (**Figs. S4** and **S5**). However, the frequencies of some metabolite exchanges differed by more than 10 percentage points between pathway definitions. For e_coli_core, uptake of oxygen and CO_2 and secretion of succinate, CO_2, 2-oxoglutarate, and glutamate were overrepresented in ECMs relative to MPs. Uptake of CO_2 was also overrepresented in EFPs relative to MPs, while secretion of succinate was underrepresented. Comparing ECMs and MPs for iIT341, uptake of glucose and H^+ and secretion of urea, carbonic acid (H_2CO_3), CO_2, threonine, lysine, ethanol,

acetate, and water were overrepresented in ECMs, while secretion of ammonium and fumarate were underrepresented. Glucose uptake and carbonic acid secretion were overrepresented in EFPs relative to MPs, while ammonium and fumarate secretion were underrepresented.

Applying random flux sampling to boundary reactions in e_coli_core and iIT341, we obtained probability distributions over metabolite exchange fluxes that were consistent between two different samplers (**Figs. S6–S9**). These probability distributions gave an overview of the growth-supporting flux space, but they could not be directly decomposed to the elementary metabolite conversions provided by pathways. Notably, for e_coli_core, none of the 100,000 samples included CO_2 uptake, which was feasible in the model and part of several pathways. By sampling fluxes separately for each of the 34 e_coli_core MPs, we found that CO_2 uptake fluxes were indeed accessible for samplers and that sampling could complement pathway analysis by allowing detailed analysis of individual pathways' flux spaces (**Fig. S10**).

3.2 Metabolite exchanges in a microbial community

To investigate intermicrobial metabolite exchange, we built a microbial community model with 132 (15 boundary) reactions by connecting core models of three individual microbes to each other: *Synechococcus* spp. (syn) with 37 (8 boundary) reactions, a filamentous anoxygenic phototroph (fap) with 47 (10 boundary) reactions, and a sulfate-reducing bacterium (srb) with 44 (11 boundary) reactions (Taffs *et al.*, 2009). We enumerated EFMs, ECMs, EFPs, and MPs for the individual models and the community model (**Fig. 5**). Pathway counts consistently decreased from fap to srb to syn, and from EFMs to ECMs to EFPs to MPs. Counting unique growth-supporting flux patterns, we found 41 EFMs, 34 ECMs, 10 EFPs, and seven MPs for fap; 209 EFMs, five EFMs, four ECMs, three EFPs, and three MPs for srb; and a single EFM, ECM, EFP, and MP for syn. We saw the same decrease in pathway counts for the community model, from 559 EFMs to 248 ECMs to 60 EFPs to 34 MPs. These community pathways included both intermicrobial and environmental metabolite exchanges. EFMs, ECMs, EFPs, and MPs included the same metabolites in expected uptake and secretion combinations (Taffs *et al.*, 2009) with the exception of the fap EFMs and ECMs, which included two metabolite exchanges that were not part of any EFPs or MPs (**Fig. S11**).

Comparing ECMs, EFPs, and MPs, we found that pathways from the individual models were shorter than the community model pathways, and that EFMs, ECMs, EFPs, and MPs were sub- and supersets of each other, as we saw for microbial species (**Fig. 6**). Also in line with results from microbial species, we found MP subset counts decreasing from one to 12, EFM superset counts ranging from one to 135, and EFM superset counts decreasing, on average, from EFMs to ECMs to EFPs to MPs. We found positive correlation between EFM, ECM, EFP, and MP subset counts and between EFM and ECM superset counts, and length correlated positively with subset counts but negatively with superset counts (**Figs. S12** and **S13**).

In both the individual models and the community model, we found a core set of essential metabolite exchanges that were included in all pathways as well as similar metabolite exchange frequencies for EFMs, ECMs, EFPs, and MPs (**Fig. 7** and **Figs. S14** and **S15**). For the individual models, uptake of acetate, light, and CO_2 by fap and secretion of CO_2 by srb were more frequent in ECMs than in MPs. Uptake of light, glycolate, and acetate by fap were overrepresented in EFPs relative to MPs, while uptake of oxygen by fap was underrepresented. For the community model, uptake of acetate by fap and srb, secretion of acetate by syn, uptake of light by fap, and uptake of H_2 from the environment were overrepresented in ECMs compared to MPs. Uptake of H_2 by fap and from the environment were overrepresented in EFPs relative to MPs, while CO_2 secretion by fap was underrepresented. As observed for environmental metabolite exchanges in microbial species, the frequencies of intermicrobial metabolic interactions were consistent between pathway definitions. EFMs, ECMs, EFPs, and MPs all agreed on all of the following metabolite exchanges: acetate and ammonia from syn to fap and srb, CO_2 from fap to srb and syn and from srb to fap and syn, glycolate from syn to fap, H_2 from fap to srb and from srb to fap, and oxygen from syn to fap.

Random flux sampling of the individual and community models produced probability distributions over environmental and intermicrobial metabolite exchange fluxes (**Figs. S16** and **S17**). This showed that some metabolite exchanges were constrained to narrow flux ranges, especially in the community model, while others were free to vary across several orders of magnitude. The flux probability distributions tended to be wider in the individual models than in the community model. For example, ammonia production and secretion fluxes were uniformly distributed across the feasible flux range in the individual syn model, but constrained to a narrow range of small fluxes in the community model.

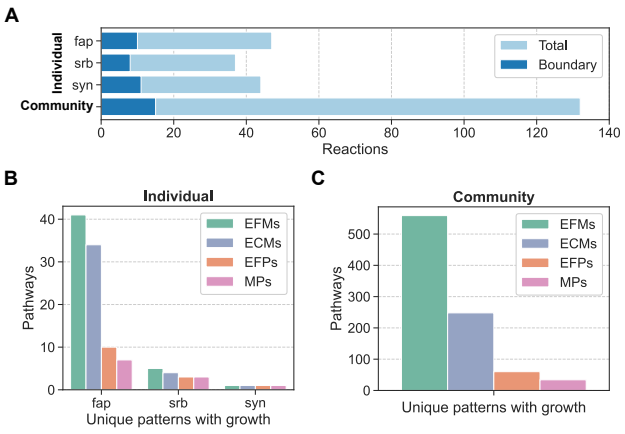


Fig. 5. Number of reactions and enumerated pathways for a microbial community. (**A**) Number of reactions and boundary reactions in individual and community models. (**B**) Number of enumerated EFMs, ECMs, EFPs, and MPs in individual models (unique patterns with growth). (**C**) Number of enumerated EFMs, ECMs, EFPs, and MPs in community model (unique patterns with growth).

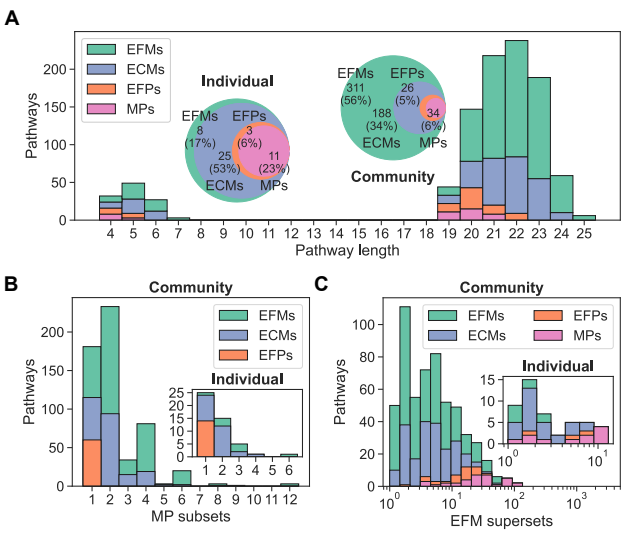


Fig. 6. Comparison of pathways for a microbial community. (**A**) Distribution of EFM, ECM, EFP, and MP lengths for individual and community models. Bars are stacked on top of each other. Venn diagrams of EFMs, ECMs, EFPs, and MPs are shown with number of unique patterns with growth and percent of total. (**B**) Number of MPs that are subsets of EFMs, ECMs, and EFPs for individual and community models. Bars are layered behind each other. (**C**) Number of FPs that are supersets of FPs, ECMs, and MPs for individual and community models. Bars are layered behind each other.

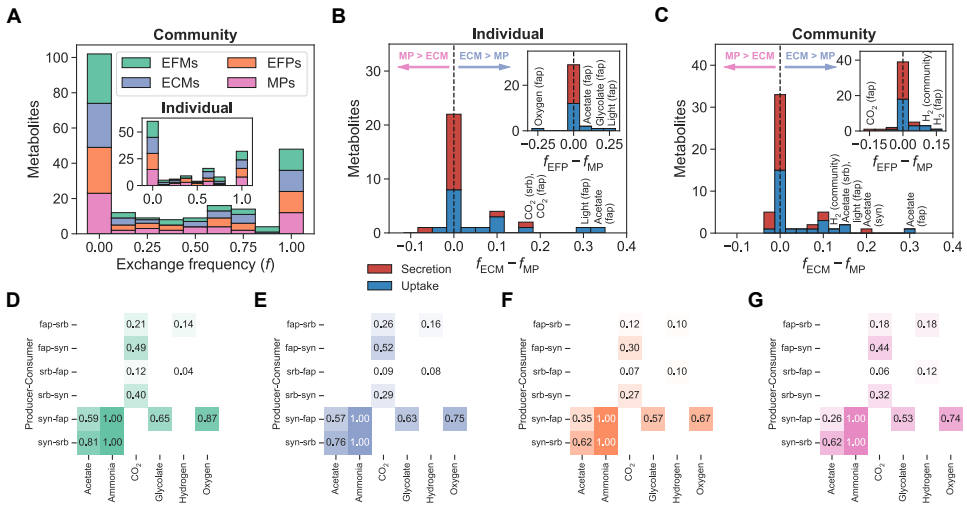


Fig. 7. Metabolite exchange frequencies for a microbial community. (**A**) Metabolite exchange frequency (f) in ECMs, EFPs, and MPs for individual and community models. (**B**) Differences in metabolite exchange frequencies between ECMs and MPs (f_{ECM} - f_{MP}) and between EFPs and MPs (f_{EFP} - f_{MP}) for individual models. (**C**) Differences in metabolite exchange frequencies between ECMs and MPs (f_{ECM} - f_{MP}) and between EFPs and MPs (f_{EFP} - f_{MP}) for community model. (**D**) Metabolic interactions from EFMs. (**E**) Metabolic interactions from ECMs. (**F**) Metabolic interactions from EFPs. (**G**) Metabolic interactions from MPs. Rows are producer-consumer pairs, columns are metabolites, and each cell shows an exchange frequency.

4 Discussion

We demonstrate how different pathway definitions relate to each other, both in models of individual microbial species and in a community model. Distributions of pathway counts, pathway lengths, and degree of overlap between pathways followed a consistent trend across single- and multi-species models – EFMs $\geq$ ECMs $\geq$ EFPs $\geq$ MPs – reflecting a hierarchy in terms of metabolite exchange patterns. EFMs generally yield a larger set of exchange patterns than ECMs because EFMs are guaranteed to be the minimal set of conformal generators for the flux cone of the full network, not for the conversion cone of the subnetwork of metabolite exchanges. However, the ECMs correspond precisely to the minimal subset of EFMs that conformally generate the cone of exchange fluxes, which is equivalent to the conversion cone. The EFPs are also a subset of the EFMs because EFPs are defined as the subnetwork flux patterns from EFMs that cannot be built from others without cancellations. We show that EFPs are a subset of the ECMs as well, and that the same EFPs can be identified from the ECMs as from the EFMs. EFPs do not direcly generate the conversion cone but they are the elementary building blocks of the ECMs and parts of EFMs that do. Finally, the MPs are a subset of the EFMs, ECMs, and EFPs, thus forming a core set of flux patterns that are support-minimal and shared by all pathway definitions. Like the EFPs, the MPs do not generate the conversion cone, and MPs cannot be combined to build all EFMs, ECMs, or EFPs either. Instead, the MPs correspond to the support-minimal subspace of the flux and conversion cones. If the full flux space is support-minimal, EFMs, ECMs, EFPs, and MPs have equal flux patterns. An example of this is the syn model, for which we found a single flux pattern that is simultaneously an EFM, ECM, EFP, and MP.

Pathways found by different methods generally included the same metabolites in similar combinations of uptake and secretion despite large differences in the total number of pathways enumerated. This is illustrated by the distributions of subsets and supersets, which display an increasing degree of overlap with other pathways the more pathways are enumerated. On average, the number of MP subsets increases while the number of EFM or ECM supersets decreases with the number of pathways, from EFMs to ECMs to EFPs to MPs. We interpret a pathway's subset counts as the number of other pathways that are needed to summarize it, and its superset counts as the number of other pathways it summarizes. Superset counts were orders of magnitude larger and grew faster with model size than subset counts, which were very well-preserved between models differing in size and structure. The number of EFMs and ECMs grow rapidly with model size, but the many additional pathways tend to summarize few EFPs and MPs. Conversely, EFPs and MPs tend to summarize many of the additional EFMs and ECMs. EFPs always had exactly one MP subset and vice versa, indicating that EFPs and MPs are closely related to each other. Specifically, the EFPs consist of the support-minimal pathways that are shared by all pathway definitions, i.e., the MPs, plus a minimal number of additional pathways that are needed to construct all EFMs and ECMs.

Further similarities between pathway definitions can be noted in the individual and pairwise frequencies of metabolite exchanges, which were highly conserved within each model. This shows that different pathway enumeration methods provide similar predictions of network capabilities and thus allow consistent biological interpretation. In line with this, large differences in pathway counts between methods were explained by small sets of overrepresented metabolites. For example, most of the 123,717 ECMs from iIT341 that were not MPs were accounted for by 11 metabolite uptakes and secretions that were not strictly required for growth. Notably, glucose uptake was about 60 percentage points more frequent in ECMs than in MPs, reflecting the fact that *H. pylori* can synthesize pyruvate from other sources than glucose such as alanine (Doig *et al.*, 1999). As MPs are the only pathways required to be support-minimal, they omit metabolites that are not essential for each pathway, explaining why glucose was not frequently featured in MPs compared to ECMs and EFPs. Conversely, secretion of ammonium was more frequent in MPs than in ECMs and EFPs because pathway length is minimized by breaking down as many nitrogenous waste products as possible to ammonium instead of secreting compounds such as urea, threonine, or lysine.

Our results from microbial species also held for microbial communities, which have become a recent focus in constraint-based modeling (Heinken *et al.*, 2021). In these models, one is often especially interested in analyzing interactions between species and therefore focus on extracellular activities, making it a natural application for metabolite exchange enumeration. However, microbial community models also differ from models of microbial species in terms of both size and strucure: they consist of multiple single-species models that are combined into the same model and connected through their exchange reactions. This means that the single-species exchange reactions become internal exchange reactions in the microbial community model, which also includes its own external exchange reactions that allow metabolite exchange with the environment. Despite this, we found that distributions of pathway counts, pathway lengths, and degree of overlap between pathways for internal and external

exchanges in the community model were very similar to those we saw for external exchanges in the single-species models. We also found the same hierarchy of pathways from EFMs to ECMs to EFPs to MPs. Thus, we found that all pathway definitions included in this study could be generalized to a subnetwork that included both external and internal exchanges, indicating that EFMs, ECMs, EFPs, and MPs are all applicable to general subnetworks and thus that our results hold more broadly.

Random flux sampling provides flux probability distributions for metabolite exchanges that can complement enumerated pathways, which we also saw in this study. However, some enumerated pathways did not have a corresponding flux distribution obtained by flux sampling. One example of this was uptake of CO_2 by *E. coli*, which can indeed fixate CO_2 using only native enzymes (Satanowski *et al.*, 2020). One possible explanation is that some subspaces corresponding to specific flux patterns are diminishingly small in comparison to the rest of the solution space, making them hard to access for samplers and requiring a very high number of samples to be detected. As a way to avoid this problem and an example of complementary pathway enumeration and flux sampling, we show that flux sampling can be applied separately to each pathway to obtain individual flux probability distributions.

Due to increasing pathway counts, enumeration should become progressively harder from MPs to EFPs to ECMs to EFMs. However, the computational efficiency and scalability of enumeration algorithms and their implementations differ greatly between pathway definitions. EFM and ECM enumeration is based on efficient implementations of the double description method (Terzer and Stelling, 2008) or lexicograhic reverse search (Buchner *et al.*, 2023), but still scales poorly because the number of flux vectors grows combinatorially with network size. EFP counts were generally much smaller – and appeared to grow more slowly with network size – than EFM and ECM counts but the current implementation of EFP enumeration is based on a mixed-integer linear program (MILP) that includes two binary variables for each reaction in the subnetwork (Kaleta *et al.*, 2009). This MILP scales poorly with subnetwork size and we therefore identified EFPs from enumerated EFMs and ECMs. MP enumeration also involves binary variables, but these are separated from the continuous flux variables into a binary integer program (BIP) that is alternated with multiple linear programs (LPs), which is is more efficient than a MILP (Song *et al.*, 2017). Combined with the minimal number of MPs, this helps make MP enumeration relatively scalable. Ultimately, the choice of pathway definition and enumeration method should be guided by biological questions and the size of the system of interest. For example, if one is interested in detailed analysis of all feasible metabolite exchange fluxes in a small network, one should enumerate EFMs or ECMs. If the patterns of these fluxes are sufficient, EFPs should allow scaling to larger subnetworks. Support-minimal pathways suffice for many key applications (Klamt *et al.*, 2017), and in these cases MPs would be the natural choice.

Pathway definitions are related to each other through a hierarchy of flux patterns that seems to generalize to arbitrary subnetworks and constraints. This helps explain the consistency of our metabolite exchange predictions across models and pathway definitions as well as the relative scalability of EFM, ECM, EFP, and MP enumeration. It also suggests that pathway definitions can be more formally unified into a common framework, which could facilitate efficient and scalable unbiased analysis. Specifically, elements of enumeration methods for different pathway definitions could be combined with each other as well as with flux sampling.

Acknowledgements

We thank Tom Clement, Daan de Groot, Bianca Buchner, and Jürgen Zanghellini for discussions and help with *ecmtool*, Bas Teusink and the rest of the Systems Biology lab at Vrije Universiteit Amsterdam for hosting us for research stays, Teshome Dagne Mulugeta for helping us set up software on the Orion cluster, Axel Theorell for help with random sampling, and Ross Carlson for help with the microbial community model.

Funding

This work was supported by the Research Council of Norway grants 248792 (DigiSal) and 248810 (Centre for Digital Life Norway).

References

Bordbar, A. *et al.* (2014). Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, **15**(2), 107–120.

Buchner, B. *et al.* (2023). ecmtool: fast and memory-efficient enumeration of elementary conversion modes. *Bioinformatics*, **39**(3), btad095.

Clement, T. J. *et al.* (2021). Unlocking Elementary Conversion Modes: ecmtool Unveils All Capabilities of Metabolic Networks. *Patterns*, **2**(1).

Correia-Melo, C. *et al.* (2023). Cell-cell metabolite exchange creates a pro-survival metabolic environment that extends lifespan. *Cell*, **186**(1), 63–79.

Doig, P. *et al.* (1999). Helicobacter pylori Physiology Predicted from Genomic Comparison of Two Strains. *Microbiology and Molecular Biology Reviews*, **63**(3), 675–707.

Ebrahim, A. *et al.* (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, **7**(1), 74.

Fallahi, S. *et al.* (2020). A comparison of Monte Carlo sampling methods for metabolic network models. *PLOS ONE*, **15**(7), e0235393.

Gollub, M. G. *et al.* (2021). Probabilistic thermodynamic analysis of metabolic networks. *Bioinformatics*, **37**(18), 2938–2945.

Haraldsdóttir, H. S. *et al.* (2017). CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, **33**(11), 1741–1743.

Heinken, A. *et al.* (2021). Advances in constraint-based modelling of microbial communities. *Current Opinion in Systems Biology*, **27**, 100346.

Kaleta, C. *et al.* (2009). Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Research*, **19**(10), 1872–1883.

Klamt, S. and Stelling, J. (2002). Combinatorial complexity of pathway analysis in metabolic networks. *Molecular Biology Reports*, **29**(1-2), 233–236.

Klamt, S. *et al.* (2017). From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLOS Computational Biology*, **13**(4).

Lewis, N. E. *et al.* (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, **10**(4), 291–305.

Mahadevan, R. and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, **5**(4), 264–276.

Megchelenbrink, W. *et al.* (2014). optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLOS ONE*, **9**(2), e86587.

Nielsen, J. and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell*, **164**(6), 1185–1197.

Orth, J. D. *et al.* (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus*, **4**(1).

Reimers, A. M. and Reimers, A. C. (2016). The steady-state assumption in oscillating and growing systems. *Journal of Theoretical Biology*, **406**, 176–186.

Satanowski, A. *et al.* (2020). Awakening a latent carbon fixation cycle in Escherichia coli. *Nature Communications*, **11**(1), 5812.

Schellenberger, J. and Palsson, B. Ø. (2009). Use of randomized sampling for analysis of metabolic networks. *Journal of Biological Chemistry*, **284**(9), 5457–5461.

Schilling, C. H. *et al.* (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*, **15**(3), 296–303.

Simensen, V. *et al.* (2022). Experimental determination of Escherichia coli biomass composition for constraint-based metabolic modeling. *PLOS ONE*, **17**(1), e0262450.

Song, H.-S. *et al.* (2017). Sequential computation of elementary modes and minimal cut sets in genome-scale metabolic networks using alternate integer linear programming. *Bioinformatics*, **33**(15), 2345–2353.

Taffs, R. *et al.* (2009). In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Systems Biology*, **3**(1), 114.

Terzer, M. and Stelling, J. (2008). Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**(19), 2229–2235.

Thiele, I. *et al.* (2005). Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *Journal of Bacteriology*, **187**(16), 5818–5830.

Urbanczik, R. and Wagner, C. (2005). Functional stoichiometric analysis of metabolic networks. *Bioinformatics*, **21**(22), 4176–4180.

Zanghellini, J. *et al.* (2016). *Toward Genome-Scale Metabolic Pathway Analysis*, chapter 3, pages 111–123. Wiley.

Øyås, O. and Stelling, J. (2020). Scalable metabolic pathway analysis. *bioRxiv*.

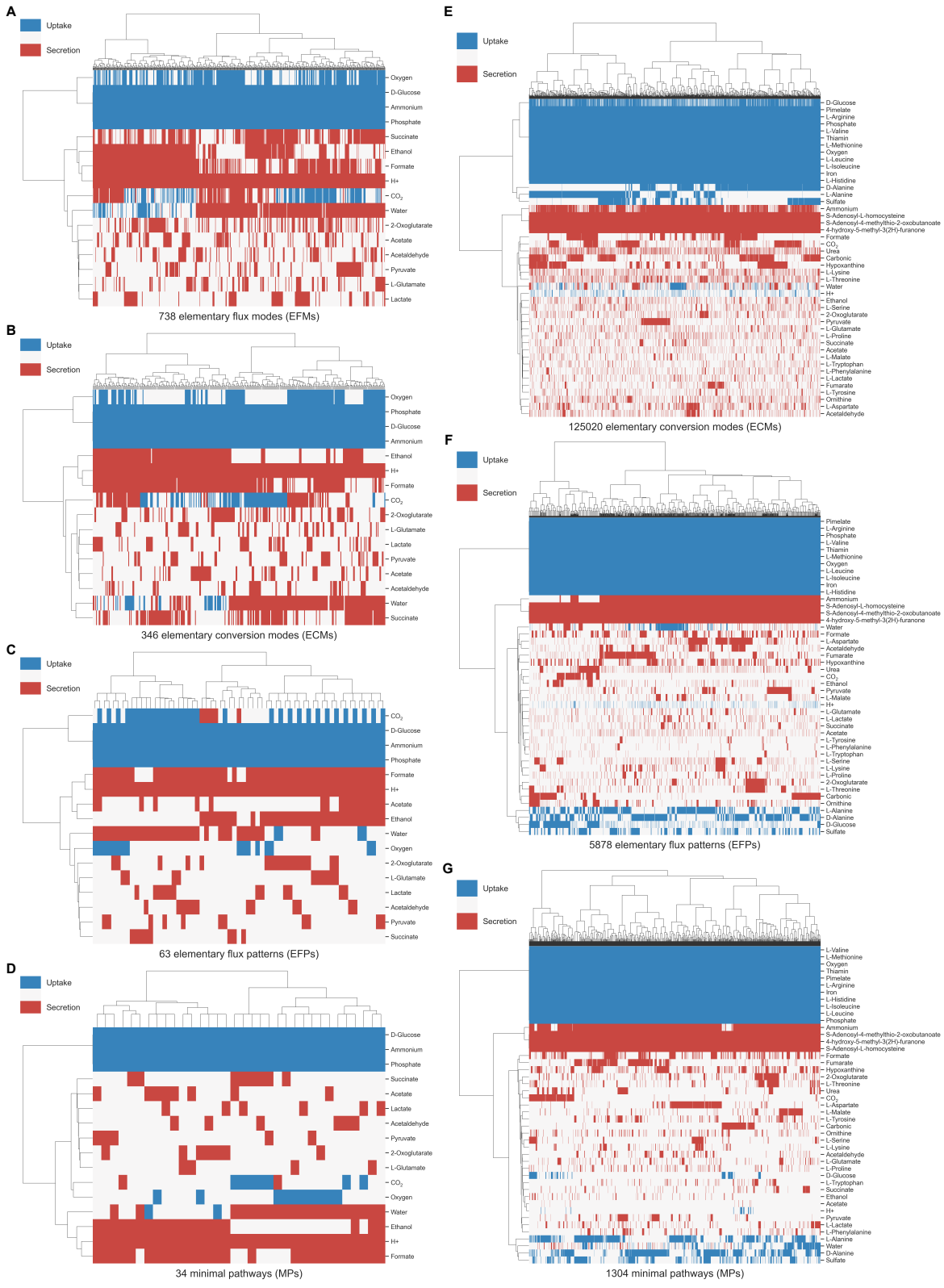


Fig. S1. Clustered heatmaps of (**A**) EFMs for e_coli_core, (**B**) ECMs for e_coli_core, (**C**) EFPs for e_coli_core, (**D**) MPs for e_coli_core, (**E**) ECMs for iIT341, (**F**) EFPs for iIT341, and (**G**) MPs for iIT341. Rows are metabolites, columns are pathways, and each cell indicates metabolite uptake (blue) or secretion (red) in a pathway. Rows and columns are clustered by Ward's minimum variance method. Unique growth-supporting flux patterns are shown.

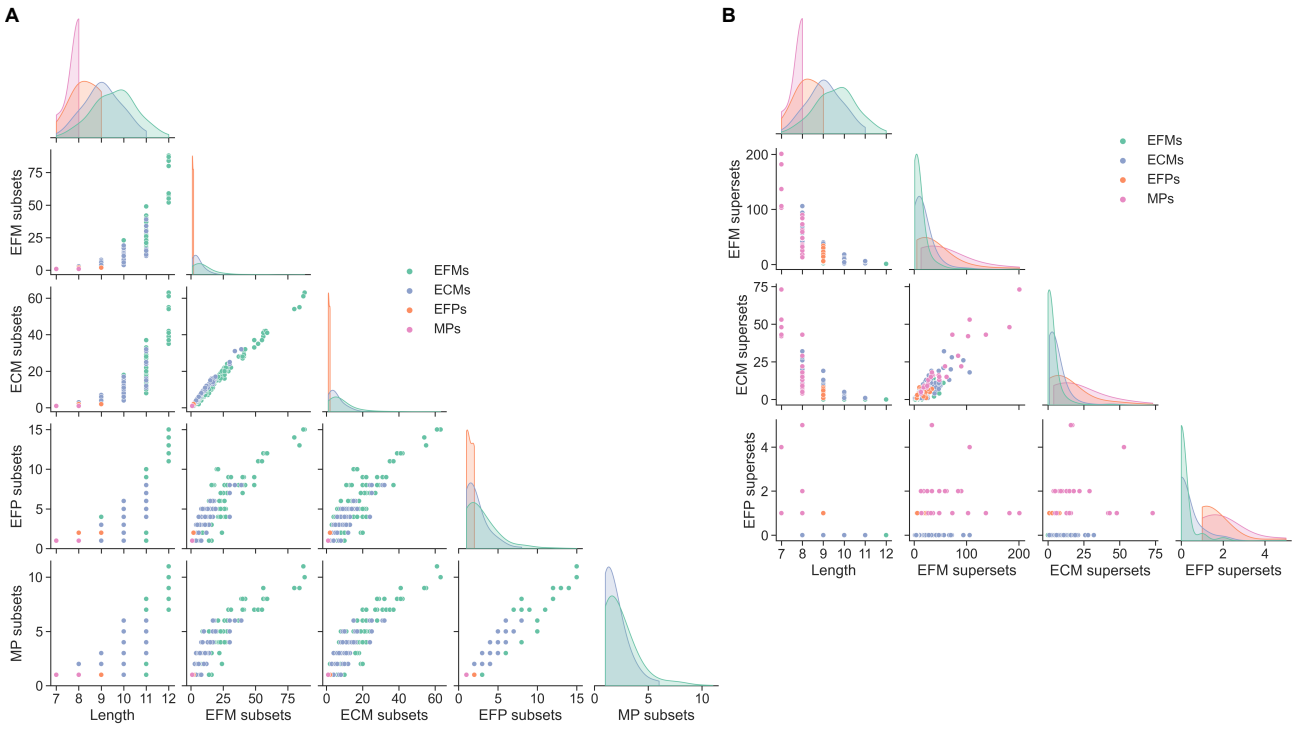


Fig. S2. Pairwise relationships between (**A**) pathway length and number of EFM, ECM, EFP, and MP subsets and (**B**) pathway length and number of EFM, ECM, and EFP supersets for e_coli_core.

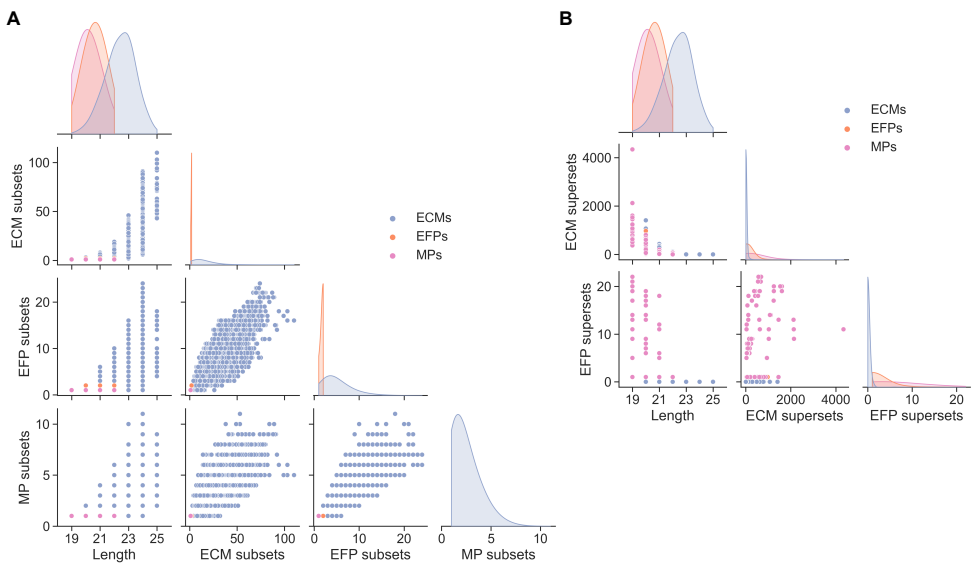


Fig. S3. Pairwise relationships between (**A**) pathway length and number of ECM, EFP, and MP subsets and (**B**) pathway length and number of ECM and EFP supersets for iIT341.

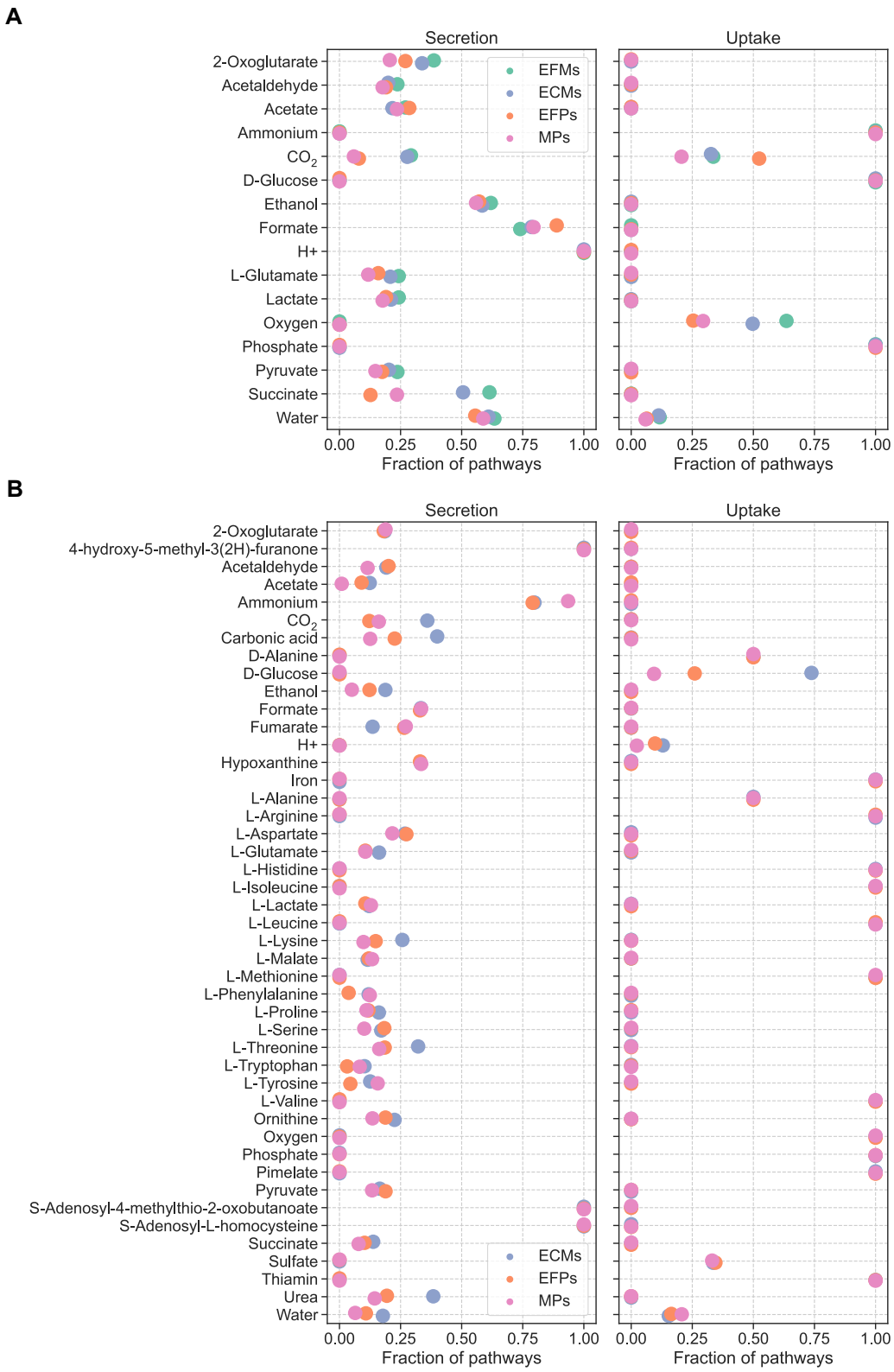


Fig. S4. Metabolite exchange frequencies (fraction of pathways including secretion or uptake of each metabolite) for (**A**) e_coli_core and (**B**) iIT341.

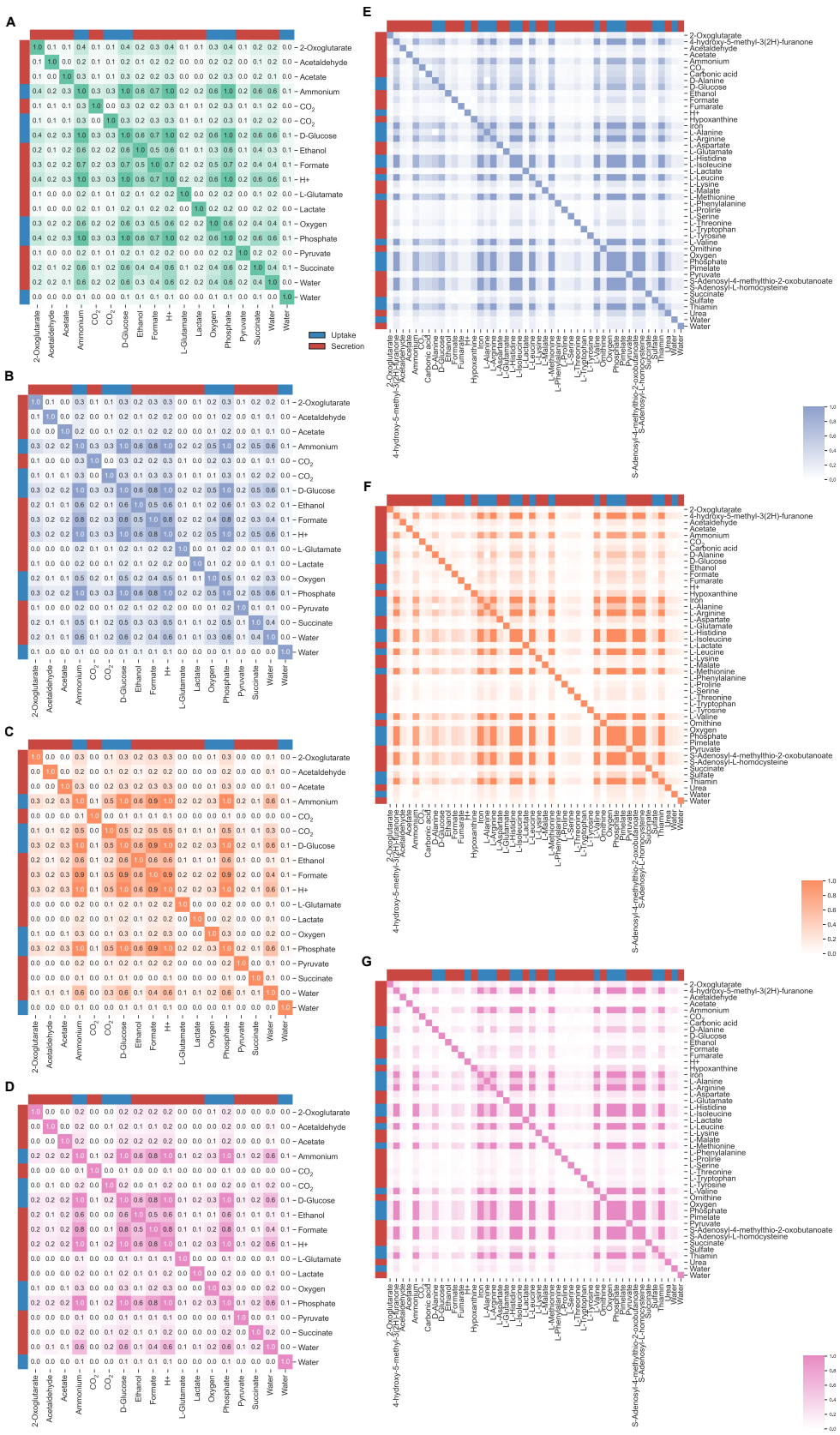


Fig. S5. Pairwise metabolite exchange frequencies (fraction of pathways including secretion or uptake of each metabolite pair) for (**A**) EFMs for e_coli_core, (**B**) ECMs for e_coli_core, (**C**) EFPs for e_coli_core, (**D**) MPs for e_coli_core, (**E**) ECMs for iIT341, (**F**) EFPs for iIT341, and (**G**) MPs for iIT341. Rows and columns are metabolites and each cell indicates the pairwise exchange frequency of two metabolites.

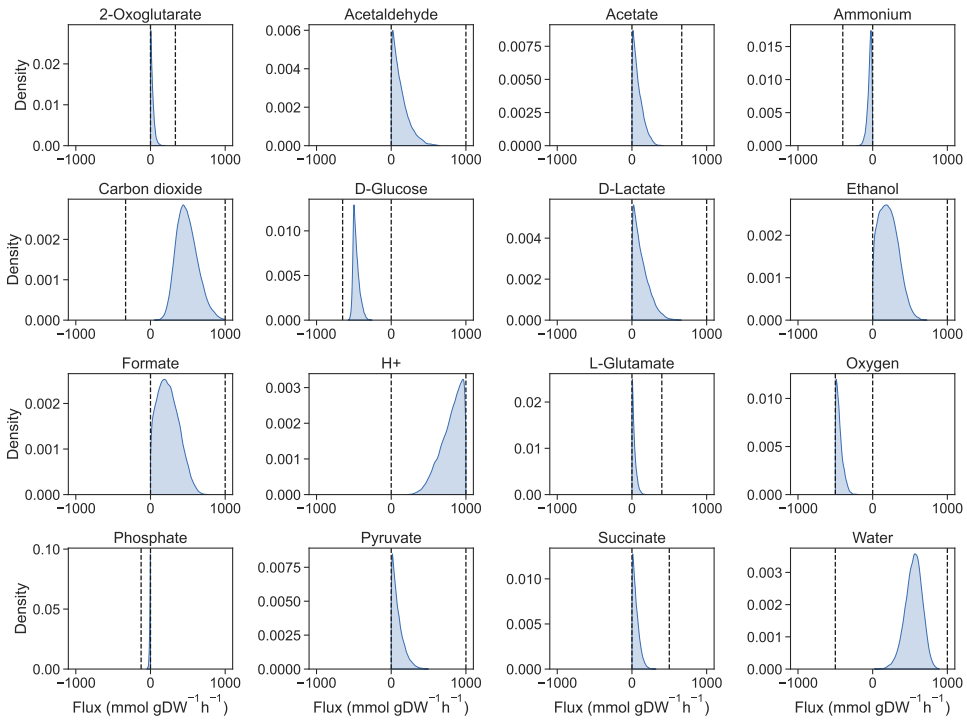


Fig. S6. Flux probability distributions for metabolite exchanges in e_coli_core from 100,000 random flux distributions sampled with PTA. Dashed lines indicate feasible flux ranges from FVA.

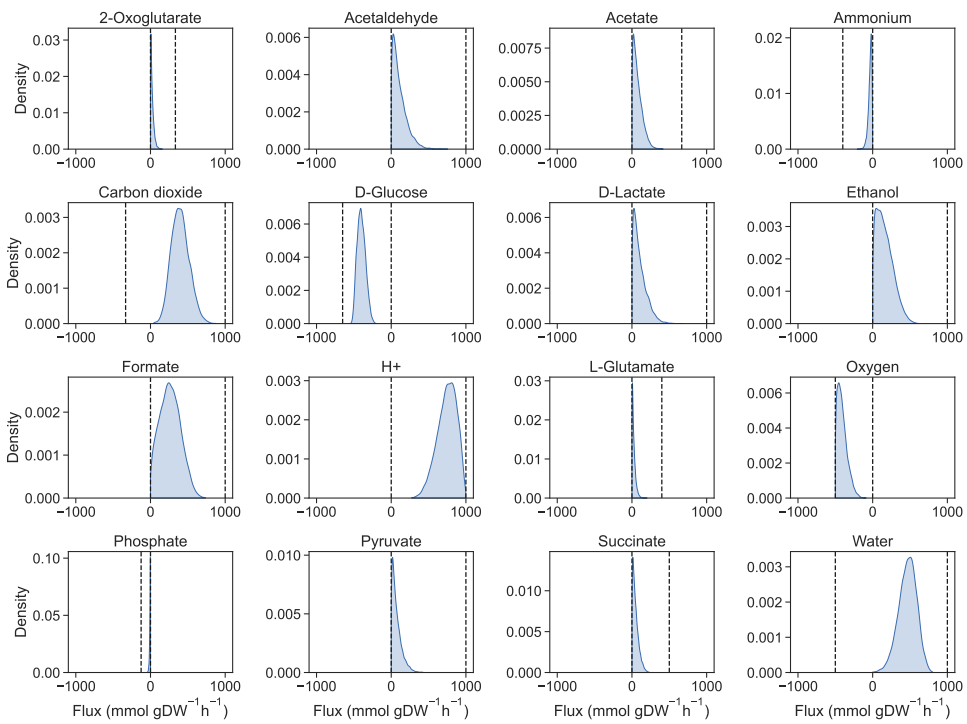


Fig. S7. Flux probability distributions for metabolite exchanges in e_coli_core from 100,000 random flux distributions sampled with OptGP. Dashed lines indicate feasible flux ranges from FVA.

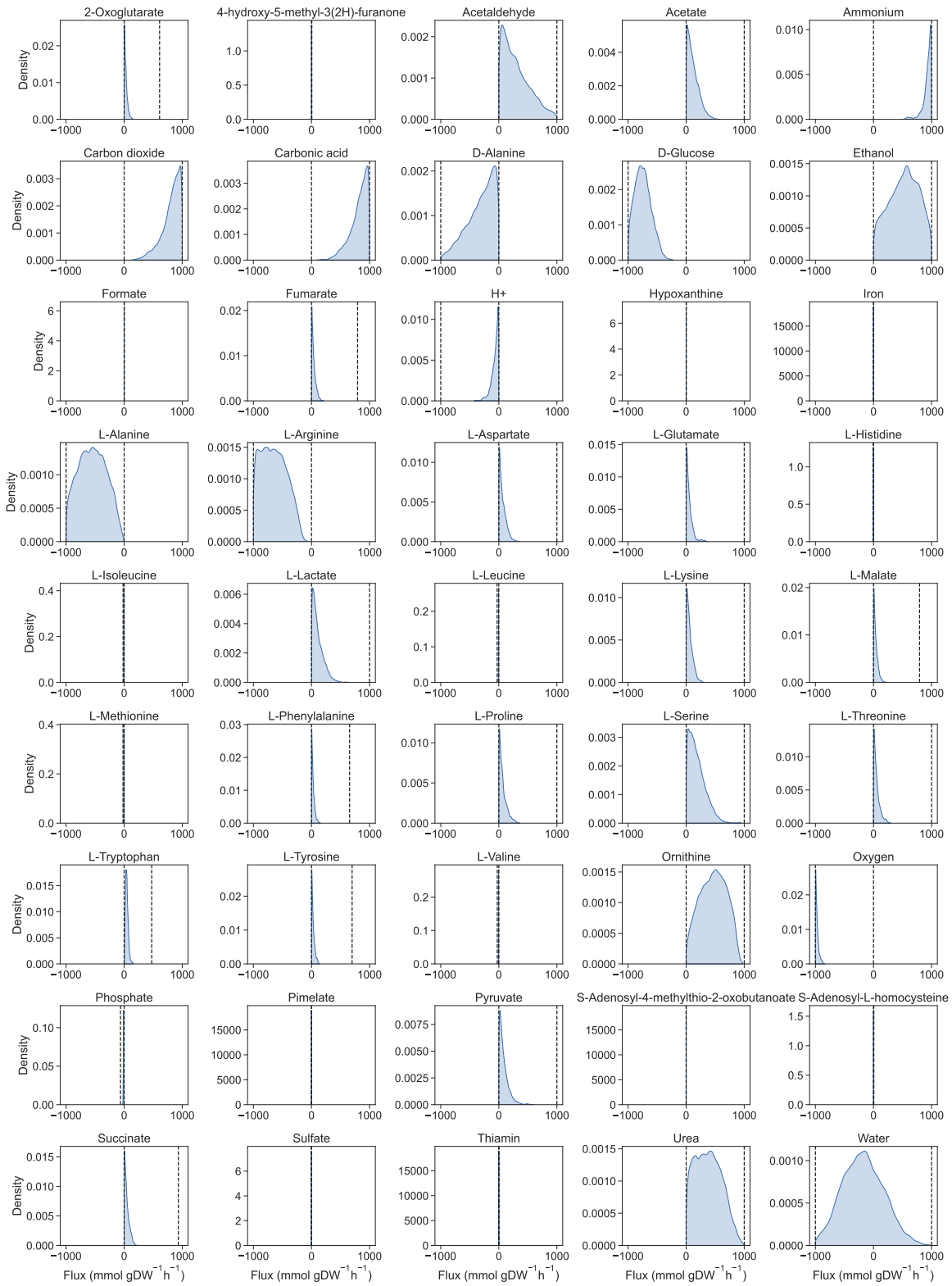


Fig. S8. Flux probability distributions for metabolite exchanges in iIT341 from 100,000 random flux distributions sampled with PTA. Dashed lines indicate feasible flux ranges from FVA.

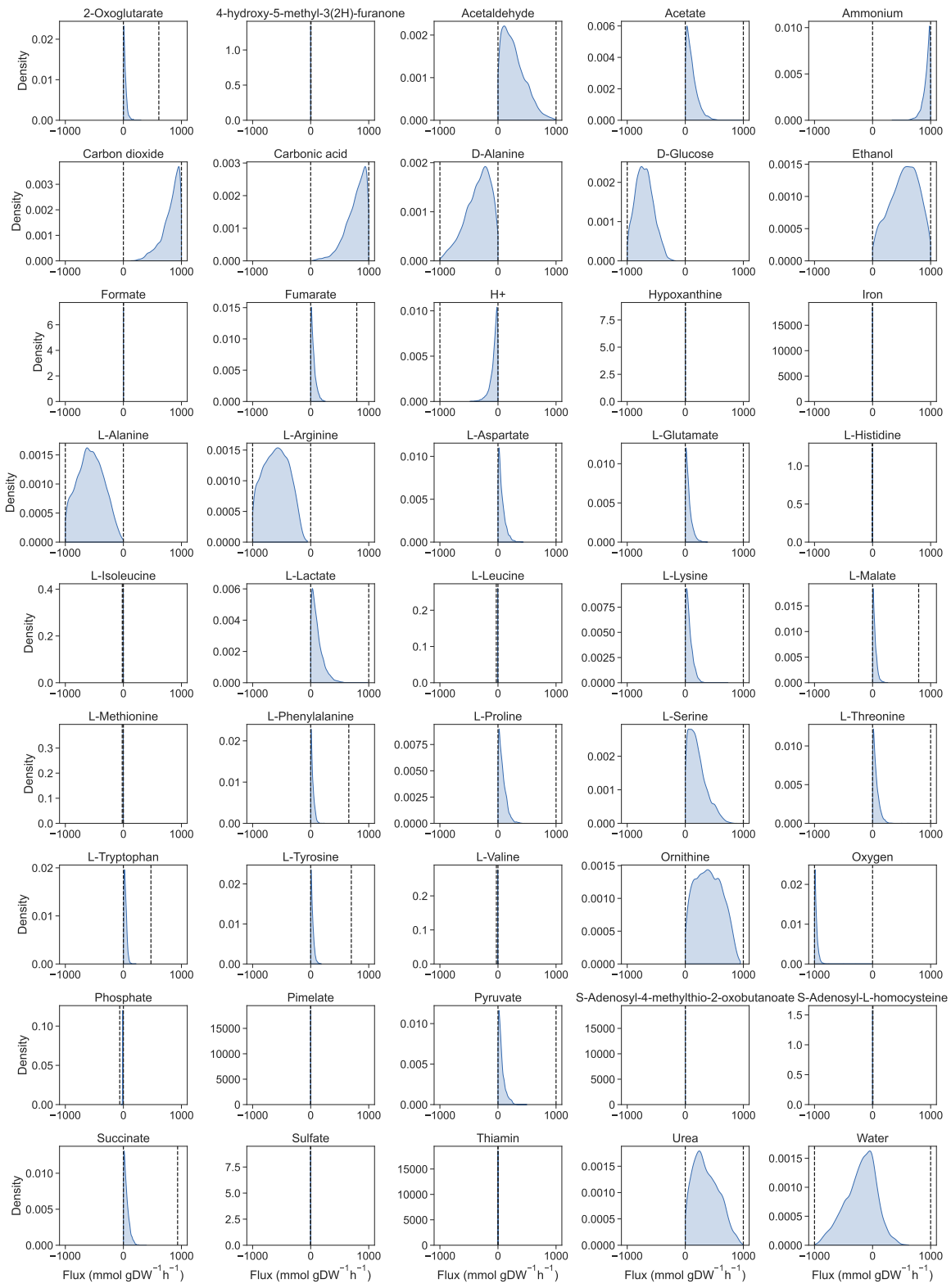


Fig. S9. Flux probability distributions for metabolite exchanges in iIT341 from 100,000 random flux distributions sampled with OptGP. Dashed lines indicate feasible flux ranges from FVA.

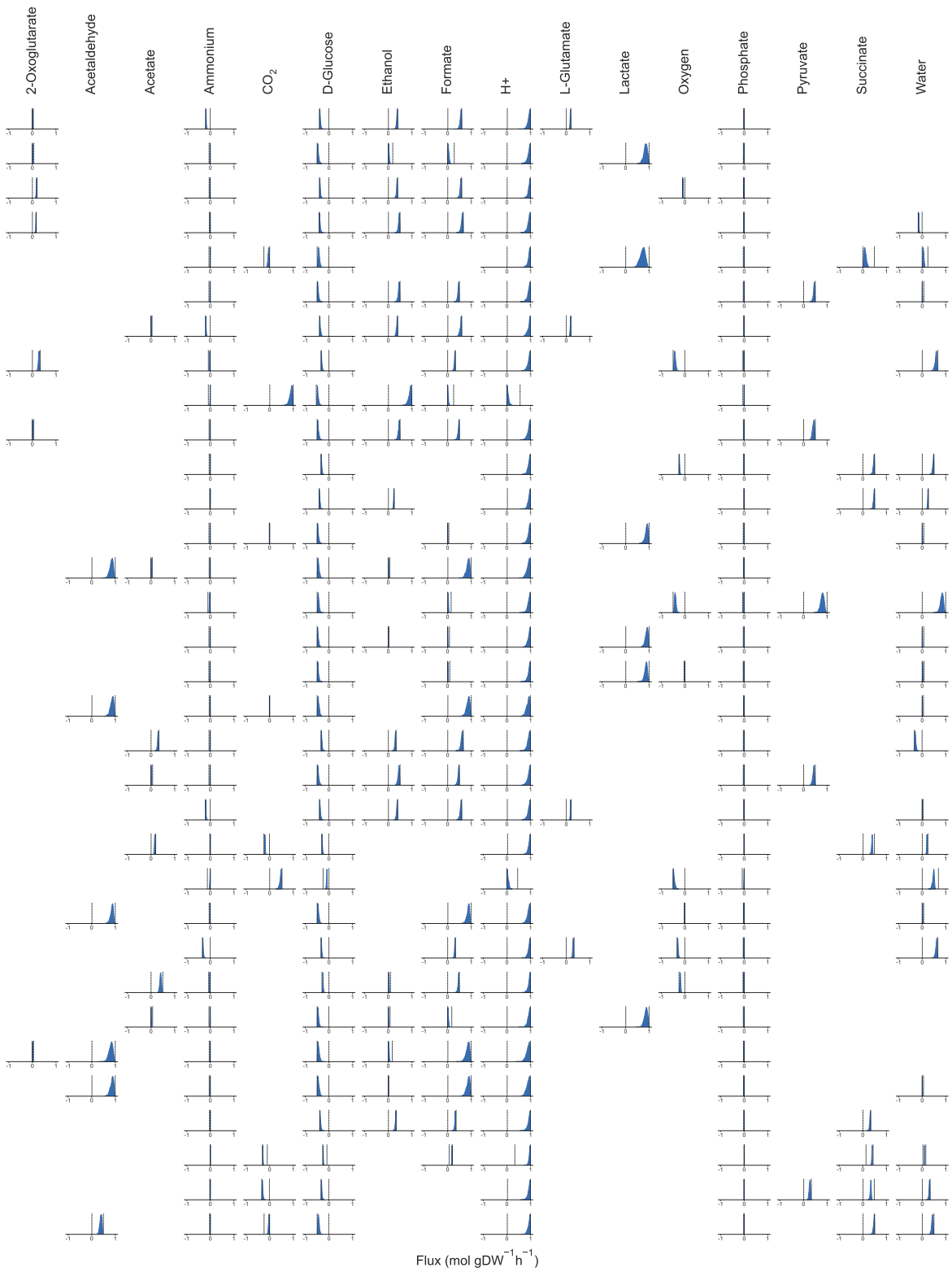


Fig. S10. Flux probability distributions for metabolite exchanges in each MP from e_coli_core. OptGP was used to sample 1,000 random flux distributions for each MP. Rows are MPs and columns are metabolites. Dashed lines indicate feasible flux ranges from FVA.

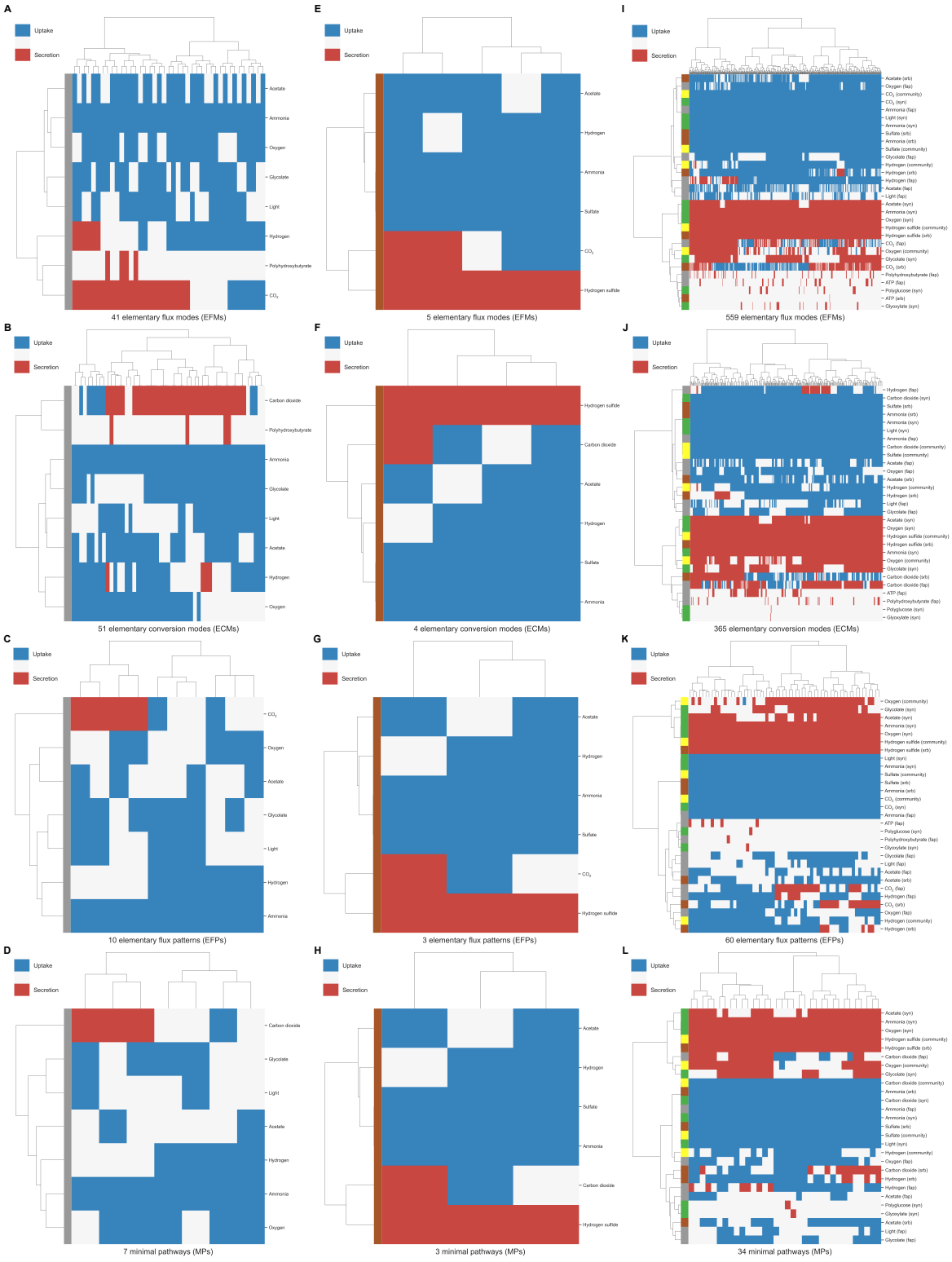


Fig. S11. Clustered heatmaps of (**A**) EFMs for fap, (**B**) ECMs for fap, (**C**) EFPs for fap, (**D**) MPs for fap, (**E**) EFMs for srb, (**F**) ECMs for srb, (**G**) EFPs for srb, (**H**) MPs for srb, (**I**) EFMs for syn, (**J**) ECMs for syn, (**K**) EFPs for syn, and (**L**) MPs for syn. Rows are metabolites, columns are pathways, and each cell indicates metabolite uptake (blue) or secretion (red) in a pathway. Rows and columns are clustered by Ward's minimum variance method and rows are colored to indicate individual microbes and the community (yellow). Unique growth-supporting flux patterns are shown.

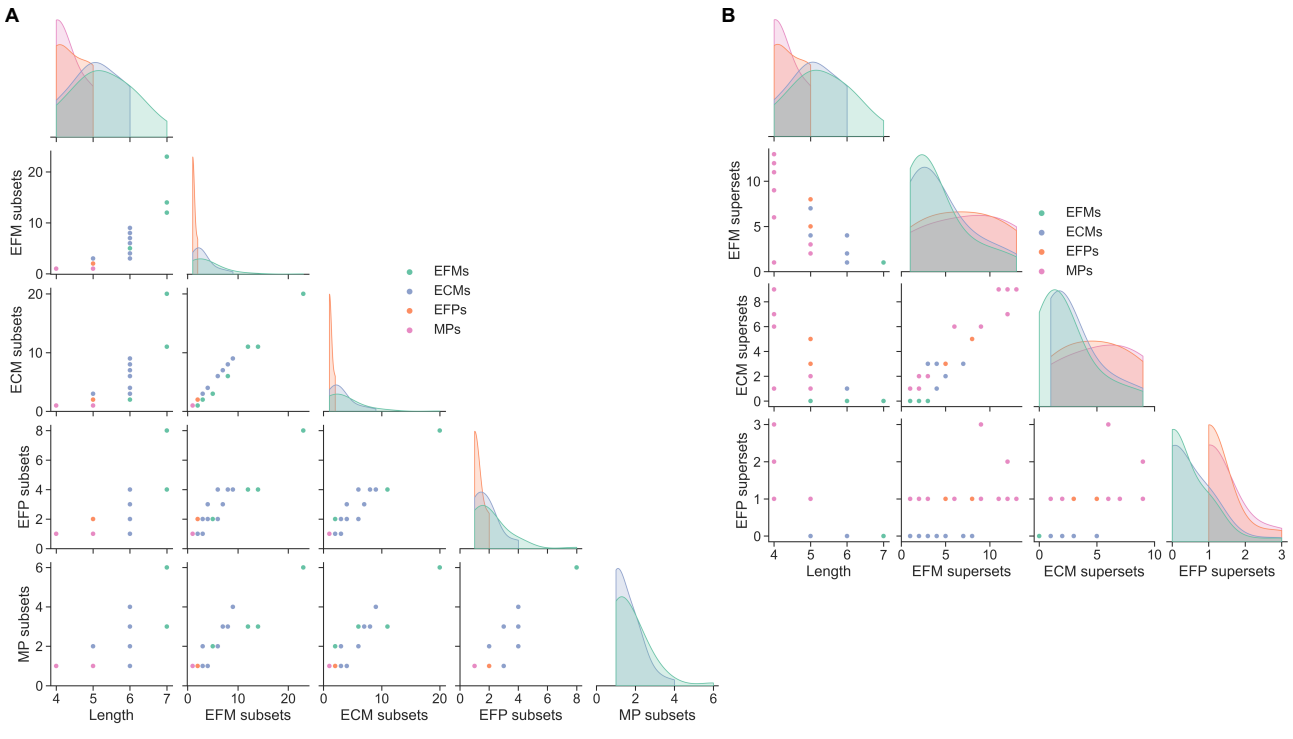


Fig. S12. Pairwise relationships between (**A**) pathway length and number of EFM, ECM, EFP, and MP subsets and (**B**) pathway length and number of EFM, ECM, and EFP supersets for fap, srb, and syn individually.

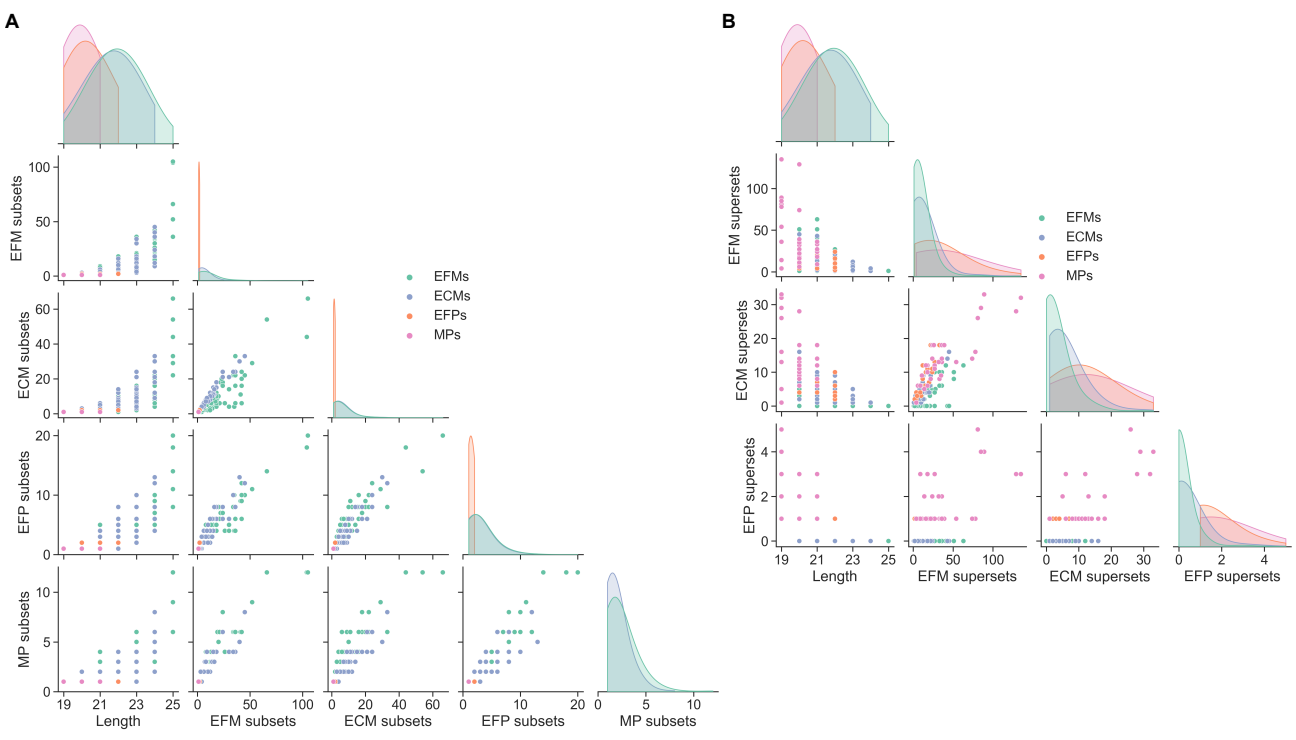


Fig. S13. Pairwise relationships between (**A**) pathway length and number of EFM, ECM, EFP, and MP subsets and (**B**) pathway length and number of EFM, ECM, EFP, and MP supersets for the microbial community model.

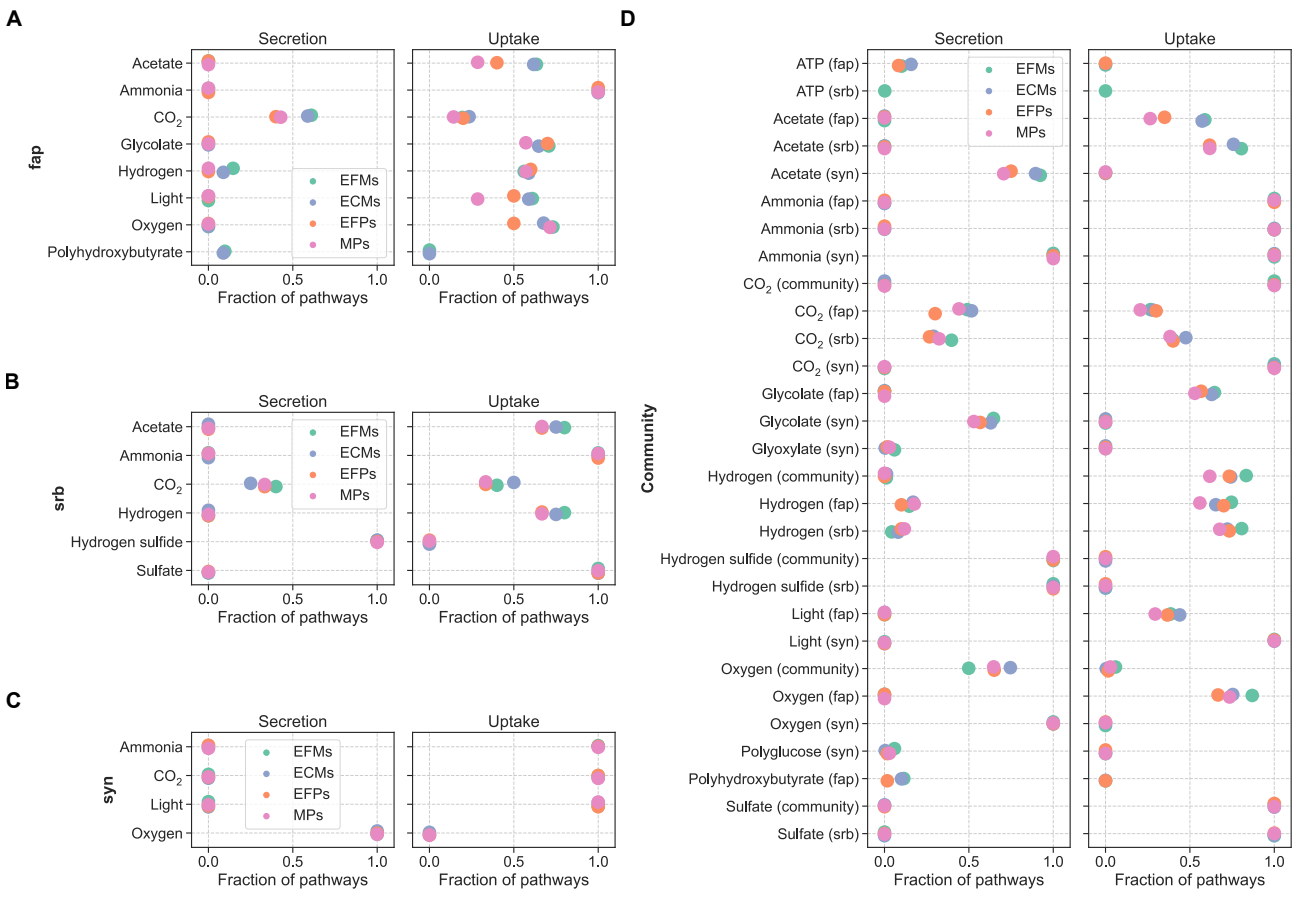


Fig. S14. Metabolite exchange frequencies (fraction of pathways including secretion or uptake of each metabolite) for (**A**) fap, (**B**) srb, (**C**) syn, and (**D**) the microbial community model.

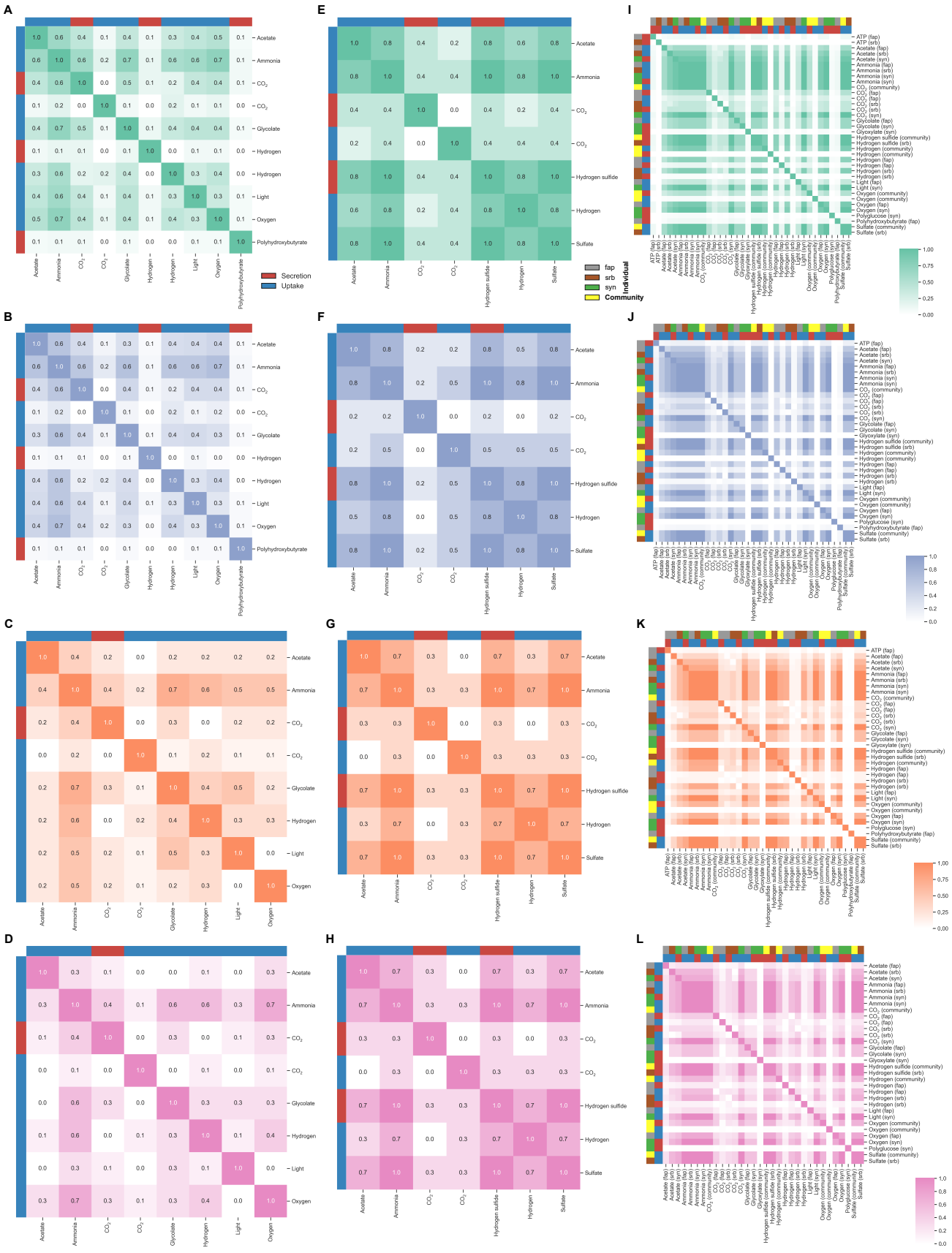


Fig. S15. Pairwise metabolite exchange frequencies (fraction of pathways including secretion or uptake of each metabolite pair) for (**A**) EFPs for e_coli_core, (**B**) ECMs for e_coli_core, (**C**) MPs for e_coli_core, (**D**) MPs for iIT341, and (**E**) MPs for iIT341. Rows and columns are metabolites and each cell indicates the pairwise exchange frequency of two metabolites.

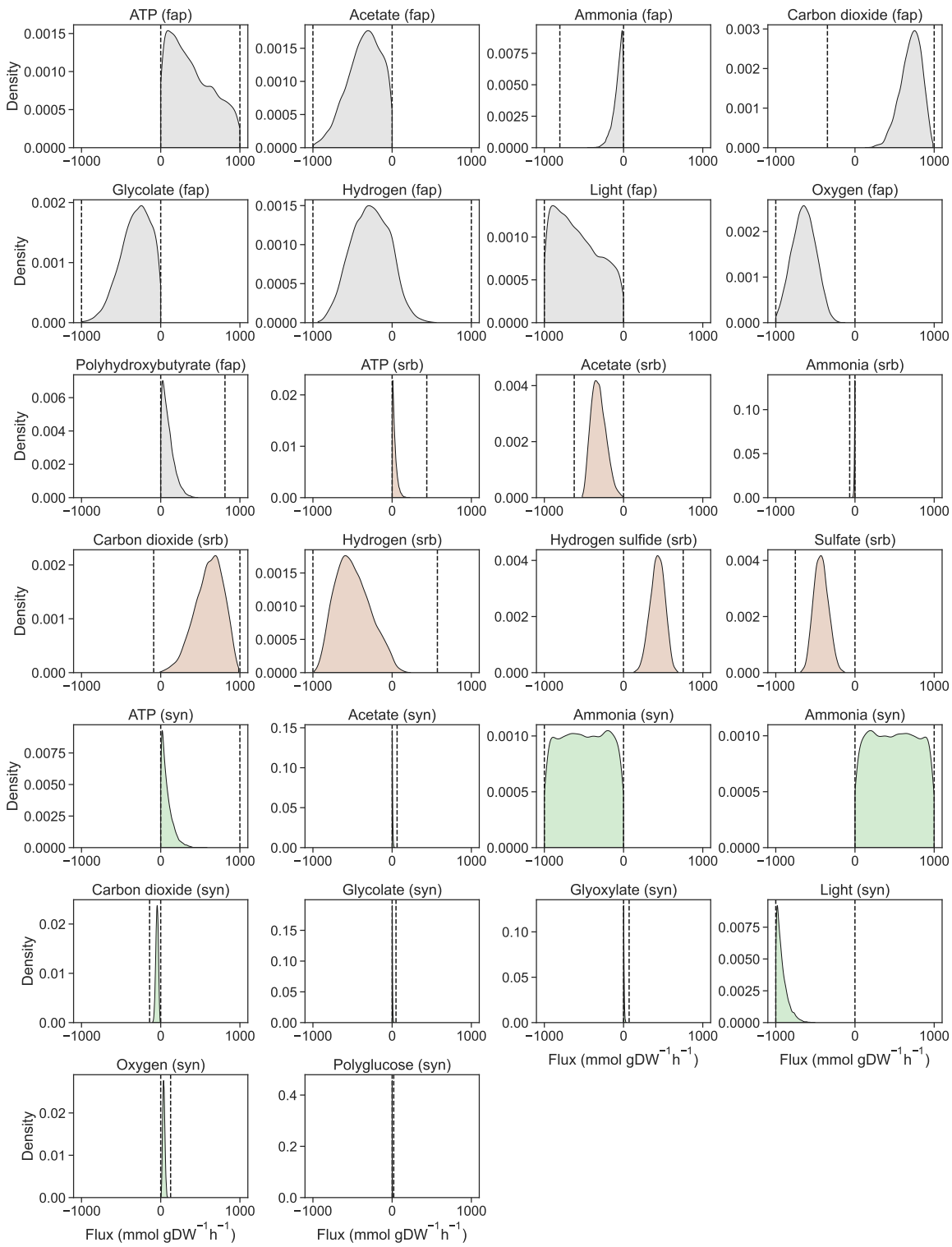


Fig. S16. Flux probability distributions for metabolite exchanges in fap, srb, and syn individually from 100,000 random flux distributions sampled with OptGP. Dashed lines indicate feasible flux ranges from FVA.

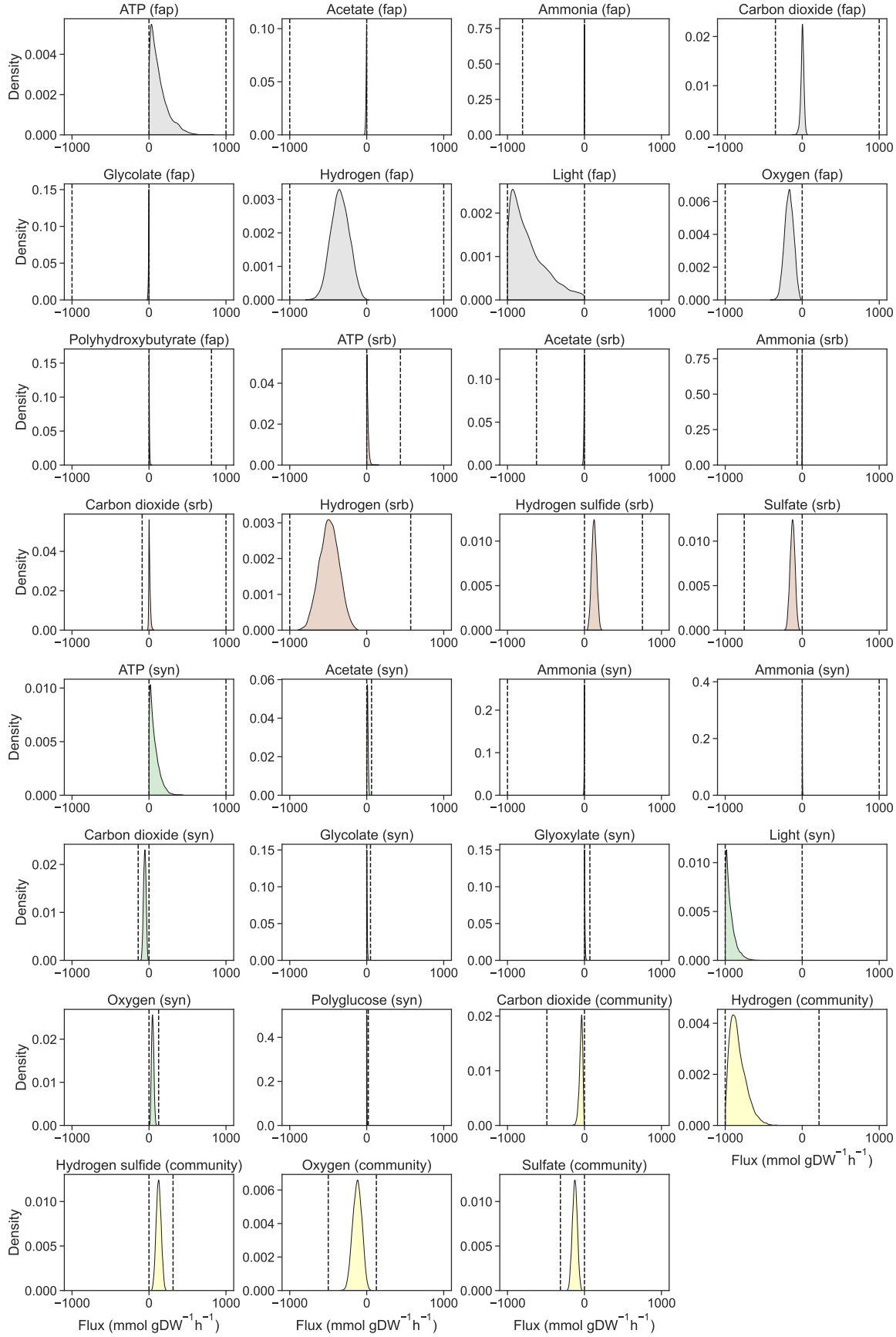


Fig. S17. Flux probability distributions for metabolite exchanges in the microbial community model from 100,000 random flux distributions sampled with OptGP. Dashed lines indicate feasible flux ranges from FVA.