



HAL
open science

Using automated reasoning to explore the metabolism of unconventional organisms: a first step to explore host–microbial interactions

Clémence Frioux, Simon Dittami, Anne Siegel

► To cite this version:

Clémence Frioux, Simon Dittami, Anne Siegel. Using automated reasoning to explore the metabolism of unconventional organisms: a first step to explore host–microbial interactions. *Biochemical Society Transactions*, 2020, 48 (3), pp.901-913. 10.1042/BST20190667. hal-02569935

HAL Id: hal-02569935

<https://hal.science/hal-02569935>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using automated reasoning to explore the metabolism of unconventional organisms: a first step to explore host-microbial interactions

Clémence Frioux^{1,2,3}, Simon M. Dittami⁴, Anne Siegel¹

¹Univ Rennes, Inria, CNRS, IRISA, 35000, Rennes, France

²Inria Bordeaux Sud-Ouest, 33405 Talence, France

³Quadram Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UQ, UK

⁴Sorbonne Université/CNRS, Station Biologique de Roscoff, UMR 8227, Integrative Biology of Marine Models, CS 90074, Roscoff, France

May 11, 2020

Abstract: Systems modelled in the context of molecular and cellular biology are difficult to represent with a single calibrated numerical model. Flux optimisation hypotheses have shown tremendous promise to accurately predict bacterial metabolism but they require a precise understanding of metabolic reactions occurring in the considered species. Unfortunately, this information may not be available for more complex organisms or non-cultured microorganisms such as those evidenced in microbiomes with metagenomic techniques. In both cases, flux optimisation techniques may not be applicable to elucidate systems functioning. In this context, we describe how automatic reasoning allows relevant features of an unconventional biological system to be identified despite a lack of data. A particular focus is put on the use of Answer Set Programming, a logic programming paradigm with combinatorial optimisation functionalities. We describe its usage to over-approximate metabolic responses of biological systems and solve gap-filling problems. In this review, we compare steady-states of Boolean abstractions of metabolic models and illustrate their complementarity via applications to the metabolic analysis of macro-algae. Ongoing applications of this formalism explore the emerging field of systems ecology, notably elucidating interactions between a consortium of microbes and a host organism. As a first step in this field, we will illustrate how the reduction of microbiotas according to expected metabolic phenotypes can be addressed with gap-filling problems.

Keywords: Metabolic networks — gap-filling — systems biology — systems ecology — community selection — non-model organisms

1 Introduction

Systems biology and metabolism Systems biology consists in considering an organism or an interacting group of organisms as a whole rather than studying its components individually [1]. This can be contemplated through the exploration of genome-scale metabolic networks (GSMs), which contain all biochemical reactions and pathways that are expected to occur in a cell. The first published metabolic networks were associated with *model* organisms characterised by low biological complexity or widely studied by the community: *Haemophilus influenzae* [2], *Escherichia coli* [3], *Arabidopsis thaliana* [4]. The construction of these models was enabled by extensive literature-based curation and expert knowledge together with experimentation, including the possibility of genetic alterations.

Metabolic networks of unconventional organisms Since then, the rise of sequencing technologies paved the way to the study of the metabolism of thousands of organisms, with strong heterogeneity between and within taxonomic groups [5, 6]. Most of these sequenced genomes correspond to *unconventional* or

non-model organisms, that is, species with little available background knowledge, or that are difficult to cultivate so that most or all of the available information is derived from their genome sequence (possibly from metagenomic studies). The *de novo* prediction of the functions of unknown proteins is generally not contemplated for such organisms, and the large phylogenetic distance from related model organisms frequently limits the transferability of predictions from these models [7]. In the same vein, the validation of predicted functions is difficult and experiments involving unconventional organisms mostly involve perturbations of their environment. In model organisms, on the contrary, it is possible to perform genetic modifications in order to investigate the role of a targeted unknown protein. As a result, disentangling the metabolism of unconventional organisms necessitates methods that differ from those applicable to well-studied organisms.

Metabolic network models are used to mathematically predict the growth rate [8, 9] or the optimum yield (function of linear combination of rates) [10, 11] of the organism by solving optimisation problems [12, 13]. A GSM reconstruction is obtained first by taking into account all genomic information related to the organism and then by manually refining the draft according to specific knowledge about the organism (adding missing reactions, removing the ones that were falsely inferred) until the mass-balanced equilibrium of internal metabolites and co-factors is satisfied and growth can be adequately predicted [14].

In this framework, the so-called *gap-filling* step that consists in adding reactions to obtain a relevant biomass prediction, is crucial. However, especially in unconventional organisms, this gap-filling step may also bear negative impacts. A first drawback is the risk of adding reactions for which the considered organism has no associated gene, either because the associated enzyme is not annotated, or because the considered function is performed by a specific pathway that has not yet been identified. A second drawback is related to the underlying hypothesis of ensuring biomass production. Indeed, GSMs are considered to be valid when they accurately predict a biomass production. The composition of this biomass function can be confirmed experimentally, or derived from models [15, 16, 17], in accordance with literature. However, the objective is much more difficult to define when the organism cannot be cultured individually as is the case for many unconventional organisms. One example is the difficulties to grow in axenic conditions one of the most studied brown algae: *Ectocarpus siliculosus*. This alga, a model among the stramenopiles [18, 19] was the first brown alga to be fully sequenced [20] and has been extensively studied, notably at the metabolic scale [21, 22]. Yet, at least in standard culture media, *E. siliculosus* does not grow properly in axenic conditions, that is individually, without its associated symbionts [23]. Axenic culture leads to altered physiology and morphology of the alga, which can be restored after inoculation of bacterial isolates. These biotic dependencies are commonly observed when trying to culture unconventional organisms. For bacteria, in particular, the auxotrophies in their ecological niches are difficult to identify [24]. This advocates for prudence in the interpretation of gap-filling results when focusing on a single unconventional organism, and more generally, for organisms that live in symbiosis where several objective functions may have to be combined [25].

Host-microbe interactions In the last decade, increasing knowledge about the wide range of interactions occurring between hosts and microorganisms or between microorganisms in their natural environment has led to expanding the concept of systems biology to what can be called *ecosystems biology* or *microbial systems ecology* [26, 27]. *Microbiotas* are communities of microorganisms that can be found in a given environment [28], or in association with a specific *host* species, such as the highly-studied human gut microbiota [29]. The organisms that form these microbiotas are frequently unculturable using standard laboratory techniques and fall into the definition of unconventional organisms [30]. As a consequence, they are less studied experimentally, or even not studied at all with all information about them provided by omics data. In addition to interacting with their abiotic environment, these symbiotic organisms have mutual interactions, leading to communities with distinct physiology and phenotypes.

The recent technical improvements in metataxonomics and metagenomics have allowed for an increased focus on these complex communities including the unculturable part of living organisms [31]. This led to an expansion of the limits of the tree of life [32] and is now providing biologists and modellers with an unprecedented amount of genomic data regarding host-microbial systems [33], inducing a change of paradigm in the study of biological species. Recent research advocates that an organism can no longer be considered

as an independent individual but rather as a member of a complex ecosystem [34]. Highly studied host-microbial systems include the human gut [35], the rhizosphere [36], arthropods [37] or marine organisms [38]. The impact of this change of paradigm on the field of metabolic networks is high in systems biology. The goal of this mini-review is to revisit the role of *gap-filling methods* introduced in the process of construction of a metabolic network. We discuss in particular how the study of species communities can be reformulated as a gap-filling problem and why the steady-state hypothesis may have to be leveraged by a more qualitative system abstraction. This leads us to introduce new optimisation problems derived from the gap-filling problem that can be efficiently applied to the study of the metabolism for individuals and communities of unconventional organisms.

As the reader will find, the translation of biological concepts and insights into formal problems is highly dependent on slight differences in mathematical abstraction that can result in divergent predictions. This necessitates the use of precise notation to accurately specify different abstractions. We accept that not all readers will be familiar with such notation, but we expect that the accompanying text will make the general principles and aims of the approaches accessible.

2 Steady-state and Boolean frameworks

A metabolic network is composed of reactions that transform metabolic compounds into other metabolic compounds. From the information associated with a metabolic network, mathematical approaches should theoretically enable modelling the dynamic response of a cellular metabolism in a given medium with ordinary differential equation (ODE) models [39]. In practice, the parameters of these models cannot be numerically fitted because of both non-linearities and a lack of experimental possibilities of manipulating a metabolic system. Kinetic models therefore target parts of the metabolism [40] but are not contemplated at genome-scale. In order to overcome this obstacle, several modelling abstractions have been introduced, such as the *steady-state* abstraction [41] relying on Flux Balance Analysis (FBA) [9] and the *Boolean* abstraction based on network expansion [42]. Both abstractions predict the family of reactions (and compounds) that can be activated (produced) from the compounds of the extracellular medium.

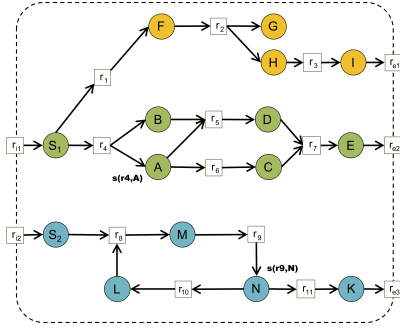
Formal definition of a metabolic network Let us introduce several notations for the sake of clarity of the theoretical background. We formally describe a metabolic network by a bipartite directed graph $G = (R \cup M, E, stc)$, where R and M stand for reaction and metabolite nodes. When $(m, r) \in E$ (respectively $(r, m) \in E$), with $m \in M$ and $r \in R$, the metabolite is called a *substrate* (respectively *product*) of the reaction r . The edge labels $stc : E \rightarrow \mathbb{R}$ describe the *stoichiometric coefficients* of the considered compounds in the considered reactions. Such coefficients are gathered in the *stoichiometric matrix* \mathbf{S} . Each reaction $r \in R$ is associated with a variable v_r which represents the reaction flux activity. The complete vector $\mathbf{v}(\mathbf{m})$ is known as the rate laws of the system [43]. It is bounded by a vector of lower bounds \mathbf{b}_l and upper bounds \mathbf{b}_u [44]. Media compounds representing available nutrients are formally denoted by S , with $S \subseteq M$.

Steady-state abstraction Here, internal compounds are assumed not to be accumulated so that the system behaviour is constrained by linear relations $0 = \frac{d\mathbf{m}(t)}{dt} = \mathbf{S}\mathbf{v}(\mathbf{m}(t)) = \mathbf{S}\mathbf{v}$, where the $\mathbf{v}(\mathbf{m}(t))$ function represents the rate laws of the system [43]. The steady-state hypothesis is known to be valid for relatively short time slots (several minutes) assuming that regulatory transcriptional and signalling timescales are disjoint [45]. In this context, *activated reactions* from a medium are those for which an admissible flux can be carried in at least one flux distribution of the system. The steady-state abstraction is widely used to construct and analyse metabolic models [46, 47, 48] and their association in small communities [49, 50, 51]. In the following, as detailed in Fig. 1, activated reactions from a medium S will be denoted by the set $active_G^s(S)$.

Boolean abstraction In the growth phase, the steady-state hypothesis of equilibrium for internal metabolites is not expected to be met [52, 53], although FBA-based predictions can be accurate [54]. The *Boolean*

	steady-state abstraction	Boolean abstraction
Underlying dynamics	$\frac{d\mathbf{m}(t)}{dt} = \mathbf{S} \mathbf{v}(\mathbf{m}(t))$	$\phi(\bar{\mathbf{m}}) = \bigvee_{\substack{r \in R \\ m \in \text{product}(r)}} \bigwedge_{s \in \text{substrate}(r)} \bar{s}$
r_{obj} is activated if there exists such that (steady-state) (activation) (culture medium) (boundaries)	$r_{obj} \in \text{active}_G^s(S)$ $\mathbf{v} \in \mathbb{R}^{\text{size}(R)}$ $\mathbf{S}\mathbf{v} = 0$ $\mathbf{v}_{r_{obj}} > 0$ $\mathbf{v}_{\text{import}}(s) \geq 0$ for all $s \in S$ $\mathbf{b}_l \leq \mathbf{v} \leq \mathbf{b}_u$	$r_{obj} \in \text{active}_G^b(S)$ $\bar{\mathbf{m}} \in \{0, 1\}^{\text{size}(M)}$ $\phi(\bar{\mathbf{m}}) = \bar{\mathbf{m}}$ $\mathbf{m}_{obj} = 1$, for all $m_{obj} \in \text{substrate}(r_{obj})$ $\phi(s) = 1$ for all $s \in S$ $\bar{\mathbf{m}}$ has the smallest number of 1's with the above properties

(a)



(b)

Reactions	Activated reaction in the steady-state abstraction	activated reaction in the Boolean abstraction
r_1, r_2, r_3, r_{e_1}	no	yes
$r_4, r_5, r_6, r_7, r_{e_2}$	yes iff $\frac{s(r_4, B)}{s(r_4, A)} = 1/2$	yes
$r_8, r_9, r_{10}, r_{11}, r_{e_3}$	yes iff $s(r_9, N) \neq 1$	no

(c)

Figure 1: Different abstractions of metabolism. (a) Formal constraints associated with the steady-state and the Boolean abstractions of a metabolic network. In both cases, the system response is computed by first searching for fixed points of equations on real or Boolean values and then constraining the search with conditions related to the metabolic network properties. (b and c) A metabolic network consisting of eleven internal reactions, two import reactions, and three export reactions. All the stoichiometric coefficients are assumed to be equal to 1 except for the products of r_4 ($s(r_4, A)$, $s(r_4, B)$ are left free) and the input of r_9 ($s(r_9, N)$ is left free). Then activation of each reaction depends both on the chosen abstraction and on the values of the stoichiometric coefficients.

hypothesis is modelled by rules which activate each reaction as soon as all its substrates have been made available by the activation of former reactions. In this framework, activated or inactivated reactions, as well as producible or unproducible compounds can have either 0 or 1 as discrete values. A metabolic compound is producible either if it is a medium compound or if it is the product of an activated reaction. A reaction is activated only if all its substrates are producible. This rule is applied until no additional reaction can be activated, leading to a steady-state of the system that models the capability of metabolic activation from the medium compounds. This is the main concept underlying network expansion introduced by Ebenhöf and colleagues [42, 55] that has been applied to the study of metabolic networks both individually [56, 57, 58, 59] and in communities [60, 61, 62, 63]. Please note that, according to Boolean rules, reversing a flux requires to change the value of the input fluxes, so that the Boolean hypothesis implicitly assumes that metabolic fluxes cannot be reversed during the response to a perturbation.

Formally, this hypothesis can be modelled as follows. Let $\bar{m} \in \{0, 1\}$ be a Boolean variable associated with each metabolic compound $m \in M$ in the network. The associated Boolean dynamic is defined by $\phi(\bar{\mathbf{m}}) = \bigvee_{r \in R | m \in \text{product}(r)} \bigwedge_{s \in \text{substrate}(r)} \bar{s}$. Metabolic compounds that are producible from nutrients S are those whose associated Boolean variable equals 1 for any fixed point of ϕ having nonzero values for nutrients. Equivalently, they are obtained by computing the fixed point of ϕ having the minimum number of nonzero values while having nonzero values for nutrients. A reaction is considered activated if all its substrates are

producible. Such reactions are denoted by $active_G^s(S)$ (see the formalisation in Fig. 1). As shown in [42], activated reactions in a Boolean framework can be computed recursively.

Differences between abstractions Although these two concepts of activated reactions seem close in terms of dynamics, it appears that they differ slightly in the way they model the impact of internal cycles. Differences are shown in Fig. 1. Reactions r_{i_1} and r_{i_2} ensure the import of nutrients S_1 and S_2 from the extracellular medium. In the Boolean abstraction, imports initiate the production of I and E through the activation of the reactions r_3 and r_7 . In the steady-state abstraction, however, I is not producible because the activation of r_2 would prevent the mass-balance equilibrium of G , that constitutes a dead-end metabolite. In addition, E can be produced if and only if the ratio $\frac{s(r_4,B)}{s(r_4,A)} = 1/2$; otherwise, internal metabolites would accumulate. Another difference between the Boolean and the steady-state abstractions is related to the production of K . In the monotonous abstraction, the reaction r_8 cannot be activated because it requires the production of L , which depends on r_8 itself and therefore cannot be directly produced from the medium. Therefore, compound K cannot be produced. On the contrary, K can be produced by the system under the steady-state abstraction because the cycle $\{r_8, r_9, r_{10}\}$ can be self-activated and produces K , as well as L as soon as the linear constraints allow v_{11} to have a nonzero value. According to the linear constraints, this is possible if and only if $s(r_9, N) \neq 1$, which correspond to the case when the cycle does not require all the produced compounds to self-regenerate.

In order to reconcile these different interpretations we can introduce a concept of *hybrid activation* corresponding to verifying both the steady-state and the Boolean activation conditions. This is the most stringent notion of activation which verifies both that fluxes can occur in a metabolic network according to mass-balance but also that all internal cycles can be fed by import reactions. However, Fig. 1 shows that only a few reactions in a metabolic network may satisfy these properties. Formally, we define that $r_{obj} \in active_G^h(S)$ if $r_{obj} \in active_G^b(S) \cap active_G^s(S)$.

These examples confirm, as discussed in [64, 65], that the steady-state abstraction of a metabolic network is highly sensitive to the network stoichiometry and to putative accumulations of internal metabolites. Under the assumption that a few cofactors are present at the initiation of the system, the Boolean abstraction has similar predictions as the steady-state abstraction in terms of metabolite production while being more resilient to network inaccuracies. In this sense, it appears suitable for a preliminary study of the metabolism of unconventional organisms. However, the Boolean abstraction fails to take into account *de novo* synthesis of compounds involved in cycles (for example co-factors) and mass-balance equilibrium. The treatment of these metabolites differs between studies: they can be added into the list of available metabolites or removed from reactions [66, 67, 68, 69]. Therefore, while the main characteristics of a metabolic network can be identified with network expansion, the steady-state abstraction is required to elucidate behaviours linked to stoichiometry constraints.

3 Gap-filling problems for individual and community modelling

As explained in the introduction, a metabolic network built from genome annotations or orthology searches is often unable to predict either biomass production or the production of experimentally observed metabolic compounds. This is either due to errors or missing knowledge in the genome annotation procedures or denotes the inability of a species to grow without other symbiotic species. In this section, we detail why the extended concept of the *gap-filling procedure* is useful to address both the issue of curating a single-species metabolic network and of studying the role of metabolic complementarities between species.

The generic principle underlying gap-filling algorithms of metabolic networks is to perform a selection of reactions within a reactions database in order to restore the functionality of a model with respect to an expected objective. The selection of reactions is often performed according to a parsimonious principle aiming at minimising the number of modifications of the system. When possible, predicted reactions to be added to the system are validated with genome-based or knowledge-based studies, so that the algorithms can be run iteratively if the predicted reactions appear to be irrelevant [70]. The result of the gap-filling

method is both dependent on the framework used to define a functional model (steady-state or Boolean) and on the database in which reactions are picked-up.

Formalisation of the gap-filling problem Following a parsimony principle, the *metabolic gap-filling problem* aims at selecting, in a database of putative reactions, a set of reactions with a minimum size such that all reactions that are experimentally known or expected to be activated are predicted *in silico* to be activated from the growth medium according to the new metabolic network.

The formal description of this problem is as follows. We denote by $active()$ the abstraction of reaction activation in a metabolic network. It can rely on the steady-state formalism or the Boolean one. The following definitions are independent from the chosen formalism. We consider a metabolic network $G = (R \cup M, E, stc)$. We denote by $G' = (R' \cup M', E', stc')$ a database of putative reactions, that is, formally, another metabolic network. For any set of reactions $R'' \subseteq R' \setminus R$, the extended metabolic network of G according to R'' is naturally defined to be: $extension(G, G', R'') = ((R \cup R'') \cup (M \cup M''), E \cup E'', stc'')$ with $M'' = \{m \in M' \mid r \in R'', m \in \text{substrate}(r) \cup \text{product}(r)\}$, $E'' = E' \cap ((M'' \times R'') \cup (R'' \times M''))$ and $stc'' = stc \cup stc'$. Let us assume that the growth medium is composed of the metabolic compounds $S \subseteq M$ and that $R_T \subseteq R$ is a set of reactions that are experimentally known or expected to be activated. To avoid too many solutions, the metabolic gap-filling problem selects a set of reactions $R'' \subseteq R' \setminus R$ with a minimum size such that all reactions in R_T are activated from the growth medium, in the extension of the initial metabolic network G with respect to the new reaction set R'' .

$$\begin{array}{ll} \text{Gap-filling optimisation problem} & \begin{array}{l} \text{minimize} \quad \text{size}(R'') \\ \text{subject to} \quad R_T \subseteq active_{extension(G, G', R'')}(S). \end{array} \end{array}$$

In this optimisation problem, the search space to explore is the set of the $2^{\text{size}(R')}$ combinations of reactions present in the completion database R' . Several frameworks have been introduced to solve this problem or its derivatives [71]. In the steady-state abstraction, the problem is solved as a linear programming (LP) optimisation problem [72, 73]. Regarding the Boolean abstraction, Boolean satisfiability problem (SAT)-based algorithms can be used to efficiently solve the underlying combinatorial optimisation problem [58]. Constraint propagation approaches are used to solve the problem associated with the hybrid activation framework [74] that associates steady-state and Boolean abstractions. As a direct consequence of the differences between the abstractions highlighted in the previous section, the result of the gap-filling procedure is highly sensitive to the considered algorithm: gap-filling based on Boolean abstractions does not take internal cycles into account, whereas steady-state abstractions focus on appropriately handling import and export reaction to balance mass-equilibria.

An underlying hypothesis of gap-filling algorithms is that the considered species are able to sustain their growth with the available nutrients. Although this hypothesis is valid for many prokaryotes studied experimentally, it is less straightforward, especially for animals and plants, which harbour major biotic interactions [75]. This is further evidenced by the difficulties encountered in cultivating unconventional organisms, notably eukaryotes, in axenic conditions, highlighting the metabolic dependency of those species to other organisms [76, 23]. Microbial interactions can provide their host or other symbionts with metabolic compounds that are otherwise costly to produce [77]. Such interactions are then suspected to be involved in the evolution of metabolic interdependencies within symbiotic communities [78, 24].

Gap-filling a gene-soup microbiota to select reduced communities Revisiting metabolism at the systems ecology scale consists in studying complementarities between species enabling a community to collectively operate metabolic functions. This can also be viewed as a gap-filling problem in which interactions between species are expected to be identified within several metabolic networks which each serve as a database of metabolic functions to fill gaps in the other networks. Similar to other analyses of communities [79], several levels of modelling can be considered for such gap-filling. The community model can be compartmentalised or grouped together. The latter is described as *gene-soup* (or lumped or mixed-bag): a microbiota is represented as a *meta-organism* or *pan-metabolism* in which all metabolites and all reactions of the organisms are

gathered in a unique abstracted compartment. This gene-soup framework allowed for instance the evaluation of the effect of obesity and inflammatory bowel disease in the gut microbiota [67]. This framework is also very useful for the selection of artificial communities based on metabolic complementarities, in order to reduce the complexity of native microbiota and to test biological hypotheses on symbiotic interactions. To that end, we define a *reduced microbiota* to be a family of species that has the equivalent targeted metabolic properties as a complete microbiota.

Formally, we assume that G_1, \dots, G_N are the metabolic networks of the organisms constituting a specific community and that we aim at activating a family of metabolic functions $R_T \subset \cup R_i$ from a growth medium $S \subset \cup M_i$. The *gene-soup metabolic network* is defined by $\text{geneSoup}(G_1..G_N) = (\cup M_i \cup \cup R_i, \cup E_i, \cup \text{stc}_i)$. In this framework, the *community selection problem* corresponds to selecting the minimal number of organisms whose gene-soup metabolic network is collectively able to activate the objective reaction(s).

Community selection problem

$$\begin{aligned} & \text{minimize} && \text{size}(\{G_{i_1}..G_{i_L}\}) \\ & \{G_{i_1}..G_{i_L}\} \subset \{G_1..G_N\} \\ & \text{subject to} && R_T \subseteq \text{active}_{\text{geneSoup}(G_{i_1}..G_{i_L})}(S). \end{aligned}$$

This problem is nothing but a gap-filling problem applied to an empty network and considering the full gene-soup metabolic network as a database. Instead of considering the effect of adding reactions from an external database to an initial network as we did before, we explore how combining several metabolic networks impacts the activation of the targeted reactions. Organisms are selected to be added to an empty system (alternatively, to the metabolic network of a host organism) such that the targeted reactions are activated in the mixed-bag selected community. Therefore, instead of minimising the number of reactions added to the network, we optimise the number of organisms whose metabolic networks contain the relevant reactions to ensure the targeted objective. The search space is therefore the 2^N sub-families of the microbiota species G_1, \dots, G_N .

In practice, solving this problem is often less complex than solving the gap-filling problem for individual organisms, and notably, one solution can be found in a simplified steady-state model [66]. However, many different combinations of bacteria are expected to realise the targeted metabolic objective due to the functional redundancy of microbiotas [80, 81], therefore the space of solutions can be large. We confirmed this hypothesis in [63] by performing a complete enumeration of solutions to the optimisation problem in the Boolean framework, evidencing a strong redundancy in artificially-reduced communities associated with the gut microbiota. 86.5% of source-product pairs connected by a metabolic pathway in the gene-soup gut microbiota could be equivalently operated by more than 100 minimal artificial communities, and 49.8% could be operated by more than 1,000 equivalent minimal communities.

Gap-filling a compartmentalised microbiota to identify metabolic complementarity The gene-soup framework is intrinsically limited by the fact that the cost of interactions between species is not taken into account, nor are they precisely identified. To address this obstacle, the *compartmentalised framework* is designed such that all metabolic networks of the microbiota form compartments that additionally share an external compartment. *Exchange reactions* between compartments operate the export and import of metabolic compounds between the different metabolic networks. Compartmentalised communities have been modelled for the design of growth media [82], the inference of interactions in communities [51, 83], or the study of microbiota evolution [25, 84]. In this framework, *compartmentalised reduced microbiota* are defined to be families of organisms that have metabolic properties equivalent to the initial community, while taking into account the energetic cost of metabolic exchanges between species (import and export). The weight of exchange reactions can be a transport cost (if available) or they can be considered to have equal weights if no information allows differentiating them.

Formalising this concept requires introducing $\mathcal{E} \subset \text{exchg}(G_1..G_N) = \{(r_m, i, j) | m \in M_i \cap M_j, i \neq j\}$ a family of metabolic exchanges allowed in the microbiota and possibly the host. The associated compartmentalised metabolic network, denoted by $\text{cptModel}(G_1..G_N, \mathcal{E}) = (\overline{M}, \overline{R}, \overline{E}, \overline{\text{stc}})$ is defined by associating each metabolite and reaction with the index of the organism it belongs to: metabolites in \overline{M} are denoted by (m, i) , reactions in \overline{R} are denoted by (r, j) , and edges in \overline{E} are denoted by (e, i, j) . The stoichiometric

function is naturally deduced from the individual species' metabolic networks. Compartmentalising seeds, through the operation $\text{cptSeed}(G_1..G_N, S) = \cup_{1..N}(S \cap M_i \times \{i\})$ models the fact that not all organisms have all the seeds in their individual metabolic networks. The reduced microbiota is characterised by the fact that it requires a minimal weight of exchange reactions in \mathcal{E} to ensure the activation of the targeted reactions.

$$\begin{array}{ll}
 \text{Exchange reactions problem} & \text{minimize} \quad \text{weight}(\mathcal{E}) \\
 & \mathcal{E} \subset \text{exchg}(G_1..G_N) \\
 & \text{subject to} \quad \forall r_T \in R_T, \exists i \leq N, \\
 & \quad (r_T, i) \in \text{active}_{\text{cptModel}(G_1..G_N, \mathcal{E})}(\text{cptSeed}(G_1..G_N, S))
 \end{array}$$

Differences between gene-soup and compartmentalised microbiota for community reduction are illustrated in [63]. As before, this problem can be seen as a gap-filling problem: the reference database in which reactions are searched for is now the family \mathcal{E} of all possible exchanges between organisms. The size of the search space is $4^{\sum_{i < j} \text{size}(M_i \cap M_j)}$, which is larger than the search spaces considered in the two previous optimisation problems. This problem was solved in [69] by using a Steiner graph approach and applied to small size communities. For larger size communities, a heuristic has been introduced by combining the solving of the gene-soup and the compartmentalised gap-filling problems, with good results on microbiota containing hundreds of species [63].

4 Solving optimisation problems associated with gap-filling

The three optimisation problems stated above can be addressed with logic solving approaches. Answer Set Programming (ASP) [85] is a declarative approach oriented toward both knowledge processing with a non-monotonous logic programming approach and combinatorial optimisation problem-solving, such as the optimisation problems presented above when considering the Boolean abstraction. Similarly to the solving of linear problems using LP, the problem is formulated in a dedicated language while the solving of the problem is left to the solver [85]. A first advantage of ASP is its high-level modelling language: problems are formulated according to a first-order propositional logic which provides expressive power and flexibility in problem descriptio. As shown in [57, 74], this flexibility for extending a problem statement allowed formulating an ASP program to solve the three gap-filling problems in a unified framework [86]. A second advantage of ASP is the high performance underlying solvers [87], designed to take advantage of SAT-based solving techniques, enriched with the capability of exploring the space of solutions, for instance by performing the intersection or the union of all solutions in addition to the more computationally demanding alternative of enumerating them. This appeared to be very useful for exploring the search space of the gap-filling problems instead of selecting a single solution [58, 63].

Roughly, in ASP, the focus is on the problem specification and reasoning rather than the algorithmic part. A problem is expressed as a set of logical rules (clauses) $h : -b_1, \dots, b_m \text{ not } b_{m+1}, \dots, \text{not } b_n$, where each b_i and h are literals. In fact, each proposition is a predicate, encoded by a function whose arguments can be constant atoms or variables over a finite domain. A rule states that the head h is proven to be true if the body of the rule is satisfied, i.e. b_1, \dots, b_m are true and it cannot be proven that b_{m+1}, \dots, b_n are true. By default, all atoms are supposed to be false unless a rule proves that it is true. Optimisation rules can be described with specific predicates. Together, the syntax allows for formulating a very large panel of combinatorial optimisation problems and possibly to combine them in a unified formalism.

The main limitation of ASP resides in the fact that is ill-suited for solving linear problems such as those yielded by the steady-state abstraction. This limitation has been partially solved in [86] by relying on the theory reasoning capacities of an ASP solver that allows extending ASP to express and solve linear constraints in addition to combinatorial constraints. These technologies are therefore very promising to provide a general framework to model and solve all optimisation problems related to metabolism.

5 Applications to unconventional organisms: the example of macroalgae

The example of macroalgal metabolism illustrates how the combinations of different semantics and different gap-filling problems can help elucidate the characteristics of unconventional species at the metabolic scale, from their individual metabolism to host-microbial interactions.

Using Boolean abstraction to shed light on the evolution of the algal metabolic processes

Brown algae (part of the stramenopiles) are important members of marine ecosystems. *E. siliculosus* is a model to study the biology of these organisms [18]. Following the publication of its genome in [20], a first reconstruction of its metabolic network was published four years later [21]. The authors used the Boolean abstraction to gap-fill the network resulting in 44 reactions identified as sufficient to activate the production of all the targeted metabolic compounds and the algal biomass. Such computation of Boolean gap-filling can be calculated in a computational time of a few minutes [58]. The study of the metabolic network shed light on the evolution of metabolic processes. It suggested that *E. siliculosus* has the potential to produce phenylalanine and tyrosine from prephenate and arogenate, but does not possess a phenylalanine hydroxylase as found in other stramenopiles. It also possesses the complete eukaryote molybdenum cofactor biosynthesis pathway, as well as a second molybdopterin synthase that was most likely acquired via horizontal gene transfer from cyanobacteria.

Combining abstractions of metabolism to capture algal properties

Applied to other macroalgae, we noticed, however, that the Boolean abstraction may not always be sufficient to model the metabolic properties of a network. In [88], the study of the *Cladosiphon okamuranus* metabolic network highlighted that, although 67 reactions were sufficient to produce biomass according to the Boolean abstraction, they failed to explain the biomass production according to a steady-state framework. A method for solving the hybrid (Boolean and steady-state) problem was implemented [74] and enabled the identification of a single missing reaction needed to degrade one component that was not part of the biomass function, but accumulated during biomass production. Orthologues of proteins known to catalyse this reaction in other organisms were identified in *C. okamuranus*. The complete approach allowed studying biosynthetic pathways for carotenoids production, highlighting both reactions preserved through evolution and the specificities related to brown algae.

Similar difficulties were encountered in the study of the red alga *Chondrus crispus* [89]. For this organism, an exhaustive confrontation of the experimentally detected metabolites and knowledge databases of metabolic reactions evidenced that for many compounds, biosynthetic pathways could not be inferred with gap-filling algorithms, because of incomplete biochemical knowledge and incomplete conservation of biochemical pathways during evolution. Specific methods based on the Boolean abstraction were required to infer reactions by analogy with metabolic transformations occurring in other plants. Results suggest that even metabolic pathways previously considered as conserved, like sterol or mycosporine-like amino acid (MAA) synthesis, undergo substantial turnover, evidencing a phenomenon termed '*metabolic pathway drift*' - i.e. the fact that a given phenotype can be conserved even if the underlying molecular mechanisms are changing.

An observation from these examples and more generally from the study of high-quality GSM reconstructions is the persistence of the need for curation and the use of several semantics, methods and tools combined with each others [90, 91, 92, 22]. The development of new automatic procedures [93, 94, 95, 96] (compared in [97]) facilitates reconstructions but cannot fully substitute refinements operated by experts [14]. Indeed, the publications of high-quality GSMs mention the curation effort and literature-based improvements made to the models [98, 99]. For the algal GSMs presented above, the combination of abstractions was crucial as it enabled the assessment of the completeness of the model at each reconstruction step, thereby pinpointing the metabolic pathways that required curation.

From metabolic network gap-filling to suggestions of host-microbial metabolic complementarities The study of the metabolic network of *E. siliculosus* also shed light on the importance of symbionts for the metabolism of the algal host. As explained in [21], a gap-filling algorithm was used to fill the metabolic network of *E. siliculosus*. The GSM was further analysed in [58], and the coenzyme A biosynthesis pathway was notably scrutinised. Gap-filling suggested that the reaction producing beta-alanine was required to initiate the production of vitamin B5 (pantothenic acid), a precursor of the pathway. However, genome-based studies concluded that no corresponding gene could be identified in the algal genome for the enzyme EC 4.1.1.11 producing beta-alanine. The corresponding gene was nonetheless present in a bacterium, *Candidatus* Phaeomarinobacter ectocarpus, known to live symbiotically with the alga [100], suggesting a putative host-microbial complementarity regarding beta-alanine and other putative interactions. The absence of EC 4.1.1.11 in brown algae was confirmed in [22], although this study also highlighted a potential alternative biosynthetic pathway from 3-aminopropanal.

Computation and indirect validation of reduced microbiota The above example confirms that the result of gap-filling algorithms at the individual scale should be considered cautiously, in light of a putative role of biotic interactions. This is consistent with the results of [23] which stated that, cultured in axenic condition, *E. siliculosus* evidenced alteration of its morphology and physiology. A main objective in the context of this brown alga but more broadly for host-microbial systems, is to identify the precise mechanisms of metabolic interactions.

In this direction, the non-compartmentalised community selection problem was applied to *E. siliculosus* and ten bacteria isolated from the algal microbiome to identify metabolic complementarities. As shown in [101], the algorithm allowed predicting consortia of three bacteria that would best complement the algal metabolism. Co-culture experiments were set up with a subset of these consortia to monitor algal growth as well as the presence of key metabolites. Although bacterial communities were only modified (and not fully controlled) in the experiments, the data demonstrated a significant increase in algal growth in cultures inoculated with the selected consortia, suggesting that metabolic complementarity is a good indicator for beneficial metabolite exchanges in microbiota. These results constitute a promising application of community reduction and selection of microorganisms based on metabolic complementarity. However, this study also experimentally observed the evolution of the algal microbiome after bacterial inoculation, demonstrating the presence of bacteria that were undetected in the axenic medium. This highlights new experimental challenges to test the predictions made by automated reasoning approaches when working with unconventional organisms.

6 Perspectives

The rise of high-throughput and cost-effective sequencing paved the way for the study of the metabolism of thousands of organisms with little available background knowledge, many of which are difficult to cultivate. A change of paradigm is occurring in GSM reconstruction with the extensive use of automatic methods that are under active development in order to reduce the need for manual curation and costly experiments to refine the networks, and to account for the limited knowledge on non-cultured organisms. From now on, metabolic modelling appears conceivable for 'unconventional' or 'non-model' organisms for which the main available data is the genomic sequence. In such cases, the precision of GSM reconstruction is impaired by the lack of data available for these organisms or close relatives, impeding the use of modelling approaches based on flux optimisation and compelling the development of new approaches for their analysis.

The combination of several abstractions of metabolism (Boolean, steady-state and hybrid) can be a response to investigate the metabolic capabilities of an unconventional organism, and should be combined when possible to better understand its metabolism. In particular, approaches of logic programming and Boolean abstractions of metabolic networks are promising to predict the metabolic capacities of these organisms as well as their biological roles in symbiotic communities.

Efforts are still needed in the direction of facilitating metabolic inference for non-model organisms. In particular, the field of metabolism will benefit from methods dedicated to infer reactions beyond those present

in knowledge databases, and thereby account for the still-unknown pool of functional sequences in less-studied genomes. In addition, the microbiota context in which unconventional organisms are mostly studied has to be taken into account. Building and refining models for microorganisms that cannot be cultured remains an open challenge, and metagenomic data together with metagenome-assembled genomes need to be linked to metabolic modelling. The existence of multiple and complementary formalisms to abstract metabolism will certainly prove useful to address these challenges in the next few years.

7 Abbreviations

ASP	Answer Set Programming
GSM	Genome-Scale Metabolic network
LP	Linear Programming
MAA	Mycosporine-like amino acid
ODE	Ordinary differential equations
SAT	Boolean SATisfiability problem

8 Funding

This work was partially funded by ANR project IDEALG (ANR-10-BTBR-04) “Investissements d’Avenir, Biotechnologies-Bioressources”

9 Competing interests

The authors declare that there are no competing interests associated with this manuscript.

References

- [1] Kitano H. Systems biology: A brief overview. *Science*. 2002 mar;295(5560):1662–1664. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.1069492>.
- [2] Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry*. 1999;274(25):17410–17416. Available from: <http://www.jbc.org/content/274/25/17410.full.pdf>.
- [3] Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology*. 2003;4(9):R54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12952533><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC193654>.
- [4] de Oliveira Dal’Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in *Arabidopsis*. *Plant Physiology*. 2010;152(2):579–589. Available from: <http://www.plantphysiol.org/content/152/2/579><http://www.plantphysiol.org/cgi/doi/10.1104/pp.109.148817>.
- [5] Ellegren H. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*. 2014 jan;29(1):51–63. Available from: <https://www.sciencedirect.com/science/article/pii/S0169534713002310>.
- [6] Liu H, Deutschbauer AM. Rapidly moving new bacteria to model-organism status. *Current Opinion in Biotechnology*. 2018 jun;51:116–122. Available from: <https://www.sciencedirect.com/science/article/pii/S0958166917301714?via=ihub>.

- [7] Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*. 2017 aug;34(8):2115–2122. Available from: <https://academic.oup.com/mbe/article/34/8/2115/3782716>.
- [8] Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Current Opinion in Biotechnology*. 2003 oct;14(5):491–496.
- [9] Orth JD, Thiele I, Palsson BØ. What is Flux Balance Analysis ? *Nature biotechnology*. 2010 mar;28(3):245–248. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108565/> <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108565/pdf/nihms299330.pdf>.
- [10] Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2002 nov;99(23):15112–15117.
- [11] Schuster S, Pfeiffer T, Fell DA. Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*. 2008 jun;252(3):497–504.
- [12] Holzhütter HG. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*. 2004 jun;271(14):2905–2922. Available from: <http://doi.wiley.com/10.1111/j.1432-1033.2004.04213.x>.
- [13] Oliveira AP, Nielsen J, Förster J. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiology*. 2005 jun;5(1):39. Available from: <http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-5-39>.
- [14] Karp PD, Weaver D, Latendresse M. How accurate is automated gap filling of metabolic models? *BMC Systems Biology*. 2018 dec;12(1):73. Available from: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-018-0593-7>.
- [15] Feist AM, Palsson BO. The biomass objective function. *Current Opinion in Microbiology*. 2010;13(3):344–349. Available from: <http://www.sciencedirect.com/science/article/pii/S1369527410000512> <http://www.sciencedirect.com/science/article/pii/S1369527410000512/pdf?md5=19c270e4ffc078e9482a7e0cb45019e1&pid=1-s2.0-S1369527410000512-main.pdf>.
- [16] Chan SHJ, Cai J, Wang L, Simons-Senftle MN, Maranas CD. Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*. 2017;33(22):3603–3609. Available from: <https://watermark.silverchair.com/btx453.pdf?token=AQECAHi208BE490oan9kxW{ }Ercy7Dm3ZL{ }9Cf3qfKAc485ysgAAAccwggHDBgkqhkiG9w0BBwagggGOMIIBsAIBADCCAa>
- [17] Lachance JC, Lloyd CJ, Monk JM, Yang L, Sastry AV, Seif Y, et al. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLOS Computational Biology*. 2019 apr;15(4):e1006971. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1006971>.
- [18] Peters AF, Marie D, Scornet D, Kloareg B, Cock JM. Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. *Journal of Phycology*. 2004 dec;40(6):1079–1088. Available from: <http://doi.wiley.com/10.1111/j.1529-8817.2004.04058.x>.
- [19] Cormier A, Avia K, Sterck L, Derrien T, Wucher V, Andres G, et al. Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. *New Phytologist*. 2017;214(1):219–232. Available from: <http://doi.wiley.com/10.1111/nph.14321>.

- [20] Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature*. 2010 jun;465(7298):617–21.
- [21] Prigent S, Collet G, Dittami SM, Delage L, De Corny FE, Dameron O, et al. The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): A resource to study brown algal physiology and beyond. *Plant Journal*. 2014 oct;80(2):367–381. Available from: <http://doi.wiley.com/10.1111/tpj.12627>.
- [22] Aite M, Chevallier M, Frioux C, Trottier C, Got J, Cortés MP, et al. Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Computational Biology*. 2018 may;14(5):e1006146. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1006146>.
- [23] Tapia JE, González B, Goulitquer S, Potin P, Correa JA. Microbiota Influences Morphology and Reproduction of the Brown Alga Ectocarpus sp. *Frontiers in Microbiology*. 2016 feb;7(FEB):197. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26941722><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4765120>.
- [24] Pande S, Kost C. Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends in Microbiology*. 2017 may;25(5):349–361. Available from: <https://www.sciencedirect.com/science/article/pii/S0966842X17300525?via%3Dihub>.
- [25] Zomorodi AR, Islam MM, Maranas CD. d-OptCom: Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities. *ACS Synthetic Biology*. 2014 apr;3(4):247–257. Available from: <http://pubs.acs.org/doi/abs/10.1021/sb4001307>.
- [26] Muller EEL, Faust K, Widder S, Herold M, Martínez Arbas S, Wilmes P. Using metabolic networks to resolve ecological properties of microbiomes. *Current Opinion in Systems Biology*. 2018 dec;8:73–80. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S245231001730197X>.
- [27] Otwell AE, López García de Lomana A, Gibbons SM, Orellana MV, Baliga NS. Systems biology approaches towards predictive microbial ecology. *Environmental Microbiology*. 2018 aug;20(12):4197–4209. Available from: <http://doi.wiley.com/10.1111/1462-2920.14378>.
- [28] Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015;3(1):31. Available from: <https://microbiomejournal.biomedcentral.com/track/pdf/10.1186/s40168-015-0094-5?site=microbiomejournal.biomedcentral.comhttp://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-015-0094-5>.
- [29] Schmidt TSB, Raes J, Bork P. The Human Gut Microbiome: From Association to Modulation. *Cell*. 2018 mar;172(6):1198–1215. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29522742>.
- [30] Leonelli S, Ankeny RA. What makes a model organism? *Endeavour*. 2013 dec;37(4):209–212. Available from: <https://www.sciencedirect.com/science/article/pii/S0160932713000379?via%3Dihub>.
- [31] Xu Y, Zhao F. Single-cell metagenomics: challenges and applications. *Protein & Cell*. 2018 may;9(5):501–510. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29696589http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5960468http://link.springer.com/10.1007/s13238-018-0544-5>.
- [32] Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. 2018 nov;36(10):996–1004. Available from: <http://www.nature.com/articles/nbt.4229>.

- [33] Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature*. 2019 feb;p. 1. Available from: <http://www.nature.com/articles/s41586-019-0965-1>.
- [34] Cavaliere M, Feng S, Soyer OS, Jiménez JI. Cooperation in microbial communities and their biotechnological applications. *Environmental Microbiology*. 2017 aug;19(8):2949–2963. Available from: <http://doi.wiley.com/10.1111/1462-2920.13767>.
- [35] Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. *Nature Medicine*. 2018 apr;24(4):392–400. Available from: <http://www.nature.com/articles/nm.4517>.
- [36] Hassani MA, Duran P, Hacquard S. Microbial interactions within the plant holobiont. *Microbiome*. 2018 03;6(1):58.
- [37] Gauthier JP, Outreman Y, Mieuxet L, Simon JC. Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS one*. 2015;10(3):e0120664. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25807173><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4373712>.
- [38] Blackall LL, Wilson B, van Oppen MJH. Coral—the world’s most diverse symbiotic ecosystem. *Molecular Ecology*. 2015 nov;24(21):5330–5347. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26414414><http://doi.wiley.com/10.1111/mec.13400>.
- [39] Stalidzans E, Seiman A, Peebo K, Komasilovs V, Pentjuss A. Model-based metabolism design: Constraints for kinetic and stoichiometric models. *Biochemical Society Transactions*. 2018;46(2):261–267. Available from: <http://www.biochemsoctrans.org/content/ppbiost/early/2018/02/16/BST20170263.full.pdf><http://www.ncbi.nlm.nih.gov/pubmed/29472367>.
- [40] Mannan AA, Toya Y, Shimizu K, McFadden J, Kierzek AM, Rocco A. Integrating Kinetic Model of E. coli with Genome Scale Metabolic Fluxes Overcomes Its Open System Problem and Reveals Bistability in Central Metabolism. *PLoS one*. 2015;10(10):e0139507. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26469081><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4607504>.
- [41] Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*. 1994;60(10):3724–3731.
- [42] Ebenhöf O, Handorf T, Heinrich R. Structural analysis of expanding metabolic networks. *Genome informatics International Conference on Genome Informatics*. 2004;15(1):35–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15712108>.
- [43] Liebermeister W, Uhlenendorf J, Klipp E. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*. 2010 apr;26(12):1528–1534. Available from: <https://doi.org/10.1093/bioinformatics/btq141>.
- [44] Maranas CD, Zomorrodi AR. *Optimization methods in metabolic networks*. Wiley; 2016.
- [45] Doran PM. *Bioprocess engineering principles*. Elsevier/Academic Press; 2013. Available from: <https://books.google.fr/books?id=wZSy1DhgEXMC&pg=PA664&lpg=PA664&dq=flux+analysis+steady+state+assumption&source=bl&ots=fSemZaon-x&sig=ACfU3U2f208iQOeWdsasYAg00vL2KmS65g&hl=fr&sa=X&ved=2ahUKEwjBsa6vjariAhVJ4eAKHaEQBcYQ6AEwCHOECAkQAQ#v=onepage&q=fluxanaly>.

- [46] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC systems biology*. 2013 aug;7:74. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23927696><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3751080>.
- [47] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature protocols*. 2011 sep;6(9):1290–307. Available from: <http://www.nature.com/articles/nprot.2011.308><http://www.ncbi.nlm.nih.gov/pubmed/21886097><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3319681>.
- [48] Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*. 2019 mar;14(3):639–702.
- [49] Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. *Molecular systems biology*. 2007 mar;3(1):92. Available from: <http://msb.embopress.org/cgi/doi/10.1038/msb4100131>.
- [50] Zomorodi AR, Maranas CD. OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology*. 2012;8(2):e1002363. Available from: <http://dx.doi.org/10.1371/journal.pcbi.1002363><http://www.ploscompbiol.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pcbi.1002363&representation=PDF>.
- [51] Heinken A, Thiele I. Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. *Gut microbes*. 2015;6(2):120–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25901891><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4615372>.
- [52] Song HS, Ramkrishna D. When is the quasi-steady-state approximation admissible in metabolic modeling? When admissible, what models are desirable? *Industrial and Engineering Chemistry Research*. 2009 sep;48(17):7976–7985.
- [53] Reimers AM, Reimers AC. The steady-state assumption in oscillating and growing systems. *Journal of Theoretical Biology*. 2016 oct;406:176–186.
- [54] Mahadevan R, Edwards JS, Doyle FJ, 3rd. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical journal*. 2002;83(3):1331–40. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1302231/pdf/12202358.pdf><http://www.ncbi.nlm.nih.gov/pubmed/12202358><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1302231>.
- [55] Handorf T, Ebenhöf O, Heinrich R. Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *Journal of Molecular Evolution*. 2005;61(4):498–512. Available from: <https://link.springer.com/content/pdf/10.1007%2Fs00239-005-0027-1.pdf>.
- [56] Christian N, May P, Kempa S, Handorf T, Ebenhöf O. An integrative approach towards completing genome-scale metabolic networks. *Molecular BioSystems*. 2009;5(12):1889–1903. Available from: <http://pubs.rsc.org/en/content/articlehtml/2009/mb/b915913b>.
- [57] Schaub T, Thiele S. Metabolic network expansion with answer set programming. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 5649 LNCS. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 312–326. Available from: http://link.springer.com/10.1007/978-3-642-02846-5_27<http://www.springerlink.com/index/9368L534R7V10671.pdf>.

- [58] Prigent S, Frioux C, Dittami SM, Thiele S, Larhlimi A, Collet G, et al. Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. *PLOS Computational Biology*. 2017 jan;13(1):e1005276. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1005276>.
- [59] Ravikrishnan A, Nasre M, Raman K. Enumerating all possible biosynthetic pathways in metabolic networks. *Scientific Reports*. 2018 dec;8(1):9932. Available from: <http://www.nature.com/articles/s41598-018-28007-7>.
- [60] Christian N, Handorf T, Ebenhöf O. Metabolic synergy: increasing biosynthetic capabilities by network cooperation. *Genome informatics International Conference on Genome Informatics*. 2007;18:320–329.
- [61] Ofaim S, Ofek-Lazar M, Sela N, Jinag J, Kashi Y, Minz D, et al. Analysis of Microbial Functions in the Rhizosphere Using a Metabolic-Network Based Framework for Metagenomics Interpretation. *Frontiers in Microbiology*. 2017 aug;8:1606. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01606/full>.
- [62] Opatovsky I, Santos-Garcia D, Ruan Z, Lahav T, Ofaim S, Mouton L, et al. Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment. *BMC Genomics*. 2018 dec;19(1):402. Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-4786-7>.
- [63] Frioux C, Fremy E, Trottier C, Siegel A. Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*. 2018 sep;34(17):i934–i943. Available from: <https://academic.oup.com/bioinformatics/article/34/17/i934/5093211>.
- [64] Ebenhöf O, Handorf T, Kahn D. Evolutionary changes of metabolic networks and their biosynthetic capacities. *Systems biology*. 2006 sep;153(5):354–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16986316>.
- [65] Kruse K, Ebenhöf O. Comparing flux balance analysis to network expansion: producibility, sustainability and the scope of compounds. *Genome Informatics*. 2008;20:91–101. Available from: http://www.worldscientific.com/doi/pdf/10.1142/9781848163003_{_}0008http://www.ncbi.nlm.nih.gov/pubmed/19425125.
- [66] Eng A, Borenstein E. An algorithm for designing minimal microbial communities with desired metabolic capacities. *Bioinformatics*. 2016;32(13):2008–2016. Available from: <http://bioinformatics.oxfordjournals.org/content/32/13/2008.long>.
- [67] Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2012 jan;109(2):594–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22184244http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3258644>.
- [68] Cottret L, Milreu PV, Acuña V, Marchetti-Spaccamela A, Stougie L, Charles H, et al. Graph-Based Analysis of the Metabolic Exchanges between Two Co-Resident Intracellular Symbionts, *Baumanniacicadellincola* and *Sulcia muelleri*, with Their Insect Host, *Homalodisca coagulata*. *PLoS Computational Biology*. 2010 sep;6(9):e1000904. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1000904>.
- [69] Julien-Laferrière A, Bulteau L, Parrot D, Marchetti-Spaccamela A, Stougie L, Vinga S, et al. A Combinatorial Algorithm for Microbial Consortia Synthetic Design. *Scientific Reports*. 2016 jul;6:29182. Available from: <http://www.nature.com/articles/srep29182>.

- [70] Pan S, Reed JL. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Current Opinion in Biotechnology*. 2018 jun;51:103–108. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29278837><http://linkinghub.elsevier.com/retrieve/pii/S0958166917302045>.
- [71] Karp PD, Weaver D, Latendresse M. How accurate is automated gap filling of metabolic models? *BMC Syst Biol*. 2018 06;12(1):73.
- [72] Satish Kumar V, Dasika MS, Maranas CD. Optimization based automated curation of metabolic reconstructions. *BMC bioinformatics*. 2007 jun;8:212. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933441/><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933441/pdf/1471-2105-8-212.pdf><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1933441&tool=pmcentrez&rendertype=abstract>.
- [73] Thiele I, Vlassis N, Fleming RMT. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics (Oxford, England)*. 2014 sep;17:2529–2531.
- [74] Frioux C, Schaub T, Schellhorn S, Siegel A, Wanko P. Hybrid metabolic network completion. *Theory and Practice of Logic Programming*. 2019 jan;19(01):83–108. Available from: https://www.cambridge.org/core/product/identifier/S1471068418000455/type/journal_article.
- [75] Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, et al. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*. 2015 dec;528(7582):364–369. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26633631><http://www.nature.com/doi/10.1038/nature16192>.
- [76] Provasoli L, Pintner IJ. Bacteria induced polymorphism in an axenic laboratory strain of *Ulva lactuca* (Chlorophyceae). *Journal of Phycology*. 1980 jun;16(2):196–201. Available from: <http://doi.wiley.com/10.1111/j.1529-8817.1980.tb03019.x>.
- [77] Johns NI, Blazejewski T, Gomes AL, Wang HH. Principles for designing synthetic microbial communities. *Current Opinion in Microbiology*. 2016;31:146–153. Available from: <http://www.sciencedirect.com/science/article/pii/S136952741630025X>.
- [78] Mas A, Jamshidi S, Lagadeuc Y, Eveillard D, Vandenkoornhuysen P. Beyond the Black Queen Hypothesis. *ISME J*. 2016 09;10(9):2085–2091.
- [79] Henry CS, Bernstein HC, Weisenhorn P, Taylor RC, Lee JY, Zucker J, et al. Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction. *Journal of Cellular Physiology*. 2016 nov;231(11):2339–2345. Available from: <http://doi.wiley.com/10.1002/jcp.25428>.
- [80] Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O’Connor MI, et al. Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*. 2018 apr;p. 1. Available from: <http://www.nature.com/articles/s41559-018-0519-1>.
- [81] Sambamoorthy G, Raman K. Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics*. 2018 sep;34(17):i981–i987. Available from: <https://academic.oup.com/bioinformatics/article/34/17/i981/5093206>.
- [82] Klitgord N, Segrè D. Environments that induce synthetic microbial ecosystems. *PLoS Computational Biology*. 2010;6(11):e1001002.
- [83] Chan SHJ, Simons MN, Maranas CD. SteadyCom: Predicting microbial abundances while ensuring community stability. *PLoS Computational Biology*. 2017 may;13(5):e1005539. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1005539>.

- [84] García-Jiménez B, García JL, Nogales J. FLYCOP: Metabolic modeling-based analysis and engineering microbial communities. In: *Bioinformatics*. vol. 34. Oxford University Press; 2018. p. i954–i963. Available from: <https://academic.oup.com/bioinformatics/article/34/17/i954/5093244>.
- [85] Gebser M, Kaminski R, Kaufmann B, Schaub T. Answer Set Solving in Practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2012;6(3):1–238. Available from: <http://www.morganclaypool.com/doi/abs/10.2200/S00457ED1V01Y201211AIM019>.
- [86] Janhunen T, Kaminski R, Ostrowski M, Schellhorn S, Wanko P, Schaub T. Clingo goes linear constraints over reals and integers. In: *Theory and Practice of Logic Programming*; 2017. p. 872–888. Available from: <https://arxiv.org/pdf/1707.04053.pdf>.
- [87] Gebser M, Kaminski R, Kaufmann B, Schaub T. Clingo = ASP + Control: Preliminary Report. In: *Technical Communications of the Thirtieth International Conference on Logic Programming (ICLP'14)*; 2014. p. 1–9. Available from: <http://arxiv.org/abs/1405.3694>.
- [88] Nègre D, Aite M, Belcour A, Frioux C, Brillet-Gueguen L, Liu X, et al. Genome-Scale Metabolic Networks Shed Light on the Carotenoid Biosynthesis Pathway in the Brown Algae *Saccharina japonica* and *Cladosiphon okamuranus*. *Antioxidants (Basel)*. 2019 Nov;8(11).
- [89] Belcour A, Girard J, Aite M, Delage L, Trottier C, Marteau C, et al. Inferring Biochemical Reactions and Metabolite Structures to Understand Metabolic Pathway Drift. *iScience*. 2020 feb;23(2).
- [90] Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLoS Computational Biology*. 2013 jun;9(6):e1003081. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1003081>.
- [91] Imam S, Schäuble S, Valenzuela J, López García De Lomana A, Carter W, Price ND, et al. A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. *Plant Journal*. 2015;84(6):1239–1256. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26485611><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4715634>.
- [92] Campodonico MA, Vaisman D, Castro JF, Razmilic V, Mercado F, Andrews BA, et al. Acidithiobacillus ferrooxidans’s comprehensive model driven analysis of the electron transfer metabolism and synthetic strain design for biomining applications. *Metabolic Engineering Communications*. 2016;3:84–96. Available from: <http://www.sciencedirect.com/science/article/pii/S2214030116300098>.
- [93] Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, et al. Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*. 2016 sep;17(5):877–890. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5036846/pdf/bbv079.pdf><https://arxiv.org/pdf/1510.03964.pdf><https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbv079>.
- [94] Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*. 2018 jul;36(7):566–569. Available from: <http://www.nature.com/doi/10.1038/nbt.4163>.
- [95] Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Computational Biology*. 2018 oct;14(10):e1006541. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1006541>.
- [96] Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*. 2018 sep;46(15):7542–7553. Available from: <https://academic.oup.com/nar/article/46/15/7542/5042022>.

- [97] Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*. 2019 aug;20(1):158. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1769-1>.
- [98] Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*. 2016 nov;35(1):81–89. Available from: <http://www.nature.com/doifinder/10.1038/nbt.3703>.
- [99] Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*. 2018; Available from: <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky992/5146204>.
- [100] Dittami SM, Barbeyron T, Boyen C, Cambefort J, Collet G, Delage L, et al. Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi", a new candidate genus of Alphaproteobacteria frequently associated with brown algae. *Frontiers in Genetics*. 2014 jul;5:241. Available from: <http://journal.frontiersin.org/article/10.3389/fgene.2014.00241/abstract>.
- [101] Burgunter-Delamare B, KleinJan H, Frioux C, Fremy E, Wagner M, Corre E, et al. Metabolic Complementarity Between a Brown Alga and Associated Cultivable Bacteria Provide Indications of Beneficial Interactions. *Frontiers in Marine Science*. 2020 feb;7:85.