Metabolic modelling approaches for describing and engineering microbial communities

Beatriz García-Jiménez a, b, Jesús Torres-Bacete a, Juan Nogales a, c, *

a Department of Systems Biology, Centro Nacional de Biotecnología (CSIC), 28049 Madrid, Spain
b Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA), Universidad Politécnica de Madrid (UPM), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus de Montegancedo-UPM, 28223-Pozuelo de Alarcón, Madrid, Spain
c Interdisciplinary Platform for Sustainable Plastics towards a Circular Economy-Spanish National Research Council (SusPlast-CSIC), Madrid, Spain

Microbes do not live in isolation but in microbial communities. The relevance of microbial communities is increasing due to growing awareness of their influence on a huge number of environmental, health and industrial processes. Hence, being able to control and engineer the output of both natural and synthetic communities would be of great interest. However, most of the available methods and biotechnological applications involving microorganisms, both in vivo and in silico, have been developed in the context of isolated microbes. In vivo microbial consortia development is extremely difficult and costly because it implies replicating suitable environments in the wet-lab. Computational approaches are thus a good, cost-effective alternative to study microbial communities, mainly via descriptive modelling, but also via engineering modelling. In this review we provide a detailed compilation of examples of engineered microbial communities and a comprehensive, historical revision of available computational metabolic modelling methods to better understand, and rationally engineer wild and synthetic microbial communities.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction .................................................................................................................. 227
2. Learning from nature: functional and stability-based design of synthetic microbial consortia ........................................................................................................... 227
   2.1. Unidirectional non-distributed .............................................................................. 228
   2.2. Multidirectional non-distributed ........................................................................ 228
   2.3. Unidirectional distributed ................................................................................... 228
   2.4. Multidirectional distributed ............................................................................... 231
3. The long journey of community-based modelling: From ecological to genome-scale models ................................................................................................................. 231
   3.1. Ecological models ............................................................................................. 231
   3.2. Individual-based modelling ............................................................................... 231
   3.3. Genome-scale metabolic models (GEMs) ......................................................... 232
4. Dynamics-based classification of descriptive approaches to metabolic modelling ...................................................................................................................... 232
   4.1. Static/Unified methods ...................................................................................... 232
   4.2. Static/Multi-part methods .................................................................................. 233
   4.3. Dynamic methods .............................................................................................. 233
   4.4. Practical applications of descriptive modelling methods .................................. 236
5. Engineering metabolic modelling: Design and optimization .................................... 236
   5.1. Production .......................................................................................................... 237
   5.2. Pathway distribution ......................................................................................... 237
   5.3. Community stability .......................................................................................... 237
1. Introduction

Microbes play a pivotal role in fields as diverse as human health, environmental science and biotechnology. Focus on the latter, microbial production of chemicals has become increasingly attractive across industry due to its role in delivering sustainable manufacturing technology. Microbial biotechnology platforms integrating systems and synthetic biology tools have successfully contributed to delivering a large portfolio of chemical compounds [1–7]. Early applications of systems metabolic engineering to maximize metabolite production focused on engineering single competitive strains, and thus faced hurdles such as metabolic burden and heterogeneity [8,9]. Consequently, production of target chemicals is not always cost-effective and great efforts must be made to improve yield and productivity. The use of microbial consortia has thus been promoted as an alternative to overcome these limitations [10,11] because cooperation among several strains allows microbial communities to function at higher levels of complexity than individual cells. Pathway modularization allows distribution of metabolic reactions among highly specialized strains, thus reducing genetic load requirements per individual. Increased bioproduction performance and efficiency in source transformation can be achieved using different substrates, and/or synthesizing products in parallel, and/or avoiding intermediate metabolite accumulation. In addition, robustness provided by microbial communities avoids environmental stresses [12–14].

The advantages described have underpinned recent progress in analysing, understanding, designing and developing both natural and synthetic microbial communities. Such progress has been applied to improving health, food and chemical production and to dealing with environmental challenges. Therefore, microbial biotechnology will probably lead to microbial community engineering using species selection, manipulation of strain ratios and/or genetic engineering of community members. However, metabolically engineering microbial communities entails certain unresolved challenges [15], e.g. defining co-culture conditions and growth compatibility, and selecting the cross-feeding metabolites among different strains in the consortium. Microbial community modelling emerged with the need to improve the knowledge and understanding of interactions among heterogeneous cells [16]. Such interactions can be described using metabolite exchange modelling, where the actual metabolites and the extent to which they are exchanged need to be defined. Community interactions include cross-feeding, competition for nutrients, symbiotic relationships (such as plant–microorganism) or parasitism (such as human–pathogen) or multi-tissues. Therefore, these pioneering efforts have highlighted the need for novel computational-system approaches in order to facilitate more rational designs. Here we review pioneering achievements in the field of microbial consortia-based bioprocesses and available computational metabolic modelling tools that provide a better understanding and support rational engineering of microbial communities.

2. Learning from nature: functional and stability-based design of synthetic microbial consortia

In nature, microorganisms are involved in a large array of complex interactions with other organisms and their environment, thus contributing to stability and functionality. Among others, such complex relationships traditionally include neutralism, commensalism, amensalism, mutualism, predation and competition. These natural relationships have been profusely used as mechanisms for the establishment of synthetic metabolic interactions when designing synthetic microbial communities (SMC) [13,17,18] (Fig. 1A). Therefore, the two main questions emerging when designing a SMC are: i) how will the microbial community structure be established to ensure the consortium’s stability? and ii) how will relationships be established within the SMC to drive the community’s output?

The stability of the population in a microbial community involves complex interactions among its components. These relationships are usually established through primary interaction mechanisms i.e., interactions that are strictly necessary to deliver the consortium’s output. However, the relationships within a given microbial consortium are not exempt from the emergence of secondary interactions which, although contributing to the community’s stability and/or functionality, are not strictly necessary. For instance, these primary and secondary interactions are easily visible in SMCs formed by *Synechococcus elongatus* and *Pseudomonas putida*. The production of sucrose by the cyanobacterium provides an organic carbon source to the heterotrophic partner and is the primary interaction within the consortium, while the production of O2 derived from *S. elongatus’* photosynthetic activity and the CO2 released by *P. putida* would be secondary interactions, as additional sources of CO2 and O2 exist in the consortium [19,20]. In terms of stability, and considering only primary interactions, microbial relationships can be grouped into two main categories: unidirectional and multidirectional. Unidirectional interactions are those in which the population of one of the consortium’s components is regulated (either positively or negatively) by another component. In contrast, multidirectional interactions are those in which all of the community’s microbial components actively interact with each other to support the stability of the entire consortium via positive or negative feedback (Fig. 1B). Similarly, relationships within the community also determine to a great extent the SMC’s functionality and level of complexity. Overall, functionality as a function of complexity level can be categorized as non-distributed or distributed. Non-distributed functionality implies low complexity levels and often only one member of the consortium is responsible for the final community’s output. On the contrary, more complex SMC outputs require the involvement of several community partners within the consortium. Labour is split in a distributed process, thus increasing the consortium’s efficiency. SMC can therefore be classified into four main categories according to the relationships contributing to stability and
functionality: Unidirectional Non-Distributed, Multidirectional Non-distributed, Unidirectional Distributed and Multidirectional Distributed (Fig. 1B). In order to define the field’s state-of-the-art, in the following sections we have categorized and contextualized current efforts on synthetic microbial consortia engineering based on the above classification. A detailed review of outstanding examples is summarized in Table 1.

2.1. Unidirectional non-distributed

This category includes the simplest community, in which stability is provided by a single unidirectional relationship (e.g., one microbial component is responsible for feeding the other, either directly or by feedstock processing) while the second strain is in charge of the whole consortium’s functionality. Unidirectional Non-Distributed SMC designs provide significant advantages over single cultures by joining complementary metabolic traits of the cognate microbial partners. Therefore, SMC expand the scope of the target bioprocess in terms of either access to new feedstock and/or providing additional biosynthetic properties. In this scenario, a single component of the consortium addresses the catabolism of a complex feedstock (e.g., xylan, cellulose, syrup, etc.) to release low-complexity carbon sources. These easy-to-uptake carbon sources are subsequently used by the other component of the consortium, which is in charge of delivering the non-distributed biotechnological output. Many consortia fitting these criteria have already been constructed, mainly for the production of biofuels. For instance, the fungi Trichoderma reesei was used to hydrolyse cellulose in a co-culture with Escherichia coli, which was in charge of synthesizing isobutanol [21]. Similarly, the co-culture of two clostridium species (C. thermocellum and C. saccharoperbutylicetonium) were used for the production of butanol [18], while a consortium made up of two specialized strains of E. coli was used to produce ethanol from xylan [25]. Within this category, an interesting group of consortia are those in which one of the components is the producer of the primary carbon source. For instance, the cyanobacteria S. elongatus has been profusely used as sucrose producer in co-cultures with heterotrophic organisms such as P. putida and E. coli for the production of polyhydroxyalkanoates and 3-hydroxypropionic acid, respectively [19,20].

2.2. Multidirectional non-distributed

These SMC add an additional layer of complexity. In these consortia, the stability of the community is achieved through the establishment of co-dependency relationships between community microorganisms, while only one consortium component is in charge of the labour. Co-dependency relationships allow better control of the stability of microbial populations. An example of a co-dependency interaction is metabolite cross-feeding, where all members of a consortium are responsible for feeding each other. For example, Sgobba et al [23] developed a multidirectional non-distributed consortium for cadaverine production in which a lysine auxotroph E. coli strain released glucose from starch, feeding C. glutamicum that in turn produced lysine for E. coli. Multidirectional relationships can also be established by competition mechanisms, e.g. in a co-culture of Bacillus subtilis and Streptomyces sp. Mg1, the growth of B. subtilis stimulated the production of chalcomycin A by Streptomyces sp Mg1, which is an inhibitor of B subtilis’ growth [24]. An interesting approach to stabilizing SMC by means of competition mechanisms is the introduction of cheaters strains. For example, when Pseudomonas aeruginosa and Burkholderia cenocepacia were co-cultured under iron-limiting conditions, despite both microorganisms possessing the ability to secrete siderophores, only P. aeruginosa achieved stable growth. When mutant strains of P. aeruginosa unable to secrete siderophores were used instead, both microorganisms grew at the expense of the siderophores secreted by B. cenocepacia [25].

2.3. Unidirectional distributed

This category adds complexity as all members in the SMC contribute to the system’s functionality. Engineering Distributed SMC relies mainly on splitting complex biosynthetic pathways among two or more microbial strains. The division of microbial labour allows resource optimization, thus reducing metabolic burdens and increasing the efficiency of the whole process. In an interesting example E. coli and the yeast S. cerevisiae were used for the production of oxygenated taxanes. Both strains were engineered so that E. coli produced taxadiene, which in turn was used by S. cerevisiae to produce oxygenate taxane [26]. In this SMC, E. coli utilized
Recent examples of engineering Synthetic Microbial consortia.

B. García-Jiménez, Jesús Torres-Bacete and J. Nogales Computational and Structural Biotechnology Journal 19 (2021) 226–246

Table 1

Recent examples of engineering Synthetic Microbial consortia.

| Microorganism                        | Interaction                                                                 | Goal to optimize                                      | C-source          | Yield                                      | Ref. |
|--------------------------------------|-----------------------------------------------------------------------------|--------------------------------------------------------|-------------------|--------------------------------------------|------|
| S. elongatus                         | Escherichia coli                                                           | - Sucrose production in presence of 2,4-DNT            | CO₂               | 1.2 g/L sucrose at 120h.                  | [19] |
|                                      |                                                                            | - 2,4-DNT cleaning                                    |                   | 250 mM, 2,4-DNT cleaning at 24 h.         |      |
|                                      |                                                                            | - PHA production                                      |                   | 5.1 mg/L day PHA                          |      |
| S. elongatus                         | Clostridium thermocellum                                                   | - 3-HP production                                     | CO₂               | Up to 68.29 mg/L, 3-HP at 7 days          | [20] |
|                                      |                                                                            | - Sucrose production                                  |                   | 600 mg/L sucrose at 144 h                 |      |
| Klebsiella pneumoniae                | Shewanella oneidensis                                                      | - Lactate production                                  |                   | 2.1-times increase lactate production     | [139]|
|                                      |                                                                            | - Flavin production                                   |                   | 7.9-time increase flavin production       |      |
|                                      |                                                                            | - Inoculum ratio                                      |                   | Inoculum ratio 1:10                       |      |
|                                      |                                                                            | - Electric power                                       |                   | 19.9 mW/m² power density                 |      |
| Pseudomonas putida                   | B. subtilis holdsyroses sucrose in fructose and glucose, producing propionic acid. They are used by R. eutroph, producing PHA or poly (3-hydroxybutyrate-co-3-hydroxyvalerate) [P(3HB-co-3HV)]. | - Biomass production                                  |                   | - Biomass 3.79 g dcw/L                   | [140]|
|                                      |                                                                            | - PHA production                                      |                   | PHA 63 w/w                               |      |
|                                      |                                                                            | - P(3HB-co-3HV) production                            |                   | P(3HB-co-3HV) 66% w/w                    |      |
| Citrobacter amalonaticus             | Sporomusa ovata                                                            | - Acetate production                                  | CO                | 0.157 mM acetate from 0.439 mM CO         | [141]|
|                                      |                                                                            | - Organic acids production                             |                   |                                           |      |
|                                      |                                                                            | - Corn stover                                         |                   |                                           |      |
|                                      |                                                                            | - E. coli biomass accumulation                        |                   |                                           |      |
|                                      |                                                                            | - E. coli acetate removal                              |                   |                                           |      |
| T. reesei                            | R. orizeae                                                                   | - Butanol production                                  | Rice straw        | 6.5 g/L butanol from 40 g/L rice straw    | [18] |
|                                      | Closstridum thermocellum                                                   | - Xylan hydrolysis                                    | Xylan             | 38.6% hydrolysis                         | [22] |
|                                      |                                                                            | - Ethanol production                                   |                   | 3.71 g/L ethanol                         |      |
|                                      | E. coli                                                                     | - Carotenoids production                               | Corn syrup        | 8.2 mg/L carotenoids                      | [144]|
|                                      |                                                                            | - Production of Lys and cadaverone or L-PA            |                   |                                           |      |
|                                      |                                                                            | - Production of Lys and cadaverone or L-PA            |                   |                                           |      |
|                                      |                                                                            | - Production of Lys and cadaverone or L-PA            |                   |                                           |      |
| E. coli                              | Corynebacterium glutamicum                                                 | - Cellulose                                           |                   |                                           |      |
|                                      |                                                                            | - Production of Lys and cadaverone or L-PA            |                   |                                           |      |
|                                      |                                                                            | - Production of Lys and cadaverone or L-PA            |                   |                                           |      |
| Sacharmyces cerevisiae, Bacillus     | Amyloliquefaciens/L fermentum produces amylase, hydrolysing starch into glucose and oligosaccharides. They are used by S. cerevisiae. Its growth stimulates the production of more amylase for B. amyloliquefaciens/L fermentum. | - α-amylase production                                 |                   | - 1.8-times increase α-amylase production | [145]|
|                                      | Amyloliquefaciens/L fermentum                                              | - Co-culture conditions                               |                   | - 1.8-times increase α-amylase production | [145]|
| Streptomyces sp. Mg18. subtilis       | In co-culture at limited iron P. putida and Streptomyces sp Mg1 to produce chalconycin A (macrolide antibiotic). Chalconycin A inhibits B subtilis growth. | - Chalcomycin A                                       | Maltose           | - n.d.                                    | [24] |
|                                        |                                                                            | - Chalcomycin A                                       |                   |                                           |      |
|                                        |                                                                            | - Chalcomycin A                                       |                   |                                           |      |
|                                        |                                                                            | - Chalcomycin A                                       |                   |                                           |      |
| P. putida - Bdellovibrio bacteriovorus|                                                                               | - PHA and PHB                                          |octanoate          | 80 % recovery in the extracellular medium | [146]|
|                                        |                                                                               | - Population fitness                                   |                   |                                           |      |
|                                        |                                                                               | - Casamino acids                                      |                   |                                           |      |
|                                        |                                                                               | - Increase in the growth of B. cereoparica            |                   |                                           |      |
|                                        |                                                                               | - Electric density                                     |                   |                                           |      |
|                                        |                                                                               | - Glucose                                             |                   | - 14-times increase of the electric density | [147]|

(continued on next page)
Table 1 (continued)

| Microorganism | Interaction | Goal to optimize | C-source | Yield | Ref. |
|---------------|-------------|------------------|----------|-------|------|
| **Unidirectional Distributed** | | | | | |
| E. coli | Hydrogel compartmentalized E. coli and S. cerevisiae were co-cultured, using glucose as C-source. E. coli produces L-DOPA, that is used by S. cerevisiae to produce betaxanthins | - Stability of the compartmentalized consortium | Glucose | - Up to 10 times reutilization of the compartmentalized consortium | [148] |
| E. coli, S. cerevisiae | The rosmarinic acid biosynthetic pathway was divided in three E. coli strains, one producing caffeic acid, other salvinic acid, and a third strains that use those intermediaries to produce rosmarinic acid. All of them use glucose as carbon source. | - Rosmarinic acid | Glucose | - 172 mg/L rosmarinic acid | [30] |
| E. coli | The glutarate biosynthetic pathway from Lys was split in two E. coli strains. The first one use Lys, producing 5-aminovaleric acid, that is used by the second E. coli strain producing glutarate. | - Glutarate production | Lysine | - 43.8 g/L glutarate | [149] |
| E. coli | E. coli RES produces resveratrol from p-coumarate. The resveratrol is glycosylated by E. coli RGL. Both strains use glucose as carbon source. | - Resveratrol glucosides | Glucose | - 92 mg/L resveratrol glucosides | [28] |
| Halomonas sp. HL-48 | When both strains are growing using glucose as carbon source they compete for it. When xylose is used instead of glucose, Halomonas consumes xylose, producing metabolites that are used for Marinobacter growth. | - Growth | Xylose | - Changed from competitive to cooperative interaction the growth was improved in co-culture | [50] |
| Marinobacter sp. HL-58 | The enzymatic pathway to produce ethanol from cellulose was divided in four strains. The first produces phenylpropanoic acid, that is used for the second, producing flavonones. A third strain produces flavan-3-ols from flavonones. Finally, the last E. coli strain produces anthocyanins from flavan-3-ols. | - Inoculation ratio | Glucose | - 9 mg/L anthocyanin-3-O-glucosides | [31] |
| E. coli | The resveratrol biosynthetic pathway is divided in two E. coli strains. Both strains use glycerol as carbon source. One of them produces p-coumarate, which is used for the other to produce resveratrol. | - Resveratrol production | Glycerol | - 22.6 mg/L resveratrol in 30 hours | [29] |
| E. coli | E. coli utilizes xylose as C-source, producing acetate which is the C-source for S. cerevisiae. In parallel, E. coli is producing taxadiene, that is oxygenated by S. cerevisiae. | - Co-culture stability | Xylose | - 33 mg/L oxygenated taxanes | [26] |
| E. coli | One E. coli strain uses xylose as C-source, producing 3-dehydroshikimic acid (DHS), uses for the other strain to produce muconic acid or 4-hydroxybenzoic acid, using glucose as C-source. | - Muconic acid | Glucose | - 4.7 g/L of muconic acid | [151] |
| E. coli | The enzymatic pathway to produce ethanol from cellulose was divided in four S. cerevisiae strains. | - Ethanol production | Cellulose | - 1.25 g/L of ethanol | [17] |
| Four strains of S. cerevisiae | | | | | |
| **Multidirectional Distributed** | | | | | |
| Dietzia sp strain DQ1245-1b Pseudomonas stutzeri SLG510A3-8 | Dietzia uses hexadacene as C-source, producing hexadecanoid acid, α-ketoglutaric acid and R-3-hydroxybutanoic acid, that are used by P. stutzeri, that in turn produces glutamate and acetate for Dietzia. The consortium increase the diesel degradation | - Diesel biodegradation | Hexadacene | - 85.54 % diesel removal | [152] |
| E. coli | One E. coli strain uses xylose, producing tyrosol. The other consumes glucose and produces salidroside (from tyrosol). The relationship between both strains had been established by cross-feeding. The xylose consuming strain is Phe auxotroph, while the glucose consuming is Tyr auxotroph. | - Salidroside production | Glucose | - 6.03 g/L at 120 h fermentation | [32] |
| E. coli | One E. coli strain uses glucose as C-source, producing lysine. The other E. coli strain intakes the lysine producing cadaverine. This strain uses glycerol as carbon source | - Cadaverine production | Glycerol | - Up to 28.5 g/L with constant feeding at 40 h | [153] |
| E. coli | E. coli utilizes glucose as C-source, producing lactate and an electron donor. B. subtilis uses also glucose producing riboflavin as an electron shuttle. S. oneidensis uses the electron donor and the shuttle | - Electricity production | Glucose | - 15 days production with an efficiency of 55.7% | [154] |
Table 1 (continued)

| Microorganism | Interaction                                                                                                                                                                                                 | Goal to optimize | C-source | Yield                                                                 | Ref. |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|----------|-----------------------------------------------------------------------|------|
| ***E. coli*** | *E. coli* ferments glucose producing formate, which is used by *S. oneidensis*, producing flavins, uses by *E. coli*. Their activity increase the electric current from cathode to anode in a MFC                                                                 | - Current density | Glucose  | - Increase of the current density to 2.0 μA/cm².                       | [155]|
| ***E. coli*** | *E. coli* L is Leu auxotroph and *E. coli* K is Lys. They co-culture provide each other with the necessary amino acids, increasing the growth rate and the biomass.                                                                                                           | - Growth         | Glucose  | - 3-fold growth rate increase                                          | [156]|

2.4. Multidirectional distributed

In this category, all members of the SMC have a role in maintaining both the stability of the system and its functionality. Multidirectional Distributed SMC are significantly more complex and require substantial metabolic engineering of the involved strains to ensure stability and functionality. One example is a consortium comprising two *E. coli* strains for the production of salidroside [32]. Both strains were engineered to establish cooperative metabolite cross-feeding so that each strain complemented the other’s auxotrophy. One of the strains produced tyrosol, which was used by the other to synthesize salidroside. To avoid competition over the carbon source, both strains were also engineered so that the tyrosol producer used xylose and the salidroside producer used glucose.

Overall, it becomes apparent that increasing the complexity of SMC in terms of stability and functionality and using complex feedstocks delivers cost-effective production of increasingly complex metabolites. However, upscaling SMC design from simple Unidirectional Non-Distributed to Multidirectional Distributed is not always straightforward, but is instead a trial and error process. Thus, more holistic approaches to microbial community engineering are needed. In this sense, biotechnological applications using monocultures benefit from simpler construction and application of metabolic models.

Given the clear advantages that modelling contributes to the development of more precise and complex SMC, in the following sections we provide a comprehensive review of classical and recent computational modelling methods developed to describe and engineer natural and synthetic microbial consortia.

3. The long journey of community-based modelling: From ecological to genome-scale models

Multiple community-level modelling approaches have been developed to gain insights in the understanding of complex ecosystems [33,34]. Overall, they have been classified as ecological, individual-based and metabolic models [35,36]. Ecological models describe communities using ecological parameters such as pairwise interactions, growth rates, etc. and interactions depend mainly on correlations. Individual-based models focus on the individual rather than the population level. Finally, just as ecological models predict interactions between components of a given system, metabolic models predict interactions in a metabolic context while providing a community dynamics description. The lack of knowledge of kinetic parameters has promoted wide use of stoichiometric, constraint-based rather than kinetic models [37].

3.1. Ecological models

Ecological models focus on representing and/or discovering potential interactions among different species [38,39]. They include mainly the evolutionary game theory and non-linear dynamics where evolution is driven by stochastic processes [40]. Evolutionary game theory emerges as an adaptation of the classical game theory to biological systems after stating that the assumption that the success of one individual depends on the choices of others does not apply in biology [41]. Thus, in evolutionary game theory, natural selection and mutation are what drive change in biological communities. This theory has been used to explain the behaviour of microbial communities in terms of interactions such as cooperation [42,43] or competition [44]. A new application of game theory combined with metabolic models has been recently suggested in community modelling to infer evolutionary stable interactions by analysing the behaviour of a pair of microbes with complementary autotrophies and cross-feeding relationships [45].

Lotka-Volterra (LV) are the first non-linear dynamic systems describing biological populations with a mathematical model based on ordinary or partial differential equations. LVs are deterministic and do not consider the randomness present in nature. From a static point of view, they are used to model similarity- or regression- or rule-based networks [46]; while from a dynamic point of view, generalized Lotka-Volterra (gLV) is the main approach [47,48]. gLV requires knowledge of the growth rates and the strength of the interactions between different components of the community. gLV equations have been used to model a variety of different microbial communities including cheese fermentation communities [49], marine phage communities [50] and the human microbiome [51–53]. In this last case, gLV was extended to take into account external perturbations over time [51], and successfully applied to predicting species abundances in the community [52]. Finally, it was possible to qualitatively infer interaction types without a dynamic model, quantitative assessment of interaction strengths and growth rates when gLV was considered [53].

3.2. Individual-based modelling.

In individual-based modelling (IBM), otherwise known as agent-based modelling [54,55] microbes are individually simu-
lated as concentration state variables rather than at population-level. Each cell evolves over time following predefined probabilistic rules that introduce the randomness required to model dynamics. This approach includes genes, transcripts, proteins and metabolites, although usually just a representative subset of them is selected to reduce the complexity. Its main advantage is that they take intra-population heterogeneity into account. For example, IBM has been used to model biofilms with a microbial consortia dynamics simulator set up on a high performance computing platform [56]. A detailed review of IBM models is included in [57].

3.3 Genome-scale metabolic models (GEMs)

GEMs are structured representations of a target organism based on existing genetic, biochemical and physiological information. Therefore, GEMs represent the metabolic capabilities of a particular organism and can be used in combination with algorithms such as Flux Balance Analysis (FBA) to predict phenotype from genotype [58–60]. Before applying FBA, the metabolic network of a given organism is converted into a numerical stoichiometric matrix (S), where rows describe individual metabolites and columns describe reactions. Cells’ subscripts \( i < j \) refer to their row (i) and column (j). To predict growth (Z) and the vector \( v \) of unsolved individual flux values for each metabolic reaction, a linear programming optimization problem can be solved by maximizing a given objective function \( Z = s^T \cdot v \), subject to a set of constraints (Eq. (1)). FBA assumes the steady-state with mass-balance constraint \( (S \cdot v) = 0 \), where metabolites’ change in concentration as a function of time equals zero. Flux is constrained for each reaction by defining its lower (\( lb \)) and upper (\( ub \)) bounds in mmol/g of dry weight * h) units. \( Z \) is usually set to maximize biomass (g/L), although any metabolic reaction could be selected as a target for flux optimization. A good example is setting the bounds for a particular metabolite of interest’s exchange reaction to maximize its production rate.

\[
S \cdot v = 0 \\
\text{lb}_i \leq v_i \leq \text{ub}_i 
\]  

(1)

An important advantage of metabolic models is that they provide accuracy without requiring kinetic information [60]. Therefore, it is not surprising that this modelling approach has started to be used successfully irrespective of the level of complexity, i.e. from individual organisms right up to microbial communities. Therefore, GEMs are seen by many as optimal computational tools for optimizing SMC-based biotechnological endeavours. The GEM design procedure includes a posteriori experimental validation of the model, including a performance assessment using different carbon sources, gene essentiality and flux prediction in known conditions. Flux prediction could be validated with experimental techniques that retrieve in-vivo fluxomics data, i.e. 13C metabolic flux analysis (MFA), based on stable isotope tracing studies [61]. For instance, 13C-MFA was extended to measure metabolic fluxes at the microbial community level [62] and further improved by avoiding cell separation. This allows quantification of microbe-specific fluxes and metabolite cross-feeding rates and has been applied successfully with E. coli biofilms [63].

Community models based on GEMs have been used at microbial scale, but also to create multi-tissue models such as the human liver and to predict the effect of kinetically-modelled drugs [64]. Interestingly, this approach has also been applied to the construction of whole-plant models including the leaves, the stem, the seeds and the roots of barley [65]. The combination of several genome-scale models in a community, however, entails several challenges [66]. For instance, defining a community’s target function is a tricky point both in biological and mathematical terms. On the other hand, setting exchange rates and metabolic fluxes to constrain the models in the context of a community is a new topic. Current techniques measure the in vivo data for individual organisms and are thus not applicable to the community. Instead, the individual contribution of each organism must be measured. An additional challenge is to define the composition of the medium for combined culture taking into account the exchange of metabolites within the SMC.

Due to the increasing interest and applicability of GEMs, we focus here on the different approaches available so far to model microbial communities using these metabolic models. Two different stages can be distinguished in microbial community modelling with GEMs: The first stage includes descriptive methods to understand and describe communities. The second level of development, which took off recently, focuses on building methods to describe and engineer these microbial communities.

4. Dynamics-based classification of descriptive approaches to metabolic modelling

Descriptive methods for metabolic modelling of microbial communities are useful to describe how consortia work, to understand them and to identify relationships within the community. However, they cannot be used to engineer consortia for reasons explained below. Several attempts have been made to classify and categorize different approaches to modelling microbial communities. The first classification was based on available knowledge about the community and its complexity [67,68]. An alternative categorization was based on the scope of the community, as defined by Bosi et al [69]. A most recent classification is based on the definition of the target function (simplified linear, multi-level or non-linear function) [70]. In order to complement current approaches, we propose a classification based on microbial community dynamics (see Table 2). Thus, we classified available descriptive methods as static/unified, static/multi-part and dynamic.

4.1 Static/Unified methods

This approach considers all strains unified in a common metabolic model, with only one copy of the shared reactions and metabolites. The model is completed by adding strain-specific metabolic content and a combined community-based biomass target function. This approach, also called ‘lumped network’ or ‘enzyme soup’, is the simplest and, although only useful to have a general perspective of how the community works, it allows high scalability (Table 3). In network-based models, the unified approach would be the closest as it considers all reactions in a single graph, irrespective of stoichiometry, which is ignored in favour of topology. Network-based models consider the metabolic reactions of each of the strains in the community to plot a graph where metabolites are represented by nodes connected to each other following the direction of metabolic reactions, i.e. from substrates to products. Reactions are in turn represented by edges. This approach could be applied to poor-quality GEM reconstructions because the main source of data is the reaction sequence. Some tools or algorithms that follow this unified approach are:

- Borenstein’s group uses a graph or network-based community model representation that does not consider stoichiometry. With this unified static approach, they mainly study relationships among different microbes [71].
- Kbase is a community data-driven network reconstruction [72]. It builds a single community model rather than aggregating individual models and is focused on predicting interactions between species in a community. In the absence of data for cer-
tain species due to lack of individual cultivability, this approach uses relevant community-level data as input. Single and community modelling is carried out using the KBase software platform, including automatic gap-filling analysis by providing a particular community-based growth condition (www.kbase.us) [73].

- MO-FBA and MO-FVA, multi-objective FBA and FVA (Flux Variability Analysis) algorithm extensions to community level [74]. These methods model microbial consortia by grouping several constraint-based individual models in a large, combined stoichiometric matrix. The multi-objective feature allows weighted combination of each strain's individual objective.

4.2. Static/Multi-part methods

This category of models preserves the individual metabolic matrices and introduces a pool of metabolites, which could be defined by pre-fixed reactions (guild compartment in other classifications) or by new stoichiometric reactions after an initial optimization step (bi-level optimization). Single strain models are directly connected by exchange reactions, assuming no change in the concentrations of extracellular metabolites and no accumulation in the medium. This approach has been profusely applied to describe microbial communities (Table 3). Several algorithms that fit this category are summarized below:

- The method described by Stolyar et al [75] provided the first metabolic model of a microbial consortium and the distribution of its associated metabolic fluxes. This method has since been used in applications pursuing different objectives, such as categorizing interactions [76], estimating medium composition [77], predicting relative biomass abundances [78] and defining a host-pathogen interaction between the human alveolar macrophage and M. tuberculosis in a multi-tissue model [79].

- OptCom [80] and d-OptCom [81]: are two closely related methods focusing mainly on engineering microbial communities (see a longer description in section 5), although descriptive versions are also available.

- cFBA [78]: this method assumes a balanced and fixed growth rate for all microbes in the consortium. Subsequently, cFBA (community FBA) maximizes this growth rate using a non-linear multi-objective function. This approach implies constant species abundance ratios in the community and is applicable to cells grown in chemostat or in waste-water microbial community scenarios [66].

- Mminte [82]: this method supports the assessment of the pairwise microbial metabolic interactions that occur in a community model limited to two strains. Metabolic models are automatically reconstructed using ModelSEED and metagenomics data as input (16S rRNA sequences).

- SteadyCom [83]: this system maximizes community stability, i.e., constant growth rate across all microbes in the community, with an iterative linear programming approach. Additionally, it applies FVA to predict microbial abundances under changing uptake rates.

- CarveMe [84]: this focuses mainly on automatic reconstruction of single strains. In addition, it allows to automatically merge several single-species models into a single community model with a common or individual extracellular compartments.

- Microbiome modelling toolbox [85]: a COBRA Toolbox extension to analyze microbial communities and study interactions (intra- or with the host).

- MIMICOM [86]: this static approach predicts growth rates and fluxes from in vivo data such as species abundances in a microbiome sample. Consequently, it infers metabolic interactions among the microbiota.

4.3. Dynamic methods

Static approaches ignore temporal events. The dynamic or hybrid approaches are based on dynamic Flux Balance Analysis (dFBA) [87], which allows representations of the community's temporal evolution, including variations of metabolite concentrations and cell densities over time. This is the preferred approach to simulate microbial interactions because shared metabolites vary dynamically. It is however limited by the fact that dynamic approaches entail kinetic parameter configuration and require higher computational resources, running FBA multiple time points per strain and thus limiting the analysis to smaller sized communities.

- DMMM [88]: this was the first method that used dFBA at community level. DMMM optimizes growth rates for each strain.

- COMETS [89]: in addition to dynamic simulation of communities using dFBA, this algorithm considers the cells' spatial distribution. The biomasses and fluxes per time points reported as output can be visualized using the VisANT tool [90].

- MCM [91]: this framework simulates dynamic community models and adds statistical evaluation and parameter calibration based on experimental data. It was initially tested with a
Table 3

Applications descriptive microbial community modelling approaches. There are three blocks corresponding to the descriptive modelling approach category described in Table 2. The 'tool' column includes the name of the algorithm or method defined in that application to describe the communities, and link to the software if it is available. 'In vivo validation' column indicates if the application has been validated with in vivo data or they are in silico-based results.

| Modelled Species | Application                                                                 | In vivo validation | Tool          | Ref. |
|------------------|-----------------------------------------------------------------------------|--------------------|---------------|------|
| Static/unified   | Several anaerobic fermentative strains                                        | No                 | ad hoc        | [157]|
|                  | - 478 species                                                               | No                 | ad hoc        | [158,159]|
|                  |   - 154 human microbiome species                                            | No                 | ad hoc        | [157]|
|                  | Assorted 113 bacterial species                                              | No                 | ad hoc        | [160]|
|                  | Synechococcus spp, Chloroflexus spp, and sulfate reducing bacteria           | No                 | ad hoc        | [67] |
|                  | Clostridium cellulolyticum, Desulfovibrio vulgaris                          | Yes                | ad hoc        | [161]|
|                  | Hildenborough, and Geobacter sulfurreducens                                 | No                 | ad hoc        | [163]|
|                  | 2 naphthalene-contaminated soil communities; with 13 and 12 species, including: Achromobacter, Azospirillum, Comamonas, Chromobacter and Pseudoxanthomonas | Yes               | ad hoc        | [162]|
|                  | 261 assorted species of diverse habitats, such as soil, water and the human gut | No                 | ad hoc        | [164]|
|                  | Microbialites and microbial mats (structures similar to corals and stromatolites) | No                 | ad hoc        | [165]|
|                  | Thermosynechococcus elongatus BP-1 and Microthermus ruber                    | No                 | KBase         | [72] |
|                  | Strain A                                                                    | No                 | MO-FBA/       | [74] |
|                  | The same as Taffs et al., 2009 [67] (see above)                             | No                 | FVA           |      |
| Static/multi-part | D. vulgaris and Methanococcus maripaludis                                   | Yes                | ad hoc        | [75] |
|                  | Clostridium butyricum and Methanosarcina mazei                              | No                 | ad hoc        | [166]|
|                  | - Hepatocyte (liver), adipocyte (fat) and myocyte (skeletal muscle) human cells | No                 | ad hoc        | [79,167]|
|                  |   - leaf, stem and root of Arabidopsis thaliana cells                       | No                 | ad hoc        | [168]|
|                  |   Plasmidium falciparum and the host red blood cell (erythrocyte)            | No                 | ad hoc        | [77] |
|                  | E. coli, Bacillus subtilis, Helicobacter pylori, Salmonella typhimurium, Methanosarcina Barkeri, S. oneidensis and Methylobacterium extorquens | No                 | MO-FBA/       | [74] |
|                  | 46 pairs of auxotroph E. coli                                              | Yes                | ad hoc        | [169]|
|                  | The same as Stolyar,2007, Taffs et al., 2009 and Miller et al.,2010 [67,75,161] (see above) | No                 | OptCom        | [80] |
|                  | - Two imaginary species 'x' consuming glucose and ammonium and producing succinate and species 'y' consuming succinate, fixing nitrogen gas and excreting ammonium. | No                 | cFBA          | [78] |
|                  |   - E. coli polymorphism in Long Term Experimental Evolution experiment [170] | No                 | ad hoc        | [78] |
|                  | Geobacter metallireducens and G. sulfurreducens                             | No                 | ad hoc        | [171]|
|                  | Bacteroides thetaiotaomicron, Eubacterium rectale and Methanobrevibacter smithii | No                 | ad hoc        | [100]|
|                  | Bifidobacterium adolescents L2-32 and Faecalibacterium prausnitzii A2-165   | No                 | ad hoc        | [172]|
|                  | Ketogulonicigenium vulgar and Bacillus megaterium                           | No                 | ad hoc        | [173]|
|                  | 11 representative gut microbes (E. coli, H. pylori, Salmonella enterica, S. thermophilus, etc.) | No                 | ad hoc        | [103]|
|                  | Leptospirillum ferrphilum and Ferroplasma acidophilum                      | No                 | ad hoc        | [174]|
|                  | AOR, ammonia oxidizing bacteria: Nitrosomonas europaea,                     | Yes                | ad hoc        | [175]|
|                  | Study of interspecies electron transfer mechanisms in syntrophic associations, in genomic and transcriptomics. | Yes                | ad hoc        | [171]|
|                  | Prediction of interactions between 3 key representative bacteria in the human gut, and analysing their individual contributions to secrete SCFA. | Yes                | ad hoc        | [100]|
|                  | Predicting demand for acetate and production of butyrate, in 2 gut strains related to Chron's disease, using OptCom tool. | No                 | ad hoc        | [172]|
|                  | Understanding of vitamin C production by an artificial consortium, study of subsystems and other possible metabolites to secrete. | No                 | ad hoc        | [173]|
|                  | Study of interactions between gut microbes and human small intestinal enterocytes, under anoxic and normoxic conditions. | No                 | ad hoc        | [103]|
|                  | Study of bioleaching (oxidizing iron) in a bacteria-archaea consortium presents in natural environment, with chemo-mixotrophic growth. | No                 | ad hoc        | [174]|
|                  | Assessment of NO redox reactions contributes to N2O | Yes                | ad hoc        | [175]|

234
| Modelled Species | Application | In vivo validation | Tool | Ref. |
|------------------|-------------|--------------------|------|------|
| *N. eutropha*, *N. multiformis*, and *N. oceani*. And *NOR*, nitrite oxidizing bacteria: *Candidatus Nitrospira defluvi*, *Nitrobrobacter winogradskyi*, *Nitrobrobacter hamburgensis*, *Nitrospina gracilis.* | Exploring pairwise microbial metabolic interactions, using 16S data from microbiome studies. Evaluating a sulphate-reducing bacteria growth in gut microbiome with different diets with data from [176]. | No | MMInte | [82] |
| - Human Microbiome Project data | | | | |
| - Desulfovibrio piger, *B. thetaiotamicron*, Bacteroides caccae. Bacteroides ovatus, *E. rectale*, Marvinbryantia formateoxigenes, Collinsella aerofaciens, E. coli and Clostridium symbiosum | Maximizing community stability (common growth). Automatic reconstruction of single strain models (from 238 to 2472 reactions per model) with the possibility to merge in a community one, analysing the number of compounds that can be exchanged. | No | SteadyCom | [83] |
| - 4 E. coli auxotrophic for amino acids- Gut microbiome | Analysis of pairwise interactions (microbe-microbe) of different types (competition, parasitism, etc.) with a join matrix of GEMs, and modelling of microbial communities given the relative abundances, used to personalize community biomass reaction and simulating under different diets. | No | Microbiome modelling toolbox | [85] |
| Human gut strains from AGORA collection [102] and human cells (Brunk et al., 2018) [177] | Predicting growth rates and metabolic fluxes from microbe abundances as input. Using an heuristic optimization approach based on L2 regularization to allow different growth rates per strain. | No | MCM | [91] |
| Human gut strains from AGORA collection | | | | |
| Dynamic E. coli | Exploring the metabolic variability among bacterial strains and identifying interactions, across different single-carbon-source conditions. They use a combination of a graph-theoretic approach together with a metabolic model. | No | ad hoc | [178] |
| Clostridium acetobutylicum and Clostridium cellulolyticum | Improving bioprocessing of cellulose with a clostridial consortia, with DMMM. | No | ad hoc | [179] |
| G. sulfurreducens and Rhodotherax ferrireducens | Designing of uranium bioremediation scenarios with two competing heterogeneous species | No | DMMM [88,180] | |
| - E. coli auxotrophs | Study of impact of lactate vs acetate addition on the composition of uranium-reducing community. In-vivo validation of *E. coli* auxotrophs with Wintermute and Silver, 2010 [181] results. | Yes | d-OptCom | [81] |
| - *G. sulfurreducens*, *R. ferrireducens*, and *S. oneidensis* | Simulation of spatiotemporal dynamics of microbial communities, predicting species ratios and investigating the influence of spatial structure on competition in mutualistic systems, and with a competitor between the cross-feeding pair. | Yes | COMETS [89] | |
| E. coli, *S. enterica* and *Methylbacterium extorquens* | | | | |
| Homogeneous E.coli consortia | Combining metabolic model with statistical analysis and calibration to experimental data, in this case related to Lenski’s experiment LTEE. | Yes | MCM | [91] |
| - E. coli and *S. enterica* [87] | Visualization of metabolic interaction networks between microbes in a community. | No | VisANT | [90] |
| - *B. fragilis*, *B. longum*, *C. difficile*, *E. coli*, *H. pylori* and *L. acidophilus* | Analysis of evolution. LTEE: divergence in glucose-limited conditions, with daily transfers. | No | evoFBA | [182] |
| E. coli (E. coli B, not K12) | Analysis of interactions and spatial and temporal distributions of microbes in communities using individual-based metabolic modelling. | No | BacArena | [94] |
| - Clostridium beijerinckii and *M.arkeri* | Study of cross-feeding with short-chain fatty acids from glucose in the human gut microbiome, using DMMM with spatial addition. The L. plantarum GEM is converted in a ‘supra-model’ increased by pathways crucial in carbohydrate fermentation in the colon. | No | ad hoc | [183] |
| - Anaerostipes caccae, *B. thetaiotamicron*, *Bifidobacterium longum*, *Blautia producta*, *Clostridium ramosum*, *E. coli* and *Lactobacillus plantarum* | Study of the dynamics of nitrification-derived N oxide production, with aerobic ammonia- and nitrite-oxidizing bacteria, using DMMM. | Yes | ad hoc | [184] |
| L. plantarum | Combining diauxic shift in two homogeneous subpopulations, combining ordinary differential equations (ODE) with GEMs. | No | Daphne | [95] |
| N.s europaea and *N. winogradskyi* | Simulation of heterogeneous microbial communities behaviour over time with ODE and GEMs under perturbations, i.e. changes in availability of metabolites and biomass of different strains. | No | MMODES [96] | |
| E.coli | | | | |
| - *F. prausnitzii* and *B. adolescentis* | | | | |
| - *P. aurecens*, *H. stevensi*, *Halobacillus sp.* | | | | |
homogeneous *E. coli* community and subsequently with species assemblages in nitrifying and methanogenic bioreactors [92,93].

- BacArena [94]: this combines metabolic modelling with individual-based modelling instead of using population-based modelling (one model per strain with a certain amount of biomass). Therefore, BacArena supports modelling of metabolically heterogeneous populations where each individual cell is represented by a unique metabolic model depending on its spatial resource allocation. In BacArena, metabolite diffusion (implemented with partial differential equations) produces gradient concentrations resulting in spatial niches where different metabolic pathways are activated. It is also able to predict novel cross-feeding interactions through fermentation products. COMETS and BacArena allow spatial resolution taking diffusion parameters into account.

- Daphne [95]: Daphne combines two different modelling strategies: GEM (metabolism) and ODE (Ordinary Differential Equations). ODE supports modelling the strain’s growth kinetics and the medium metabolite consumption and production dynamically. It is underpinned by a set of equations that can be solved mathematically.

- MMODES [96]: this also integrates GEMs and ODEs to simulate biomass and metabolite dynamics over time. In addition, it is possible to add perturbations using external longitudinal interventions such as changes in the medium and/or strain ratio (e.g. increasing a metabolite concentration and/or biomass of a species).

### 4.4. Practical applications of descriptive modelling methods

Overall, each of the three categories of descriptive methods described above are suitable for modelling different microbial scenarios and their interactions. The unified approach is appropriate for multiple strain systems and/or where knowledge is limited, e.g. unknown details of the individual assignment of reactions and metabolites, such as in metagenomics. Quantifying metabolic fluxes and representing inter–species interactions requires more complex approaches, such as the multi-part or dynamic models. However, the dynamic approach is the only suitable one to model medium composition and predict metabolite concentration because the multi-part feature transfers metabolites from one model strain to another. Therefore, the dynamic approach is better able to represent complex situations in microbial communities. Dynamic approaches consider all time-dependent elements, although they only can be applied to small communities because they require describing individual strains in great detail and are more time consuming than the other approaches. Nevertheless, they are recommended for engineered or synthetic microbial communities used in biotechnology applications, i.e. scenarios where species richness is generally low. Computing requirements tend to increase from Static/Unified to Static/Multi-part to Dynamic approaches. This is particularly true for Dynamic approaches because dFBA requires solving each GEM for every time slot in the time-series. Final requirements will depend on the size of the strain models (number of reactions and number of metabolites) and any additional factors, such as the length of the time-series.

Static/Unified, Static/Multi-part and Dynamic approaches have been used to model a large number of microbial communities (see ‘modelled species’ column in Table 3) in a range of scenarios; such as food biotechnology [97], human health (including GEMs for microbes, tissues and organs) [98] and marine microbiome [99].

Table 3 collates applications to microbial consortia from a descriptive point of view. Applications have been grouped according to the categories defined in Table 2. The most extended approach is the static/multi-part one, which has around twice the number of applications of static/unified and dynamic approaches. In some cases, the study defines a new computational method, while in others the methods are re-used. Sometimes the corresponding in vivo consortium has been deployed, although this is generally not the case. Microbial consortia applications are sorted according to: i) species richness (some are monoclonal populations, others are consortia comprising less than 10 strains or hundreds of heterogeneous strains, such as those present in the gut microbiota), ii) species diversity (ranging from only one cell per strain to large consortia with hundreds of cells per organism) and iii) environment, industrial bio-transformations, human health and plants.

The application of these methods to human gut microbiome modelling has received special attention [70]. Despite improvements in sequencing having broadened our knowledge of the components of the gut microbiota, the relationships among them and between them and the human cells remain mostly unknown. Hence, microbial community modelling techniques contribute to improving our understanding of the complex behaviour of the gut microbiome and its associations with human diseases. In many cases, only a small (less than 10) and simplified subset of representative species from the microbiome have been taken into account [100,101]. However, in recent studies the size of the modelled microorganisms has been expanded to tens or even hundreds of species. Static/multi-part applications often focus on this scenario. It is noteworthy that Thiele’s group has modelled the behaviour of the entire human gut microbiota communities using constraint-based models [102]. This approach to modelling has been used to study the interactions between gut microbes and human intestinal enterocytes under anoxic and normoxic conditions [103], to predict levels of short chain fatty acids used to treat Crohn’s disease [104] and to determine whether metformin treatment increases agmatine production by gut microbiota, explaining changes throughout the host's lifespan [105]. More recently, microbiome modelling has been used to develop human organ models [106]. Finally, in multi-omics modelling, metabolome data have begun to be combined with microbiome data [107–109]. In general, integration of different omics data, such as metagenomics, proteomics, metabolomics and fluxomics with GEMs offer wider modeling scenarios [110–112].

### 5. Engineering metabolic modelling: Design and optimization

All of the descriptive approaches discussed in the previous section are non-optimizing modelling methods. In the context individual cell modelling, following the development of descriptive methods, new tools to design and engineer high performance strains were profusely developed [113], including strain designing algorithms such as OptKnock [114], OptStrain [115], OptGene [116] and GDLS [117]. Unfortunately, progress at community level has not caught up, i.e. most of the current approaches support neither design nor optimization of natural and/or synthetic microbial communities. However, pioneering efforts in this field pave the way for future development of community-based design and optimization methods. A set goal is to be able to optimize SMC based on their final application. Key parameters/goals to be optimized include, but are not limited to, production, pathway distribution, community stability, medium composition, spatial cell organization and a combination of goals, i.e. a flexible objective. Therefore, beyond methodological classifications [59], the applications of microbial community engineering can be also grouped according to their optimization goal (see Fig. 2). In this context, there are a few tools that can be considered generic, i.e. they could be used to optimize several applications (see Table 4). However, multiple applications have been developed as *ad hoc* systems to optimize very particular tasks (Table 4). In the following sections we classify
and describe in detail the first GEM-based attempts to design microbial communities for biotechnological applications.

5.1. Production

The optimization goal in this group of applications is often maximizing production parameters in terms of either production rates, yields or titers of a specific metabolite of industrial, health or environmental interest. The outputs are parameters used to design the community that fulfills this aim, e.g., strain ratio, C-source ratio, metabolite uptakes, initial biomasses, cross-feeding rates, etc.

The main generic method in this group has been developed using OptCom’s optimization capabilities [80]. OptCom implemented two-level optimization for single strains and communities. By default, OptCom optimizes the community biomass by assuming fixed single strain growth, and returns strain ratio, substrate uptakes and secretion rates. OptCom has been applied to different microbial communities: a syntrophic association through hydrogen between *D. vulgaris* and *M. maripaludis* [80], a phototrophic microbial community based on *Synechococcus* spp. in daylight metabolism as the primary feeder [80], sub-surface anaerobic environments with electron accepting interactions [80] and communities involved in uranium reduction [81].

Apart from OptCom, some *ad-hoc* methods have been developed to optimize production: to maximize ethanol production with *S. cerevisiae* and *E. coli* [118]; to maximize flavonoid production with *E. coli* strains using a scaled-Gaussian model [119] and to maximize yield with three *E. coli* mutants in a chemostat model of competition for a simple sugar (glucose limited conditions) [120].

5.2. Pathway distribution

Methods and applications in this category support consortia engineering by fragmenting and distributing a given complex metabolic pathway between the components of a consortium. This allows division of labour through metabolite exchange, i.e. intermediate metabolites are secreted by one strain and then used by another. These are mainly graph-based methods [39] and thus allow easy identification of i) species responsible for producing a certain metabolite and ii) metabolite trafficking among strains. For example, CoMiDA [121] identifies putative sub-pathways responsible for synthetizing a target product from a series of given substrates while minimizing the number of necessary species in the community. Non-stoichiometry methods are the most usual among graph-based methods even though they are not independent and require post-processing steps (where stoichiometry is taken into account) to verify whether designs are plausible. Some approaches based on MILP (Mixed Integer Linear Programming) rather than FBA, consider stoichiometry to allocate reactions among the community’s single-strain metabolic models. This supports optimization of specific community goals (growth rate or uptake of one compound) [122]. Other approaches have been designed to expand the network with an agglomerative algorithm that adds reactions iteratively instead of fragmenting the network [123].

Generic methods have also been developed to optimize pathway distribution, including: MultiPlus (static/unified), DOLMN (static/multi-part) and BioLEGO 2 (static/multi-part). MultiPlus [124] starts with a hypergraph that integrates several GEM models and has two fixed objectives: minimizing the number of reactions and minimizing exchanged metabolites in a de novo synthesis pathway. Following a MILP optimization approach, DOLMN [125] identifies communities able to survive under constraints (e.g. limited number of reactions) that are difficult to identify manually. BioLEGO 2 [126] allows large-scale simulations of several simultaneous knockouts (KO) and runs comprehensive searches to identify the KOs maximizing ethanol yield.

5.3. Community stability

These applications predict optimal individual growth parameters resulting in stable communities over time. Two generic methods have been developed to optimize this goal, d-OptCom [81] and SteadyCom [83].

d-OptCom includes both descriptive (dynamic) and engineering approaches that are similar to those of its static version, OptCom. d-OptCom is a highly complex method, e.g. it requires a bilinear FBA solver. Optimization is based on a global search feature (BARON), kinetics parameters and additional LP constraints that need to be defined in order to configure a MILP problem with new reactions for new interactions between strains. In addition, d-OptCom is defined as a ‘comprehensive computational framework’ and does not provide any software that supports neither reproducibility nor the development of new applications. It was used to predict the optimal strain ratios in several auxotrophic pairs of *E. coli* consortia [81]. SteadyCom focuses on predicting a common growth rate for all members in a community and then expecting it to be stable. Contrary to other multi-objective methods such as d-OptCom and the flexible methods, it entails a fixed objective. SteadyCom requires linear FBA solver complexity and iterative LP-based optimization. Apart from FBA, SteadyCom is compatible with FVA. The method was applied to a multi-*E. coli* community with amino acid auxotrophy as proof of concept and a simplified human gut microbiome community that was reduced to 9 species to analyse the influence of fibre content from diet [83].

On the other hand, the *ad-hoc* CASINO toolbox [101] is a computational platform focusing on the human gut microbiome. It is designed to study metabolic interactions among microbial species and the host metabolism. From a static point of view, CASINO predicted alterations of amino-acid metabolism due to dietary interventions. This method follows a two-level optimization approach, similar to d-OptCom. It begins by maximizing growth rate at the individual species level to determine uptakes and subsequently optimizes growth rate and resource distribution at the community level. Contrary to other previous methods, CASINO requires experimental data (strain abundances, etc.) as input to configure models.

5.4. Medium composition

Applications in this category aim to predict the optimal concentration of metabolites in a given medium that deliver maximal community performance in terms of growth, production, decontamination, etc. A study by Zampieri and Sauer [127] describes an application to meet this optimization. It is based on a model that returns ideal medium composition to minimize the cost of metabolic cooperation in microbial ecosystems. The system maximized metabolite concentration of medium inputs to minimize the cost of shared essential metabolites while guaranteeing that growth was only possible within the consortium, not as individual strains. This is a comprehensive optimization approach that solves a two-level MILP problem with high computational complexity. A descriptive static/multi-part approach has also been used to optimize medium composition [77]. This used a pair of metabolic models to initially define a minimum medium containing all the metabolites required to sustain growth. Subsequently, they authors iteratively removed carbon sources to hamper growth and added new metabolites to recover growth. It was concluded that medium composition makes symbiotic relationships possible between binary pairs of 7 different strains.
5.5. Spatial organization

Methods developed to meet this objective predict the physical distribution of the community's strains along a 2D/3D space. The only generic method in this group is IndiMeSH [128], which dynamically modelled bacterial dispersion and nutrient diffusion in a 2D pore network based on pore size and nutrient gradients. Modelling space implies simplifying other issues, such as using a simplified versions of GEMs (reducing numbers of reactions and metabolites from thousands to hundreds) or integrating all bacterial biomass per spatial unit in a single reaction regardless of intra-species variation. IndiMeSH was applied to soil habitats using two different consortia: a syntrophic community of *E. coli* with *S. enterica*, and a multi-strain community comprising the obligate aerobic *P. putida* and the facultative anaerobic *P. stutzeri*. BacArena and MMODES, listed above in the descriptive methods, include some spatial features, although optimizing spatial organization is not their principal goal. In an ad hoc application, a combination of GEMs and partial differential equations to describe metabolite diffusion resulted in a dynamic model that was able to predict biofilm thickness [129].

Spatially Linked Microbial Consortia (SLMC) is a conceptual design to engineer consortia. Spatial distribution is optimized using isolated modules and bespoke growth media to improve control and facilitate new strain combinations. SLMC is reviewed by Sala [130] by including GEMs in the process of designing compatible synthetic communities.

5.6. Flexible optimization

This section describes the usage of methods for goal-agnostic optimization, i.e. the optimization goal and the consortium-parameters-to-be-predicted can be independently defined and remain different in each case. They support design and engineering of microbial communities by selecting the consortium configuration that best optimizes a given goal. To the best of our knowledge, FLYCOP [59] is the only system with the ability to do this. The goal can be defined flexibly depending on the consortium's functionality and a particular interest, e.g. community growth rate, stability, medium composition, etc. For example, FLYCOP can be configured to optimize medium composition by selecting metabolites and their initial concentrations from a finite list. FLYCOP's flexible approach could contribute to improving metabolic modelling of microbial communities in ways that go beyond multiple optimization goals not limited to maximizing growth rate. One example of this are applications where we would seek to maximize yields of certain a product of interest when its synthesis pathway has been split between different strains. Another example would be the comparison optimization approaches for different products as is done experimentally in [26] with different fitness functions. Another advantage of FLYCOP's flexible approach is that multiple parameters can be optimized at once rather than having to implement independent optimization processes for each parameter [118]. Besides, FLYCOP lends itself to applications with obligatory mutualistic communities where other engineering approaches do not, e.g. d-OptCom [81]. While production optimization methods are able to maximize yield at the individual cell level, FLYCOP optimizes reaction fluxes within the metabolic model. FLYCOP can manage applications involving GEMs with thousands of reactions where other methods are limited to small models. FLYCOP's flexibility can also applied to single-strain models where each individual strain has a different growth rate, thus not requiring a single growth rate for all strains in the model as other applications do [78,131]. Additionally, FLYCOP supports automatic search opti-
Table 4
Engineering modelling applications. Grouped by the optimization community goal. Focus on optimization/engineering topics. ‘Production’ group includes to optimize different community parameters (strains ratio, carbon source ratio, initial biomass, etc). GR = Growth Rate. Output means the configuration parameters that are predicted. If there is a software available, it is referred to and linked in the column ‘references’ too.

| Specific goal of optimization          | Output                          | Strains                                      | Results and additional details                                                                 | Ref.          |
|----------------------------------------|---------------------------------|----------------------------------------------|-------------------------------------------------------------------------------------------------|---------------|
| **Production**                         |                                 |                                              |                                                                                                |               |
| Maximizing ethanol production          | - carbon source ratio (glucose/xylose) | - *S. cerevisiae* (or *S. stiptis*)         | - ethanol productivity of ~ 1.08 gr/L/h                                                        | [118,185,186] |
|                                        | - mutant initial biomasses      | - *E. coli*                                  | - *In vivo* experiments to determine kinetics parameters                                        |               |
|                                        | - carbon sources ratio (glucose/glycerol)        | - *E. coli strains* (flavonoid pathway fragmented in 2 strains) | - Using a scaled-Gaussian model: carbon source ratio of 0.1 (glucose:glycerol), strains ratio of 7:3 (upstream:downstream) | [119]         |
|                                        | - strains ratio                |                                              | - Production of flavonoids to 40.7 ± 0.1 mg/L, i.e. a 970-fold improvement                     |               |
|                                        |                                 |                                              | - *Also in vivo* experiments to validate the results                                            |               |
|                                        |                                 |                                              | - Predicted (max. methane, ATP and biomass yield) and some *in vivo* data (biomass yield and ATP maintenance) | [131]         |
|                                        |                                 |                                              | - Low biomass yield per strain, vs community goal                                               |               |
|                                        |                                 |                                              | - 2 first strains consortium: 0.45 mol. methane/mol. ethanol                                     |               |
|                                        |                                 |                                              | - *In vivo* validation with literature data from [187]                                          |               |
|                                        |                                 |                                              | - *In vivo* data from [170]. Originally growing in tryptone                                    |               |
|                                        |                                 |                                              | - 3 mutants after evolution in-vivo, with different GRs                                         |               |
|                                        |                                 |                                              | - Glucose limited conditions (LTEE)                                                            |               |
|                                        |                                 |                                              | - Chemostat model of competition for a simple sugar                                            |               |
|                                        |                                 |                                              | - *In silico* model predictions for different glucose concentrations                            |               |
|                                        |                                 |                                              | - >0.0033% of acetate specialist to allow a viable consortium                                   |               |
|                                        |                                 |                                              | - *Strain ratios*: CV101:CV103:CV116 = 0.10:0.65:0.025                                        |               |
|                                        |                                 |                                              | - CV103 best takes up the limiting resource glucose, but excretes acetate and glycerol (and/or a closely-related compound, glycerol 3-phosphate) |               |
|                                        |                                 |                                              | - *In vivo* data from [75].                                                                  |               |
|                                        |                                 |                                              | - *In silico* model with OptCom                                                                | [80]          |
|                                        |                                 |                                              | - Strain ratio: 2.1 in vivo and 2.28:1 *in silico* lactate uptake = 48 µmol/h                   |               |
|                                        |                                 |                                              | - Formate and hydrogen accumulation = 0                                                         |               |
|                                        |                                 |                                              | - *Additional in silico* predictions: concentration of acetate, methane, CO2 and total biomass |               |
|                                        |                                 |                                              | - *In silico* data from [67].                                                                 | [80]          |
|                                        |                                 |                                              | - *In silico* model with OptCom                                                                |               |
|                                        |                                 |                                              | - Fluxes ratio O2/CO2 reactions: 0.03–0.07                                                     |               |
|                                        |                                 |                                              | - *Strain ratio*: 1:1:6 experiment, and from 1:5:1 to 3:5:1 with metagenomics data            |               |
|                                        |                                 |                                              | - *SYN/FAP strain ratio*: 1.5–3.5 in vivo and from 7.94 with (O2/CO2 = 0.07) to 20.26 (0.03) in silico |               |
|                                        |                                 |                                              | - *CO2*: 3.3 in vivo and 3.2 in *silico*                                                        |               |
|                                        |                                 |                                              | - *Several metabolite fluxes details in *Fig.5*                                               |               |
| **Maximizing yield**                   | - initial glucose concentration for stable consortia | - *E. coli*: -glucose specialist CV103-‘respirer’ - acetate specialist CV101-‘ fermenter’ - glycerol specialist CV116 | - *In vivo* data from [67].                                                                   |               |
|                                        | - strains ratio                | - *D. vulgaris*                              | - *In silico* model predictions for different glucose concentrations                            |               |
|                                        | - substrate uptake             | - *M. maripaludis*                           | - >0.003% of acetate specialist to allow a viable consortium                                   |               |
|                                        |                                 |                                              | - *In silico* model with OptCom                                                                | [80]          |
|                                        |                                 |                                              | - Strain ratio: 2.1 in vivo and 2.28:1 *in silico* lactate uptake = 48 µmol/h                   |               |
|                                        |                                 |                                              | - Formate and hydrogen accumulation = 0                                                         |               |
|                                        |                                 |                                              | - *Additional in silico* predictions: concentration of acetate, methane, CO2 and total biomass |               |
|                                        |                                 |                                              | - *In silico* data from [67].                                                                 |               |
| **Maximizing (together): community biomass** | - strains ratio                | - *Synechococcus* spp (SYN)                  | - *In vivo* data from [67].                                                                   |               |
| - yield per single strain (OptCom fixed goal) | - O3/CO2 ratio              | - filamentous anoxygenic phototrophs (FAP) related to *Chloroflexus* and *Roseoflexus* spp | - *In silico* model with OptCom                                                               |               |
|                                        | - sulphate-reducing bacteria (SRB) |                                              | - Fluxes ratio O2/CO2 reactions: 0.03–0.07                                                     |               |
|                                        | - strains ratio                | - *C. cellulolyticum*                        | - *Strain ratio*: 1:1:6 experiment, and from 1:5:1 to 3:5:1 with metagenomics data            |               |
|                                        | - substrate uptake             | - *D. vulgaris*                              | - *SYN/FAP strain ratio*: 1.5–3.5 in vivo and from 7.94 with (O2/CO2 = 0.07) to 20.26 (0.03) in silico |               |
|                                        |                                 | - *G. sulfurreducens*                        | - *CO2*: 3.3 in vivo and 3.2 in *silico*                                                        |               |
|                                        |                                 |                                              | - *Several metabolite fluxes details in *Fig.5*                                               |               |
| **Maximizing (together): community biomass** | - strains ratio                | - *S. onedensis* (acetate producer)           | - *In vivo* data from [180].                                                                  | [80]          |
| - yield per single strain (OptCom fixed goal) | - acetate and Fe (III) uptakes | - *G. sulfurreducens* R. ferrireducens        | - *In silico* model with OptCom                                                               |               |
|                                        | - strains ratio                | - Two first ones are uranium reducers         | - Carbon source: lactate = 5 mM                                                               |               |
|                                        | - substrate uptake             |                                              | - In ammonium excess condition ([NH4] = 400 µM)                                                |               |
|                                        |                                 |                                              | - Decrease in the biomass of the uranium-reducing species (SO, GS):                           | [80]          |
|                                        |                                 |                                              | - *Strain ratio*: max.community biomass: 0.056:0.051:0.055                                   |               |
|                                        |                                 |                                              | - *Strain ratio*: max.uranium reduction: 0.039:0.041:0.056                                  |               |
|                                        |                                 |                                              | - *Acetate (GS/RF): 14.9/1.49 when max.uranium reduction                                     |               |
|                                        |                                 |                                              | - *Fe(III) (SO/GS/RF): 28.3/110/2.06 when max.uranium reduction                             |               |
|                                        |                                 |                                              | - *Alternative optimization objective in the manuscript*                                     |               |
|                                        |                                 |                                              | (continued on next page)                                                                     |               |
### Table 4 (continued)

| Specific goal of optimization | Output | Strains | Results and additional details | Ref. |
|------------------------------|--------|---------|--------------------------------|------|
| **2 cases of study:**        |        |         |                                |      |
| - maximizing butyrate        |        |         | **- In silico model combining GEMs with a Markov Decision Process** |      |
| - production                 |        |         | - Predict how to modify the community over time to reach a state of maximum performance |      |
| - maximizing atrazine        |        |         | - Intervention for max. butyrate: insoln increase |      |
| - maximizing degradation     |        |         | - Intervention for max. atrazine degradation: depending on the microbiome state, increase of the biomass of *H. stevensii* is often |      |
| **Pathway distribution**     |        |         |                                |      |
| **Optimizing metabolite**    |        |         | **- In silico framework to design synthetic communities, evaluating which new metabolites could be secreted** |      |
| secretion                    |        |         | - secreted emergent metabolites (highlighting the most common ones), with their associated two-strain consortium and medium composition |      |
| **Secondary goal: medium**   |        |         | - *E. coli*/B. subtilis emergent secretion of both succinate and urea (see Figure S4 and S6 from the original study for more pairs and metabolites) |      |
| composition                  |        |         |                                |      |
| **Maximizing growth or**     |        |         | **- In silico model following a MILP optimization approach (higher computational cost than LP (FBA)), with a Static/Multi-part method** |      |
| compound yield               |        |         | - Given metabolic reactions to distribute |      |
| **Minimizing number of**     |        |         | **- In silico model with CoMiDA** |      |
| species                      |        |         | - Graph-based approach (not GEM) combined with Integer Linear Programming (ILP) |      |
|                              |        |         | - Given selected substrates and products, and a set of available species |      |
|                              |        |         | - Identify putative metabolic pathways from substrates to product |      |
|                              |        |         | - Glycolysis pathway, glucose → pyruvate, 284 species: minimal solution with one species was found. Also, they forced for multi-species solution |      |
|                              |        |         | - With 10,000 random pairs of substrate-product metabolites, 1–3 species are selected among 2051 species |      |
| **2 cases of study:**        |        |         | **- In silico model with MultiPlus, following static/Unified approach** |      |
| - maximizing antibiotics     |        |         | - De novo synthesis of bioactive metabolites |      |
| production                   |        |         | - Results: |      |
| - maximizing propanediol     |        |         | - Case study 1 (antibiotics): 4 solutions with 528 reactions (2 transports, 3 insertions, and 28 endogenous reactions) |      |
| methane yield                |        |         | - Case study 2 (industrial): 6 solutions with 110 reactions (1 transition and 10 endogenous reactions) |      |
| **Optimizing metabolic**     |        |         | **- In silico model with DMMM, following a dynamic approach** |      |
| exchange rates               |        |         | - Model parameters adjusted to *in vivo* data (kinetic ones, biomass, carbon and nitrogen sources ratio) |      |
| carbon/nitrogen exchange     |        |         | - Anaerobic species with hydrogen and nitrogen cross-feeding |      |
| and uptake rates             |        |         | - Co-cultures with uni- and multidirectional metabolic interactions |      |
| - kinetic                    |        |         | - The metabolic models can simulate their experimental data, in 4 different cultivation conditions (with/out NH4 and/or NO3), with distinct metabolic capabilities |      |
| parameters                   |        |         | **- In silico model with DOLMN, following a MILP optimization approach, with a Static/Multi-part method** |      |
| **Surviving under constraints** |        |         | - Results: |      |
| Cross-feeding partnerships   |        |         | - core: 91 combinations of 2 strains. Split the TCA cycle into two halves |      |
| and division of labor        | E. coli (2–3 strains) | | - full with reduced functionalities: 2207 combinations for 2 strains, and 2402 for 3 strains. At least 215 and 203 internal reactions to grow, respectively for 2 and 3 strain consortia. Loss one reaction is not compensated with adding one metabolite in the medium (nonlinear boundary) |      |
| **Maximizing ethanol yield** |        |         | **- In silico model with BioLEGO 2. Based on Microsoft Azure Cloud.** |      |
| KO in strains                | S. cerevisiae | | Analysis of two-step fermentation pathway of Ulva sp. biomass into ethanol with KOs in each strain from the consortium |      |
|                              | E. coli | | - 6,649,115 possible single KO analysed scenarios |      |
|                              |         |         | - Ethanol yield increased at 170% of WT (for 867 KO candidate pairs) |      |
Table 4 (continued)

| Specific goal of optimization                  | Output                  | Strains                              | Results and additional details                                                                 | Ref.          |
|------------------------------------------------|-------------------------|--------------------------------------|--------------------------------------------------------------------------------------------------|---------------|
| Stability                                      | Maximizing (together):  | strains ratio                        | Auxotrophic E. coli pairs: (argH-lysA) (lysA, trpC) (metE, ilvE)                               | dOptCom [81] |
| - biomass per single strain                    |                         |                                      | In vivo data from [169].                                                                         |               |
| - community biomass concentration (cells/L)    |                         |                                      | In silico model with dOptCom                                                                   |               |
| GR in auxotroph evolution                      | strains ratio           | E. coli lysine and leucine KOs       | Increased GR by 3 folds, while decreased growth in mono-culture                                | [156]         |
|                                               |                         | long-term                            | In vivo data to constrains the model                                                             |               |
|                                               |                         |                                      | Increased GR by 3 folds, while decreased growth in mono-culture                                |               |
| Common growth                                  | - strains ratio         | - E. coli (KO metE) in lactate       | Strain ratio depending on the aa uptake rate                                                    | COMETS [89]  |
| Secondary goal: spatial distribution           |                         | - S. enterica (secretes methionine)  | In silico model with COMETS                                                                   |               |
|                                               | - cross-feeding rate    |                                      | Strain ratios: E. coli:S. enterica = 75–80:25–20%                                               |               |
|                                               | - spatial distribution  |                                      | Spatial distribution: a strain competitor between cross-feeding species reduces the growth of   |               |
| GR with optimum distribution of resources      | - metabolites (amino-acids) consumption | E. rectale or F. prausnitzii, B. thetataoimicro, B. adolescentis and R. bromii | In vivo data to constrains the model                                                             | CASINO [101] |
| Common growth                                  | - strains ratio         | - 4 E. coli auxotrophic for amino acids | In silico model with SteadyCom                                                                | SteadyCom [81]|
| - community GR                                 |                         | - Gut microbiome (9 species)          | 4 E. coli case of study:                                                                        |               |
|                                               |                         |                                      | 4 E. coli case of study:                                                                        |               |
| Medium composition                             | Minimizing the cost of | Combination of nutrients allowing     | In silico model following a static/Multi-part approach                                          | [127]         |
| metabolic cooperation                          |                         | synergistic growth                   | Selected nutrients: supplementation of nucleotide precursors (maltose, xanthine and inosine)    |               |
| Spatial organization                           | Spatial Partitioning    | P. aeruginosa S. aureus (chronic wound biofilm) | In silico dynamic model combining GEM with partial differential equations | [129]         |
|                                               |                         |                                     | Results                                                                                         |               |
|                                               |                         |                                     | Tendency of the two bacteria to spatially partition, as observed experimentally. Nutrient       |               |
|                                               |                         |                                     | gradients influence (oxygen-top-aerobic, glucose-bottom-anaerobic)                               |               |
|                                               |                         |                                     | In vivo experimental validation: the predicted medium allows growth                             |               |
| Spatial Partitioning                           |                         | 2 case of study (reduced models):    | In silico model with IndiMeSH, following a dynamic approach                                    | IndiMeSH [128]|
|                                               |                         | - E. coli, S. enterica               | Study of soil habitat                                                                            |               |
|                                               |                         | - P. putida, P. stutzeri             | Compared to COMETS and experimental data                                                       |               |
| Flexible                                       | Optimizing PHA          | - initial biomass                    | In silico model with FLYCOP                                                                     | FLYCOP [59]  |
| accumulated                                    | passed production       | - S. elongatus                       | - biomass: 2, 0.2 g/L                                                                          |               |
|                                               |                         | - P. putida                          | - NH4: 0.5 mM                                                                                  |               |
| Stability maximization                         | (common growth)         | - NH4 concentration                 | - sucrose secretion rate: 40%                                                                   |               |
|                                               |                         | - sucrose secretion rate             | - PHA production: 22.43 mM/100 h                                                                |               |
| Several optimization goals:                    |                         | 4 E. coli auxotrophic for amino acids | In silico model with FLYCOP                                                                     | FLYCOP [59]  |
| maximizing yield or biomass or GR, and         |                         |                                      | - strains ratio: Ec1 = 35%, Ec2 = 10%, Ec3 = 15%, Ec4 = 40%                                     |               |
| minimizing time                                |                         |                                      | - aa secretion rate (in terms of %GR): Arg = 1.5, Lys = 2, Met = 1, Phe = 1                    |               |
|                                               |                         |                                      | In vivo data from Lenski’s experiment (LTEE)                                                     |               |
|                                               |                         |                                      | Different configurations are predicted depending on the optimization goal. A polymorphism with   |               |
|                                               |                         |                                      | 2 strains growing is the best configuration under limited oxygen conditions; else only one strain |               |

continued
mization versus the systematic assessment of different configurations. Regarding technical features, FLYCOP has linear FBA solver complexity and optimizes using a local search approach (SMAC).

Among the different categories of applications designed to engineer microbial communities using GEMs (see Fig. 2), the most widespread optimization goal is production followed by pathway distribution. Fewer examples focus on the optimization of other parameters, e.g., stability, medium composition, spatial organization and even a flexible goal. The most common consortia are two-strain (see ‘strains’ column in Table 4), although there are some with a higher number of strains, both homogeneous and heterogeneous.

6. Summary and outlook

As we realize that microbes are rarely found alone but in the context of complex communities, the need for computational tools able to provide mechanistic knowledge of how these communities work and evolve over time becomes critically important. Microbial communities are already recognized as key players in human health and they have begun to be seen as promising biocatalysts in biotechnology applications. Following the development of individual-cell modelling approaches and pioneering efforts on community-level modelling, it is largely expected that methods for the efficient analysis and engineering of such communities will spring up in the coming years. Combining GEMs with Machine Learning or Artificial Intelligence techniques have been suggested as promising developments for metabolic modeling [132,133] and its application to metabolic engineering [134].

In this context, the debate over the real objective function at community level is a long running one [135]. The general assumption is that the microbial community’s goal is to maximize growth under a natural selection scenario. However, optimizing biomass might not be the right microbial goal with genetically engineered organisms or when the environment is different from that where its evolution can occur [91]. Thus, alternative community configurations implying alternative goals are ignored by most of the available methods. It would therefore be interesting to have methods that support optimization of different community goals.

Current dynamic methods rely on the analysis of a few species and only static/unified methods can be used to analyse complex communities, thus hampering a deeper understanding of such communities. An important challenge to address in the near future is the development of tools to support dynamic analyses of large microbial communities in the context of high-quality GEMs [70]. This will require not just a larger GEM portfolio but also collecting large sets of kinetic parameters and developing novel computational methods to reduce the very time-consuming dFBA-based solving stage.

Beyond increasing the complexity of microbial community engineering using GEMs, it is important to note that the components of a given community often operate under different sets of conditions. Differences involving nutrient preference can be easily taken into account, so model-based analysis becomes very useful when defining a common medium that supports growth for all the strains in the consortium. However, GEMs cannot be used directly to model many other environmental conditions such as pH and temperature. Approaches based on the inclusion of omics data to constrain the models have shown to be an interesting alternative. In any case, applications of computational modelling to synthetic has only been tested with a set of microbes living in physiologically compatible environmental conditions.

Another important challenge in microbial community modeling is validation, i.e. proving model usability beyond the computational context. Experimental validation requires a very controlled environment to reduce microbial communities’ high complexity [136]. Validating community dynamics approaches is even more complex. Thus, experimental validation is currently viable when working with small size communities, i.e. two or three components, or synthetic communities. Validation becomes much more difficult when working with larger or natural communities (soil & gut microbiomes, etc.). In vitro simulators could be a suitable alternative [137,138].

In the long term, model-guided microbial community engineering should trend towards the development of technologies capable of predicting potential genetic modifications at the community-level (similar to what the individual-level OptKnock or GDLS algorithms are capable of). The same applies to all the useful and comprehensive COBRA-related algorithms currently used to engineer individual strains.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreements no. 814650 (SynBio4Flav) and 870294 (MixUp) and the “Severo Ochoa Program for Centres of Excellence in RD&I” from the Agencia Estatal de Investigación of Spain (grant SEV-2016-0672 (2017-2021)) to the CBGP. BG was supported by a Postdoctoral contract associated to the Severo Ochoa Program. The authors thank Clive A. Dove for critical reading of the manuscript.

References

[1] Park SY, Yang D, Ha SH, Lee SY. Metabolic Engineering of Microorganisms for the Production of Natural Compounds. Adv Biosys 2018;2(1):1700190. https://doi.org/10.1007/s12934-017-0190-0.

[2] Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nat Rev Genet 2010;11(5):360-79. https://doi.org/10.1038/nrg2775.

[3] Kumar V, Baweja M, Singh PK, Shukla P. Recent developments in systems biology and metabolic engineering of plant-microbe interactions. Front Plant Sci 2016;7:1421. https://doi.org/10.3389/fpls.2016.01421.

[4] Revuelta JL, Buey RM, Ledesma-Amaro R, Varela-Eldridge L, Miguel MJ. Bioengineering microbial consortia for the synthesis of (pro)vitamins, biopigments and antioxidants: challenges and opportunities. Microb Biotechnol 2016;9 (3):564-7. https://doi.org/10.1111/1751-7915.12379.

[5] Vallerio DA. Environmental biotechnology : a biosystems approach. 2015.

[6] P. Dvokáč P, Dvokáč P, Někůš P, Dvokáč P, Dvokáč P, Dvokáč P, Dvokáč P. Engineering pollutant-removing bacteria in the times of systemic biology. Biotechnol Adv 2017;35:845–66. 10.1016/j.biotechadv.2017.08.001.

[7] Yuan S-F, Alper HS. Metabolic engineering of microbial cell factories for production of nutraceuticals. Microb Cell Fact 2019;18(1). https://doi.org/10.1186/s12934-019-1096-y.

[8] Song H, Ding M-Z, Jia X-Q, Ma Q, Yuan Y-J. Synthetic microbial consortia: from systematic analysis to construction and applications. Chem Soc Rev 2014;43 (20):6544–81. https://doi.org/10.1039/C4CS00114A.

[9] Bernstein HC, Carlson RP. Microbial consortia engineering for cellular factories: In vitro to in silico systems. Comput Struct Biotechnol J 2012;3 (4):e2012010017. https://doi.org/10.5539/csbj.2012.10017.

[10] Wang R, Zhao S, Wang Z, Kofnas MAG. Recent advances in modular co-culture engineering for synthesis of natural products. Curr Opin Biotechnol 2020;62:65–71. https://doi.org/10.1016/j.copbio.2019.09.004.

[11] Szóbo E, Wendisch VF. Synthetic microbial consortia for small molecule production. Curr Opin Biotechnol 2020;62:72–9. https://doi.org/10.1016/j.copbio.2019.09.011.

[12] Amor DK, Bello MD. Bottom-up approaches to synthetic cooperation in microbial communities. Life 2019. https://doi.org/10.3390/life9010022.

[13] Che S, Men Y. Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. J Ind Microbiol Biotechnol 2019;46(9-10):1343–58. https://doi.org/10.1007/s10295-019-0211-4.

[14] Qian X, Chen L, Sai Y, Chen C, Zhang W, Zhou J, Dong W, Jiang M, Xin F, Ochsenreither K. Biotechnological potential and applications of microbial consortia. Biotechnol Adv 2020;40:107500. https://doi.org/10.1016/j.biotechadv.2019.107500.
Bhatia SK, Yoon JJ, Kim HJ, Hong JW, Gi Hong Y, Song HS, et al. Engineering of Gil-Sánchez I, Cueva C, Tamargo A, Quintela JC, de la Fuente E, Walker AW, Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, Phalak P, Chen J, Carlson RP, Henson MA. Metabolic modeling of a chronic Vitkin E, Gillis A, Polikovsky M, Bender B, Golberg A, Yakhini Z. Distributed synthetic microbial ecology: engineering Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, among organisms in a microbial ecosystem. 2016 IEEE 55th Conf. Decis. Control, CDC 2016, Institute of Electrical and Electronics Engineers Inc.; 2016, p. 7063–8. https://doi.org/10.1109/CDC.2016.7799357.

Handorf T, Ebenhöhl O, Heinrich R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. J Mol Evol 2005;61:498–512. https://doi.org/10.1007/s00239-005-0077-1.

Julien-Laferrère A, Bulteau I, Parrot D, Marchetti-Spaccamela A, Stougie L, Vinga S, et al. A combinatorial algorithm for microbial consortia synthetic design. Sci Rep 2016;6:20182. https://doi.org/10.1038/srep20182.

Thommes M, Wang T, Zhao Q, Paschalidis IC, Segrè D. Designing metabolic network-driven analysis of bacterial ecological strategies. Genome Biol 2016;17:103. https://doi.org/10.1186/gb-2016-17-10-103.

Ben SS, Or D, Ben Said S, Or D. Synthetic fungal consortia. Biotechnol Bioeng 2018;115:1096–100. https://doi.org/10.1002/bit.26509.

Scholz SA, Graves I, Minty JJ, Lin XN. Production of cellulosic organic acids via synthetic fungal consortia. Biotechnol Bioeng 2018;115:906–100. https://doi.org/10.1002/bit.36500.
