Review Article

Constraint-based modeling in microbial food biotechnology

Martin H. Rau and Ahmad A. Zeidan
Discovery, R&D Microbial Platform, Chr. Hansen A/S, 2970 Hørsholm, Denmark

Correspondence: Ahmad A. Zeidan (dkahze@chr-hansen.com)

Genome-scale metabolic network reconstruction offers a means to leverage the value of the exponentially growing genomics data and integrate it with other biological knowledge in a structured format. Constraint-based modeling (CBM) enables both the qualitative and quantitative analyses of the reconstructed networks. The rapid advancements in these areas can benefit both the industrial production of microbial food cultures and their application in food processing. CBM provides several avenues for improving our mechanistic understanding of physiology and genotype–phenotype relationships. This is essential for the rational improvement of industrial strains, which can further be facilitated through various model-guided strain design approaches. CBM of microbial communities offers a valuable tool for the rational design of defined food cultures, where it can catalyze hypothesis generation and provide unintuitive rationales for the development of enhanced community phenotypes and, consequently, novel or improved food products. In the industrial-scale production of microorganisms for food cultures, CBM may enable a knowledge-driven bioprocess optimization by rationally identifying strategies for growth and stability improvement. Through these applications, we believe that CBM can become a powerful tool for guiding the areas of strain development, culture development and process optimization in the production of food cultures. Nevertheless, in order to make the correct choice of the modeling framework for a particular application and to interpret model predictions in a biologically meaningful manner, one should be aware of the current limitations of CBM.

Introduction

Microorganisms have been used in making food products for millennia [1]. Besides their primary role in food preservation, by virtue of their fermentative metabolism, they contribute to improving food safety and organoleptic properties and can provide additional health-promoting effects [2]. It is estimated that fermented food represents one-third of the human diet [3], which allows it to secure, together with probiotic products, a global market value of over €100 billion [4]. Although food fermentations commonly relied on the spontaneous metabolic activities of the indigenous microbiota in the food matrix, modern fermentation-based food processing mostly employs commercial starter and secondary cultures that are produced in ton quantities per annum in industrial-scale bioreactors [5–8]. These cultures typically comprise a simple or complex consortium of live microorganisms (bacteria, yeast and/or mold) that is formulated to obtain the desired characteristics in the final food product. Both the industrial production of microbial food cultures and their application in food processing are attractive areas for applying emerging molecular and systems-based approaches constantly being exploited in other fields of industrial biotechnology.

Genome-scale reconstruction of metabolic networks combined with constraint-based modeling (CBM) is an increasingly popular approach in microbial systems biology. Although the reconstruction is primarily based on genomics data, it is iteratively refined and gap-filled using biochemical, genetic and/or physiological knowledge on the organism of interest [9,10]. Converting the network...
reconstruction into a mathematical model, commonly referred to as a genome-scale model (GEM), enables its qualitative and quantitative analyses via an array of constraint-based methods, such as flux balance analysis [11–14]. This facilitates further refinement of the network reconstruction and provides valuable insights into genotype–phenotype relationships. Over the past three decades, CBM has found various biotechnological and medical applications ranging from the basic analysis of biological network properties to identifying and simulating complex microbe–microbe and host–microbe interactions [15–20].

Although different mathematical modeling approaches have been adopted in various areas of food biotechnology [21–23], the full potential of CBM remains largely unexplored in this field. Reconstruction and analysis of strain-specific GEMs can be used to comprehensively assess the metabolic potential of industrial strains used in food processing and generate mechanistic understanding of different physiological and biotechnological traits. Understanding interactions in microbial consortia and simulating community phenotype are also of great value in designing culture blends for food fermentations and predicting the characteristics of the final food product. In the commercial production of food cultures, CBM may enable a knowledge-driven process optimization by rationally identifying strategies for growth and stability improvement.

Here, we explore different opportunities through which CBM can guide strain development, culture development and bioprocess optimization in the production of microbial food cultures (Figure 1), with focus on lactic acid bacteria (LAB) as a fundamental group of organisms used in the food industry. Challenges facing the systematic implementation of CBM in this field are also discussed.

**Mechanistic understanding of microbial physiology**

A GEM constitutes a structured organismal knowledgebase [24] that can be consulted for both simple and sophisticated queries — ranging from assessing the putative presence of a certain pathway to the detailed metabolic characterization of an organism in multiple conditions (Figure 2). Besides, new insights frequently originate from identifying inconsistencies between model output and experimental data, which may indicate missing or incorrect metabolic knowledge. The iterative cycles of model curation and experimental validation enable the generation of a refined model, together with new knowledge on unknown pathways, unannotated or misannotated genes and promiscuous enzymes [25,26]. CBM methods allow the calculation of optimal flux distributions under different environmental conditions, enabling the rapid simulation of multiple phenotypic
states. By imposing appropriate constraints [12,27] and selecting biologically relevant cellular objectives [28], it is thereby feasible to perform a systematic metabolic characterization of an organism both quantitatively and qualitatively [14].

The reconstruction of GEMs for several LAB used in food fermentations, e.g. *Lactococcus lactis* [29,30], *Lactobacillus plantarum* [31], *Streptococcus thermophilus* [32], *Lactobacillus casei* [33], *Oenococcus oeni* [34], and *Leuconostoc mesenteroides* [35], provides examples on how CBM can guide biological discoveries and improve our understanding of industrially relevant traits, such as carbohydrate utilization, amino acid auxotrophies, nutrient requirements or metabolite secretion. Constraint-based analysis of an *L. plantarum* GEM revealed how the catabolism of branched-chain and aromatic amino acids indirectly contributes to ATP generation and how the secretion of other amino acids, such as aspartate and arginine, provides a sink for the generated nitrogen [31,36]. The use of CBM and elementary mode analysis also facilitated the understanding of the adaptive growth strategy of *L. plantarum*, usually grown in a nutritionally rich milk environment, to a poor carbon source [37]. In *L. lactis*, CBM suggested that the formation of some flavor compounds is energetically favorable, hence positively affecting growth, and provided a platform for exploring flavor formation under different conditions [29]. The use of CBM also provided insights into amino acid metabolism in *L. lactis* and *S. thermophilus* and identified a unique pathway for the production of acetaldehyde, which contributes to yoghurt flavor, in the latter [30,32].

Many inherent biological constraints are not represented by reaction stoichiometry or thermodynamics alone. However, global phenotype measurements using high-throughput omics technologies provide a reflection of many of these biological constraints and can be integrated into GEMs using various approaches [38–42]. These approaches rely primarily on the detailed gene–protein–reaction mapping represented in GEMs for introducing additional constraints into the network (Figure 3). The integration of high-throughput data into GEMs can, therefore, increase the accuracy of model predictions by shrinking the solution space [43]. This enables the systematic computation of condition-specific metabolic states of the cells. The benefit of integrating omics data with GEMs is not merely unidirectional, as the latter can provide context for omics data analysis and facilitate their interpretation [38,44], which improves the chances of identifying regulatory mechanisms and possible metabolic bottlenecks [45,46]. Recently, the integrative analysis of transcriptomics data within a CBM framework in *Leuconostoc mesenteroides* (used in various food fermentations) contributed to understanding the metabolic behavior and regulatory mechanisms underlying the higher growth rate of the organism on sucrose than on glucose and generated new insights related to the influence of redox metabolism on cellular phenotype [35].

Figure 2. Genome-scale models provide a platform for the mechanistic understanding of physiology.

The comprehensive information contained within a GEM allows relevant biochemical knowledge to be easily extracted, ad hoc queries answered, hypotheses tested *in silico* through CBM and generated for subsequent *in vivo* verification. Moreover, integration of omics data can improve model predictions and vice versa GEMs can improve omics data analysis, while applying CBM allows a global qualitative and quantitative microbial metabolic characterization within a given environment.
CBM can also be integrated with comparative genomics for the identification of genetic determinants underlying strain-specific phenotypic differences [47]. This can be useful in the analysis of genetic variations arising from mutagenesis or laboratory evolution [48].

Thus, CBM provides several avenues for the systematic increase in physiological understanding, an implicit requirement for the advancement of rational strain and process development [49–51].

In silico strain design

In the food industry, microbial strains with novel or improved functional, technological and/or safety aspects are in continuous demand, either for developing new products or improving the quality of existing ones. Additionally, there is a growing interest in the use of LAB for the in situ production and delivery of functional food ingredients, such as low-calorie sweeteners, vitamins, amino acids and bioactive peptides [52–55], as well as heterologous proteins, e.g. vaccines [56]. This may require extensive strain improvement before reaching reasonable titers of the desired product.

Using GEMs as a predictive framework for identifying non-intuitive metabolic engineering strategies for strain improvement is one of the most common applications of CBM [15,16,49,57–59]. This is enabled via a variety of computational algorithms dedicated to model-based strain design (reviewed in refs [59–61]), several of which aim at finding genetic perturbations that couple the secretion of a product of interest to cellular growth, thereby forcing product formation as a requirement for increased fitness. In L. lactis, two CBM frameworks were used to find gene deletion targets for enhancing the production of diacetyl, which imparts a butter-like flavor to dairy products [30]. In silico predictions identified a combination of gene deletions that would result in 150% increase in the production of the direct diacetyl precursor (2-acetolactate), while maintaining reasonable growth. An updated version of the model was supplemented with reactions representing recombinant protein expression to identify gene targets for increasing the specific expression of heterologous proteins [62]. This is particularly relevant for using L. lactis as oral vehicles for delivering recombinant protein vaccines. Two of the in silico designs were experimentally evaluated and provided qualitative evidence on the validity of the model predictions. These approaches can easily be extrapolated to other industrially relevant traits.

CBM can also be used to systematically assess the ability to produce a range of native and non-native metabolites [63,64], which can be applied to fully catalog the potential of food microorganisms as in vivo cell factories.

A major bottleneck in food biotechnology, however, is the tight requirements of regulatory agencies currently imposed in many countries, especially in Europe [65], and the negative perception by consumers of genetically modified microorganisms. This renders the use of recombinant DNA technologies mostly inapplicable and
shifts the focus of strain improvement programs toward classical methods, such as random mutagenesis and screening, dominant selection and directed evolution [66]. To address this issue, a computational algorithm was recently developed that relies on CBM and cheminformatics tools to identify metabolite analogs that can be used to manipulate cellular phenotypes without the use of recombinant DNA technologies [67]. Metabolite analogs are compounds that mimic the structure of native metabolites and, therefore, interfere with their metabolism, which makes them suitable for selection of variants with altered phenotypes. Despite that metabolite analogs are already being used in classical strain improvement schemes [66], the new algorithm extends their applicability by identifying non-intuitive metabolite targets and retrieving corresponding chemical analogs that can provide the selection pressure to redirect the metabolic fluxes towards a product of interest [67].

As the legislations regarding genetically modified organisms are constantly being revised [7], it could be possible in the near future to apply recent genome-editing technologies for implementing metabolic engineering and synthetic biology strategies to manipulate microbial strains used in the food industry [68] and hence take full advantage of CBM in this area.

**Microbial community design**

Food cultures are commonly based on a microbial community rather than a single strain [69]. The overall function of the culture, which dictates the characteristics of the food product, is largely modulated by the metabolic interactions of the community members. Discerning the role of individual strains within a community and their intertwined metabolic dependencies is a complex experimental task [70,71], making the design of defined microbial communities with predictable and controllable properties a challenge. Typically, commercial food cultures are developed through empirical approaches, e.g. evaluating the performance of a limited number of relevant strain combinations, which cannot practically cover all possible designs nor do they benefit from the wealth of information embedded in microbial genomes.

CBM offers a rational approach to the design of food cultures through the mechanistic understanding of microbial community characteristics and simulating community composition and phenotype [71–73]. CBM techniques developed for modeling individual organisms can be extended for community-level modeling [74]. Depending on the complexity of the community and the scope of application, different frameworks can be adopted for building and analyzing community metabolic models, such as mixed-bag modeling, species compartmentalization or multi-species dynamic modeling [72,75–77].

CBM of microbial communities allows phenotype predictions at both the overall community and individual-strain levels. This allows the identification of strain-specific roles that are an intrinsic part of many food fermentations, as exemplified in the mutualistic relationship between *S. thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* in yoghurt production [78]. Identifying the phenotype of individual strains in a community context, in turn, enables the definition of functional strain groups comprising strains with similar capabilities within the community [79]. This could be particularly useful for model-guided simplification of otherwise complex communities (Figure 4).

![Figure 4. Rational microbial community design.](https://example.com/image.png)

Through community metabolic modeling, the phenotypes of conceived community designs can be predicted *in silico*. This, for instance, allows the rational modulation of existing communities, through addition, replacement or removal of individual strains for the generation of communities with improved or novel functionality. Likewise, model-guided simplification of complex communities allows the design of communities with similar functionality but lower complexity and improved practicality.
A major benefit of microbial community models lies within the ability to provide insights into potential metabolic interactions between community members [14,18,80,81], even in large communities [82]. A communal interaction in milk between the probiotic strain *Lactobacillus rhamnosus* GG and the dairy bacterium *S. thermophilus* was predicted through CBM, where the latter supplied six different metabolites for the enhanced growth of the probiotic bacterium [83]. A community metabolic model could also predict amino acid cross-feeding between yeast and LAB, which was demonstrated in grape juice [84]. Moreover, community-level models can be used to elucidate the interactions among gut microbiota as well as their interactions with diet and their host [19,85–92], a highly relevant application for developing next-generation probiotic cultures [93,94]. Other examples of predicting metabolic interactions through CBM of microbial communities are ample [14,77,95,96].

The ability of CBM to predict metabolic interdependencies provides valuable knowledge for the modulation of community structure and function [97] and makes it possible to systematically evaluate any community designs conceivable (Figure 4). Furthermore, the effect of altering the nutrient environment (i.e. food type) can be easily modeled, enabling beneficial nutrient or even strain adjustments. By tuning model optimality criteria, at the overall community or strain-specific level, towards for instance increased production of flavor compounds [98] or texturizing polysaccharides [99], new insights into advantageous community designs for improving food products can be gained.

### Rational optimization of cellular yield and stability

Industrial-scale production of microbial strains for food cultures employs advanced technologies and requires multidisciplinary knowledge on microbiology, physiology and process engineering [7,8]. Maximizing biomass yield is necessary for improving cost effectiveness [66], while process parameters should ensure stability (i.e. viability and metabolic activity) of the produced organism [6,7]. Bioprocess modeling commonly relies on unstructured models to calculate growth rates [100] and process optimization is often based on empirical approaches using either classical (one-factor-at-a-time) or statistical designs [101–103]. Implementing CBM in this context makes it possible to incorporate detailed metabolic knowledge in the process.

CBM can be used to check the consistency of complex fermentation data and analyze them in the context of the metabolic network [50]. This facilitates data interpretation and helps identifying missing end products, as was previously demonstrated in *L. plantarum* [31,36].

CBM is inherently suitable for calculating theoretical biomass and fermentation–product yields under different environmental conditions. This, potentiated by the mechanistic insights, CBM provides into microbial physiology *(vide supra)*, provides a rational means for growth medium optimization, and several examples with industrial relevance already exist [32,104–107]. Industrial-scale fermentations, however, often employ complex undefined growth media [101], which may limit the usability of CBM, as complex media can be difficult to thoroughly describe, and hence less amenable to rational improvement. Even so, CBM retains the potential of identifying growth-boosting nutrients, e.g. through sensitivity analysis [31], which can be used as supplements to complex media. Alternatively, shifting towards pure substrates to ensure reproducibility, facilitate downstream processing or comply with stricter pharmaceutical standards (e.g. for live biotherapeutics), may allow a more comprehensive application of CBM in medium optimization.

The stability of food cultures is a crucial aspect; cultures have to retain viability and activity throughout their shelf life [108], despite encountering various stressors during production and downstream processing. This can be influenced by medium composition, e.g. carbon source, and the levels of certain intracellular metabolites during fermentation were shown to affect the viability of LAB during bioprocessing [109,110]. By incorporating relevant phenotypic data into CBM, metabolic flux states correlating with viability and activity can be accurately calculated [111]. Subsequent CBM can elucidate relevant medium formulations that result in such desirable metabolic states. Though not using CBM, a similar transcriptomics-based approach was used for correlating phenotype signatures and fermentation conditions in *L. plantarum* [112]. Additionally, correlating flux states with stability may identify candidate biomarkers that can serve as readily measurable indicators of stability [113,114], a much-needed approach for the assessment of food cultures [108].

A large-scale bioreactor constitutes a heterogeneous environment that in turn can cause heterogeneity and suboptimal growth within the microbial population [115,116]. Extended GEMs, particularly dynamic resource allocation models, offer a framework for the prediction of cellular behavior in fluctuating environments [117,118], such as areas of local heterogeneity within the bioreactor. GEMs can also be integrated with existing
bioreactor and bioprocess models in a multi-scale CBM framework, which may help reveal non-trivial dependencies across different processes [119].

**Challenges and future perspectives**

CBM of microbial metabolism can provide quantitative answers to various biological questions and guide the generation of testable hypotheses that can benefit both the production of microbial food cultures and their application in food processing. Nevertheless, several challenges that hinder the full exploitation of CBM in microbial food biotechnology exist. Although extensive efforts are undertaken to facilitate the automated reconstruction [120–128] and refinement [26,129–134] of draft metabolic networks and to develop software tools for the analysis of the resulting models (reviewed in ref. [135]), the generation of high-quality, strain-specific models still requires a considerable amount of effort and knowledge. For industrial strains, this is further complicated by the lack of high-quality phenotypic data [16,49]. Such data are essential for the curation and validation of GEMs, especially in the case of LAB where gene function loss is frequently encountered [21]. Knowledge on strain-specific biomass composition may also be necessary to obtain reliable predictions in some applications, such as culture medium optimization. Advances in methods for the systematic generation of omics data and their incorporation into draft reconstruction pipelines may significantly aid in the automated generation of high-quality GEMs in the near future, which allows for more advanced analyses and provides more reliable predictions.

Other limitations inherent to traditional GEMs, which rely primarily on reaction stoichiometry and directionality as constraints, such as the absence of kinetic and regulatory information, remain also challenging. Incorporating such information in LAB GEMs for instance is essential for predicting homolactic fermentation [30,31,36]. Yet, extended CBM frameworks which account for resource allocation across cellular processes may partially overcome these limitations through the introduction of cellular capacity constraints [136–140].

Additional challenges arise when modeling food fermentations due to the heterogeneous nature of the food matrix, the complexity of some food-fermenting communities and the intricate mechanisms underlying the development of some food characteristics. For instance, during cheese ripening, the role of undefined secondary flora and the detailed biochemical changes that are necessary for flavor development can be challenging to model and directly correlate with sensory attributes [141]. In these scenarios, reducing the level of complexity by identifying key players together with the constant advancements in modeling complex microbial communities may prove helpful.

While constantly being addressed by researchers, awareness of these limitations may be a key to the successful choice of the CBM framework that fits an intended application as well as in the proper interpretation of model predictions in a biologically meaningful manner.

**Abbreviations**

CBM, constraint-based modeling; GEM, genome-scale model; LAB, lactic acid bacteria.

**Competing Interests**

M.H.R. and A.A.Z. are employed by Chr. Hansen A/S, a global supplier of food cultures and enzymes. The authors’ views presented in this manuscript, however, are solely based on scientific grounds and do not reflect the commercial interests of their employer.

**References**


© 2018 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).


© 2018 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).


