The developing infant gut microbiota: mathematical predictions of the effects of oligosaccharides

Versluis, D.M.

Chapter 6

Summarizing discussion, other works, and outlook

In this chapter we will first summarize the methods and results described in chapters 2 through 5. We then proceed to analyse the limitations of our approach, the opportunities for improvement, the field of metabolic modelling in general, and finally some future directions for this modelling approach and the study of the infant gut microbiota in general.

6.1 Thesis overview

In **chapter 1** we introduced the infant gut microbiota, its typical compositions, and some of the factors influencing it. We then introduced our modelling approach, constraint-based modelling, and discussed how it can be applied to modelling bacterial metabolism and bacterial interactions.

In **chapter 2** we examined the role of oxygen in the newborn infant gut, and introduced the multiscale mathematical model based on the previous model by Van Hoek and Merks [122]. The model was extended in several ways, and focused much more tightly on a specific *in vivo* situation: the infant gut microbiota in the first three weeks after birth. We showed how an enzymatic constraint was crucial to correctly model how the metabolism of *Bifidobacterium* spp. depends on the concentration of lactose. On the scale of the infant gut community we showed that the initial presence

of oxygen in the infant gut could explain the succession from *Escherichia coli* to *Bifidobacterium* spp. commonly observed in the infant gut microbiota. Furthermore, the model we created provided a suitable basis for further studies in the next chapters.

In **chapter 3** we extended the nutritional input of the model with prebiotic oligosaccharides, which are common in human milk and some infant formulas. We also further refined the model by using the same diffusion speed for nutrients, metabolites, and bacterial populations. Previously, populations diffused at a slightly different speed from nutrients and metabolites, which was not supported by the available data. Further refinements to the genome-scale metabolic models and the selection of species allowed us to better model cross-feeding relationships on the *Bifidobacterium* metabolites lactate and 1,2-PD. The model predicted that several different microbiota compositions can exist that are dominated by *Bifidobacterium*, and that the cross-feeding species present depends on the substrate available to *Bifidobacterium*. The prebiotic oligosaccharide 2'-FL was predicted to lead to a microbiota rich in *Bifidobacterium* with cross-feeding butyrate producers, such as *Anaerobutyricum hallii*. The prebiotic oligosaccharide GOS was also predicted to lead to a *Bifidobacterium*-rich microbiota, but with the non-butyrate producing cross-feeder *Cutibacterium acnes* instead of butyrate producers. This may have health implications, as butyrate is associated with improved infant health.

In **chapter 4** we further extended the model by including the production of mucin by the simulated gut epithelium and the extracellular digestion of both mucins and the prebiotic oligosaccharides 2'-FL and GOS. The model predicted that the decreased mucin consumption capacity observed in the microbiota of breastfed infants may be explained through the selective stimulation of the non-public goods producing *Bifidobacterium* species *Bifidobacterium longum* by 2'-FL. In contrast to the public goods producing *Bifidobacterium bifidum*, *B. longum* cannot consume intestinal mucin. Although its public goods metabolism allows *B. bifidum* to efficiently consume the prebiotic oligosaccharides in human milk as well as mucins, the model predicted that it also becomes vulnerable to having its public goods 'stolen'. Other bacteria can cross-feed on the public goods, mainly sugars, that *B. bifidum* releases into the medium. The model predicted that because *B. longum* can consume the prebiotic oligosaccharides without creating public goods, it is stimulated more in the infant gut microbiota by breastfeeding, and so reduces mucin consumption.

In **chapter 5** we further examined how the infant gut microbiota may form and reform communities by introducing disturbances into the model. These disturbances

kill much of the microbiota. The microbiota then slowly recovers to a state similar to, but often distinct from, that found in undisturbed infants. The model reproduced the increase in *E. coli* abundance and decrease in *Bifidobacterium* abundance that are often observed *in vivo*, but only for intermediate disturbance strengths. We hypothesised that this occurs because the disturbance disrupts the succession from *E. coli* to *Bifidobacterium* that we described in chapter 2. The model further predicted that supplementation with prebiotics in the model leads to disturbed simulations returning more consistently to a composition similar to that found in undisturbed simulations. We hypothesized that this occurs because prebiotics consistently lead to a large relative abundance of *Bifidobacterium* in the model. This implies that prebiotics may help shape an infant gut microbiota that recovers more consistently from disturbances.

In short, the model predicted that the development of the infant microbiota is regulated by metabolic factors as follows: the initial presence of oxygen stimulates a high abundance of Enterobacteriaceae, such as *E. coli*, with an aerobic metabolism. As oxygen is depleted, other species that grow faster anaerobically on lactose, particularly *Bifidobacterium*, become more abundant. The model further predicted that the prebiotics GOS and 2'-FL affect the development of the infant microbiota by steering the development of the microbial ecosystem to a high abundance of *Bifidobacterium*, particularly *Bifidobacterium longum*. A microbiota dominated by *B. longum*, in turn, consumes less mucin in the model. Through cross-feeding, 2'-FL also stimulates butyrate producers. Taken together, the model predicted a crucial role for oxygen and nutrition in determining the relative abundance of bacterial species. The abundance of species determines the nutrients consumed and metabolites produced in the gut, including metabolites that may have health effects.

In this chapter we will first discuss some limitations and simplifications of the model that should be emphasised, and that may be improved in future versions of this model. We will then discuss other related models and their relation to our modelling approach. Finally, we will discuss several interesting topics and questions that could be investigated with new models.

6.2 Model limitations

6.2.1 The difficulties of modelling metabolism

The model predictions are limited in several ways. We will focus on three aspects: metabolism, spatial variation, and microbial diversity. The model predictions are firstly limited in the quality of their metabolic predictions by their limited scope. The model only included a limited set of genome-scale metabolic models (GEMs), used only their carbon metabolism, and only in the context of the gut of the non-premature vaginally born infant. This means that the model cannot currently make predictions for any situations outside of this context, such as on the effect of different protein concentrations in nutrition, or the effect of disturbances on infants born by cesarean section. Even within the scope of the model it does not capture the full extent of carbon metabolism. As described in previous chapters, we use flux balance analysis (FBA) to predict metabolism. For FBA, an objective function needs to be chosen to serve as a proxy for biomass production. We have previously discussed the effects of our assumptions related to the objective function in chapter 4, but the core issue bears repeating: FBA assumes that bacterial metabolism is optimised to fulfill a specific reaction that represents biomass [13, 240]. However, the evolution of bacterial regulation has not been influenced solely by how efficiently biomass can be produced in isolation. Bacteria also, for example, produce bacteriocins [274], spores [20], adhesion proteins [163], or proteins that interact with the host immune system [17]. Such factors may influence the internal distribution of fluxes, and so may also influence the uptake of nutrients and production of metabolites. It is difficult to assess the impact of this omission on our model predictions. The uptake and production of metabolites is vital to the ecological interactions, such as competition and cross-feeding, that occur in the model, so even small changes to bacterial metabolism may have far-reaching effects. By producing bacteriocins some bacteria may even kill other bacteria [275], which may lead to a higher diversity in the infant gut [276, 274]. Furthermore, oxygen tolerance varies between species, and the presence of oxygen in the newborn infant gut requires bacteria to produce protective enzymes to mitigate oxygen toxicity [277, 278]. The GEMs used in this thesis included oxygen metabolism for relevant species, but did not include the necessity for species that do not use oxygen to protect themselves from oxygen in the environment. *In vitro*, the production of enzymes differs between species and even within subspecies for *Bifidobacterium*

[278], which means that this variation may be important to include in future version of the model. If additional factors were known exactly they could be incorporated in FBA through, for example, minimum fluxes through certain bacteriocin-producing reactions, minimum fluxes through oxygen-tolerating reactions, or techniques such as resource balance analysis [279]. This is difficult in practice, as the circumstances and extent of production in an *in vivo* context are unknown. Nonetheless, their incorporation could improve the quality of metabolic predictions.

This brings us to the next issue, which is that it is generally not known whether FBA predictions are correct, because no *in vitro* data is available to verify them. For some bacteria, the factors influencing metabolism are well documented, and the response to different nutrients is well characterised. This allowed us to verify, in chapter 2, that the predictions that the model makes for the ratio of metabolites produced by *Bifidobacterium longum* from lactose are correct, and that the predictions the model makes for how these ratios change based on the abundance of substrate are also correct [90, 103]. However, for most other species this information is not available, and the FBA predictions remain unvalidated. Wider sets of data have been used to validate GEMs [10], but these often focus largely on growth assays, which merely distinguish growth or no growth. This overlooks the effect of correctly predicting bacterial metabolites, which may greatly influence other bacteria in the ecosystem [89]. Even metabolic assays for particular conditions cannot validate whether the metabolites produced by the bacteria under different conditions are correctly predicted. Correct metabolic predictions for one substrate do not necessarily mean that the metabolic prediction for a different substrate is correct, as we saw for the *B. longum* products of 2'-FL digestion in chapter 3. Here, the model correctly reproduced the metabolic switch in *Bifidobacterium* on low vs. high concentrations of lactose, but not the very similar switch on low vs. high concentrations of 2'-FL. Predictions may be improved through incorporating additional constraints on the FBA solutions, such as rate of change constraints [116] or thermodynamic realizability constraints [110]. Work remains to be done on validating other existing predictions, to better estimate how good the overall model predictions fit reality.

6.2.2 Modelling the spatial distribution of bacteria in the human colon

The next limitation to discuss is our handling of space in the model. Most *in vitro* and *in silico* microbial models use well-mixed conditions to simulate the bacterial environment [280]. This allows for consistent results, but may miss much of the variability that occurs in the colon [281]. The colon represents a unique spatial system, where not only physical effects such as diffusion and flow impact the distribution of bacteria and metabolites, but also active mixing through contractions and peristalsis [282]. In the models used in this thesis, we included a two-dimensional space and bacterial diffusion, but not advection. This led to a spatial separation of primary and secondary consumers in our model, as we described in chapter 2, and as was also observed in earlier work by Van Hoek et al. using a similar spatial approach [122]. This modelling approach is a large simplification, and the model cannot represent or make more complex predictions that take factors such as colonic motility into account. This may influence many of the predictions of the model, as the spatial structure of populations determines whether they can cross-feed, and the model predicts that cross-feeding is crucial for some species.

It is unclear what the distribution and diffusion of bacteria actually is in the infant or adult gut. In the adult gut diffusion has been estimated at values around 3 orders of magnitude higher than what we used in the model [159, 103]. Some spatial *in silico* models of the adult gut have applied advection and similar diffusion values as [159] to bacterial populations, and predict that most bacterial metabolism occurs very close to the entrance of the gut [283], with little spatial separation. In this model bacteria exist throughout the gut, but the high rate of diffusion and advection requires the growth and metabolism of bacteria to occur nearly exclusively in the proximal colon, very close to the entrance from the small intestine.

Another *in silico* model also included estimates of viscosity and active movement of the gut [281]. This model predicted that the effect of any plausible diffusion is negligible compared to the known active movement of the gut, viscosity effects, and an assumed motility of bacteria [281]. They argue that, due to a high viscosity close to the gut wall, a small amount of bacterial motility best explains the persistence of bacteria in the gut. From this, their model predicts that most metabolism occurs in the transverse colon - roughly the middle section [281]. However, metabolism was also present in the other sections. Both of these *in silico* models lack experimental

validation. Additional *in vivo* data would help determine what model best approximates reality. In addition, these studies all model the adult colon - localization of bacteria and bacterial metabolism may be different altogether in the infant gut. Important factors for the spatial distribution of bacteria, such as stool viscosity and colonic motility, are different in infants [284, 162]. A thorough modelling of viscosity and active movement in the infant gut, coupled with *in vivo* data, may provide useful insights into the spatial distribution of metabolism and bacterial abundance specific to the infant gut. These insights could then be used to better estimate what spatial distributions of bacteria, and so what cross-feeding interactions, are plausible in the infant gut.

6.2.3 Modelling more microbial diversity

There are several ways in which the models used in this thesis cannot capture the full breadth of variation that exists between infant gut microbiotas, and how future versions may come closer. This would allow us to provide predictions for the effect of infant gut microbiota in different conditions, and may allow us to answer new questions such as: (1) What is the interaction between nutrition and premature or cesarean birth on the composition of the infant gut microbiota? (2) What is the influence of nutrition when we consider the true diversity of species in the gut, including less abundant bacterial species, Fungi, and Archeae? (3) What is the influence of nutrition on the abundance of specific strains, and how do differences between strains impact the digestion of nutrients? We will discuss each in turn.

Firstly, there is more variability between infant birth conditions than what we model. For example, infants are often born prematurely, or by cesarean section. Premature infants typically have a microbiota with a much higher Enterobacteriaceae or Bacillota abundance and lower *Bifidobacterium* abundance [200, 285]. Cesarean-born infants typically have more *E. coli* and less *Bifidobacterium* and *Bacteroides* [3, 16]. These effects may be related to both an altered colonization by bacteria, as well as altered physiological conditions in the gut. Further modelling could be done to explain why the microbiota develops differently in these infants, and how it can be steered to a healthier composition. The model could provide predictions for the interactions of nutrition with different birth conditions. Nutrition is a factor that may compensate for the typically less health-promoting cesarean section microbiota [50].

Secondly, the model only includes a small subset of all species in the infant gut. In reality, the infant gut microbiota consists of hundreds of species [3], with tens of different species co-existing at the same time [5]. We have selected GEMs of species from the most abundant genera in fecal samples [3, 229], but abundances may be different in the gut microbiota itself [37, 35]. Within the variety of species used, we also limit ourselves to bacterial species. Though their abundance is limited, Fungi and Archaea are also present, and may interact with the bacterial microbiota [200]. Further versions of the model may use more species from these groups, and so provide further predictions and insights into the effect of other factors, such as nutrition, on the abundance of these species.

Finally, genetic variation exists within species. Variation occurs even in genes very crucial to survival in the infant gut, such as genes necessary for the consumption of prebiotic oligosaccharides [286]. This variation can have large repercussions for the overall metabolism of the infant gut microbiota. For example, some infants fed prebiotics do not develop a microbiota that can digest prebiotics, because the microbiota lacks the right genes for prebiotic digestion [32]. Ultimately, these issues come down to the model not incorporating the diversity of bacterial metabolisms that exist in the infant gut, and so missing some of the metabolic diversity. To provide new predictions that apply more broadly to the diversity of infants, future version of the model should also include these metabolisms.

Two new resources for microbial modelling may allow future models to better represent these metabolisms. Firstly, the GEM collection AGORA 2 [287] is greatly expanded compared to the AGORA 1 collection we used in this thesis, and now includes more species and subspecies, as well as more detailed modelling of existing species [287]. For example, where relevant the new models include a separate periplasm compartment. Secondly, the improved thermodynamic database Equilibrator 3 [288] includes support for the dynamic generation of thermodynamic values for any nutrient or metabolite, such as complex mucin structures. This would allow reactions in GEMs to be checked for thermodynamic plausibility, and could potentially guide further refinements of GEMs.

6.3 Related modelling approaches

The work we performed on modelling the infant gut microbiota took place in the context of many other *in silico* approaches to modelling bacterial interactions. A

common question in the study of the gut microbiota is whether interactions are competitive or commensalist [289, 97, 290, 291]. Many approaches exist that focus on different aspects of interactions in the gut microbiota, and aim to answer different kinds of questions around the microbiota. A common approach is to focus on population dynamics, and how these may be influenced by internal and external factors. This is commonly modelled with a set of ordinary differential equations which describe a generalized Lotka-Volterra model [280, 292, 200]. For example, Roa et al. [200] used this approach, and used the absolute abundances of bacteria in the prematurely born infant gut as input. This allowed them to predict the positive and negative effect of bacteria on the abundance of other species, and so predict what relations between bacteria shaped the succession they observed in premature infants. However, their model could not explain how these interactions occurred [200], which is a weakness of Lotka-Volterra models in general [280].

Some other approaches make concrete assumptions or predictions on the means by which populations interact. Interactions may occur in many ways, such as through competition for nutrients, cross-feeding on metabolites [89], antimicrobial toxins produced by bacteria [276] or through phages [293]. Different models altogether focus on the eco-evolutionary aspects of bacterial interactions [294, 295]. By including mutation and selection these models aim to explain how the interactions between species have evolved in a dynamic environment. We will focus here on models that examine interactions between bacteria through a concretely modelled metabolism. Such interactions can be competitive, but through cross-feeding these interactions can also be commensalist or mutualist [89]. Metabolic modelling allows for mechanistic predictions that describe what molecules are important for certain interactions between bacterial population [280]. Many approaches exist to model metabolism [280], but we will focus here on approaches that use genome-scale metabolic model to examine how metabolic interactions can occur, and how these can interact with bacterial abundances to form communities of interacting microbes. We will discuss several important historical models that use genome-scale metabolic networks, as well as a number of state-of-the-art modelling frameworks. We will highlight how these approaches have provided insight into the interactions within the microbiota, and how they have informed our modelling choices. A persistent problem in the creation of useful models with genome-scale metabolic models is selecting realistic uptake constraints and objective functions [240]. These choices, as we will discuss, also influence the degree to which models predict competitive and cooperative interactions. The models we discuss are summarized in table 6.1.

Table 6.1: Important historical and current approaches for the genome-scale metabolic modelling of microbial communities

Name	Year	Scope	Dynamic	Spatial	Uptake constraints	Objective function	Metabolic model source	Notable aspects or conclusions
Cobra toolbox[296]	2007	Single microbes	No	No	Various	Various	Various	First widely used FBA toolbox
Stolyar et al. approach [115]	2007	Carbon metabolism of two soil bacteria species	No	No	Set manually	Total biomass or biomass of one species	Own models	First multispecies model
Zhuang et al. approach [118]	2011	Carbon metabolism of two soil bacteria species	Yes	No	Set manually, then dynamically	Species biomass	Various	First dynamic multispecies model
OptCom [297]	2012	Soil bacteria and bioreactors	No	No	Set manually	Community biomass	Various, including [115]	First multispecies framework
COMETS (2014) [121]	2014	Two- and three-species *in vitro* consortia	Yes	Yes	Set manually, then dynamically, minor components constant	Population biomass	Various	First dynamic spatial framework
DFBALAB [298, 299]	2016	Bioreactors and wound biofilms	Yes	Yes	Set manually, then dynamically	Species biomass and uptake	Various	Bacterial metabolism differs spatially
Van Hoek and Merks approach [122]	2017	Carbon metabolism of human gut microbiota	Yes	Yes	Set dynamically	Species biomass	[300] with adjustments	Bacteria can evolve spatial differentiation
SteadyCom [131]	2017	9 species from the human gut microbiota	No	No	Randomized	Equal biomass production rate by each species	Various	Bacterial composition influenced by fiber uptake.
BacArena [123]	2017	Many systems, including gut microbiota	Yes	Yes	Set manually, then dynamically	Population biomass	Magnusdottir et al. [10] (AGORA)	Bacterial diversity is increased by a mucin gradient
MAMBO [301]	2018	Various human microbiotas	Yes	No	Randomized	Population biomass	Own models	Predicts metabolomics directly from genomics
Chan et al. approach [124]	2019	5 species from the human gut microbiota	Partial	Yes	Randomized	Population biomass and equal biomass production rate	Various	Bacterial diversity is increased by an oxygen gradient
MICOM [302]	2020	Entire human gut microbiota	No	No	From [303]	Population biomass	Own models	Uses abundances as inputs to calculate growth rates and dilutions
µBialSim [304]	2020	773 species from the human gut microbiota	Yes	No	Flat rate for all, then set dynamically	Population biomass	Magnusdottir et al. [10] (AGORA)	First model to use hundreds of metabolic models dynamically
ACBM [305]	2020	Bioreactors	Yes	Yes	Set manually or from transcriptomics, then dynamically	Population biomass	Various	First model to use three-dimensional space
COMETS (2021) [132]	2021	Many microbial systems, including the human gut microbiota [306]	Yes	Yes	Set manually, then dynamically	Population biomass	Various	COMETS can cover many situations
Our approach	2022-2023	Carbon metabolism of 21 species from the infant gut microbiota	Yes	Yes	Set manually, then dynamically	Population biomass	Magnusdottir et al. [10] (AGORA)	Summarized in this chapter

Early work with genome-scale metabolic models focused on analyzing the behaviour of single species using either steady-state FBA [296] or dynamic FBA [113]. This provided useful insights into microbial metabolism, but could not predict the interactions of species. Stolyar et al. [115] created the first multi-species model to predict and explain the interactions between two soil microbes. In this model the carbon metabolisms of two GEMs were linked, so that they could directly exchange metabolites and shared one objective function. This was a 'community' objective function, which maximised either total biomass or the biomass of one of the two component species. Implicitly, this assumed that the bacterial regulation evolved to be cooperative. Stolyar et al. concluded that the bacteria were mutalistic, that syntrophy influences soil bacteria, and that this multi-species approach can give insights that were not possible with a single-species approach [115]. Later work by Zhuang et al. [118] on soil bacteria for uranium bio-remediation performed dynamic FBA on interacting GEMs, where each GEM represented a single species. They termed this approach DMMM, for 'dynamic multi-species metabolic modelling'. Zhuang et al. used multiple GEMs, each with its own objective function, so that each species optimised its own biomass production in each step of the dynamic FBA. This allowed the species to grow separately, and adjust their relationship over time [118], which in turn led to a competitive relationship[118]. Cooperation could still have occurred under these conditions, but is no longer assumed. They concluded that competition can greatly influence bacterial interactions and relative abundances. The model we presented in this thesis used a similar approach by assuming that only population-level objective functions exist, and indeed we also observed both competitive and commensalist relationships, particularly in chapters 3 and 4.

DMMM models were further developed into several frameworks which could model a wide variety of conditions, such as OptCom [297] and COMETS [121]. OptCom used a combination of species-level and community-level objective functions for its FBA, similar to Stolyar et al. The community-level objective functions are necessary in OptCom to model complex systems, such as phototrophic microbial communities. These communities exist in nature, but are not predicted by OptCom without assuming a community-level objective function [297]. This assumption again implies that bacterial behaviour is not entirely selfish, but evolved to limit their own growth in favor of a stable community, which is not necessarily true. In contrast to OptCom, COMETS did not assume a community-level objective function, instead using objective functions unique to each species and calculated per bacterial popu-

lation. This assumes that bacterial behaviour is entirely selfish, and has not evolved to limit growth in favor of stability, which is not necessarily true. COMETS could still represent diverse communities without the community-level objective functions by incorporating spatial diversity [121]. COMETS showed how complex interactions that involve both cooperation and competition can occur when species start in different spatial locations. While in principle extendable to very complex systems, the COMETS publication only included simple systems with up to three species and one colony per species, and notably did not include complex population dynamics. The environment simulated was carbon-limited laboratory medium, which allowed uptake bounds to be set easily. Further work with the COMETS framework tested many pairwise relations between bacteria, and concluded that many competitive, mutualistic, and commensalist interactions could exist in this framework [307]. COMETS has formed the basis for many current approaches, including the one in this thesis, both by using a spatial lattice and by assuming that only population-level objective functions exist. However, it does not yet explore the possibility of complex population dynamics on this lattice.

Work with DFBALAB, another framework that included dynamic FBA and spatial diversity [298], did show how metabolism and species abundances can become varied over one-dimensional space, even when the initial composition of a system is homogeneous. Like COMETS, DFBALAB used individual objective functions. DFBALAB went beyond what was shown with COMETS by demonstrating how, through competition and commensalism, spatially distinct ecological niches may form over time. These niches then lead to a self-reinforcing spatial variability [298]. Further work with DFBALAB also showed that this spatial separation may occur in biofilms [299]. We observed similar spontaneous spatial separation in our model, as we highlight in chapter 2.

Van Hoek and Merks developed a spatial dynamic FBA model that focused on the human gut [122], which the model in this thesis is directly based on. The gut is a unique environment that introduces additional complexities in modelling. As we discussed in section 6.2, there are many more unknown factors in the gut, and data is scarce. To simplify this system, Van Hoek and Merks used an objective function that only required ATP generation for growth. This focused the model to be on carbon metabolism, and circumvented the problem of setting unknown metabolite uptake constraints for rare nutrients. Like COMETS, the model calculated FBA solutions per bacterial population. All populations were derived from the same GEM,

but diverged and evolved over the run of the model, representing an overlap with the eco-evolutionary models that we discussed previously. This led to the formation of different 'species', similar to models that are initialised with distinct species, and so still allows for spatial heterogeneity of metabolism. The authors concluded that spatial heterogeneity involving both commensalism and competition can evolve spontaneously in a spatial gut environment, but that spatial disturbance of the population causes this diversity to collapse.

The SteadyCom framework, by Chan et al., also covered the human gut microbiota, but used static genome-scale metabolic models and did not include spatial or temporal variability [131]. It focused on an adult system of nine representative species, each represented by a GEM, from various sources. It used species-specific objective function, but introduced the additional constraint that all species must have an equal growth rate. This forced the solution to represent a steady state microbiota, which corresponds to the usual state of the adult human microbiota [308]. Steadycom did not limit the metabolism of its models to carbon metabolism. Instead, Steadycom uses a random sampling method to determine metabolite uptake bounds for unknown metabolites. When averaged over 1000 sets, this led to population ratios close to those found *in vivo*, but with a low bacterial diversity, as most species go extinct [131]. Further work with Steadycom concluded that the previous assumptions of equal growth could not support a more diverse microbiota [309]. By assuming that bacteria did not always grow optimally, the number of commensalist relationships increased and many more species could become abundant [309]. This had a similar effect as the community objective functions in OptCom. The authors noted that a spatial approach might also lead to sub-optimal growth, and so allow for more diversity [309]. Steadycom was later combined with a spatial and dynamic FBA approach, which indeed allowed it to predict a more varied human gut microbiota without assuming suboptimal growth, and furthermore predict that a spatial gradient of oxygen from the gut wall increased bacterial diversity in the gut microbiota [124]. This is similar to our results in chapter 2, where we describe how a temporal gradient of oxygen leads to a higher bacterial diversity.

The spatial dynamic FBA framework BacArena was similar in construction to COMETS, but focused more on the heterogeneity in bacterial populations by modelling individual bacteria that have different metabolisms [123],. When adapted to model a section of the human gut, the model predicted that a gradient of mucin concentration from the gut wall creates new niches, and may so explain some of the

microbial diversity in the human gut [123], as we also examined in chapter 4. The BacArena framework was also used to predict what nutrients were important for competition and cooperation in this system. Initial uptake constraints were set by determining a set of essential nutrients, and setting uptake constraints for these at 1µM, while all other uptake constraints were set at 0.1µM. These constraints allowed all species to grow in the model, without explicitly trying to model all nutrient concentrations in the gut. BacArena represented an advanced modelling framework that includes spatial and temporal variability, and provides many of the same features as our model, although its features focused on representing bacterial heterogeneity.

The new 2021 version of the COMETS modelling framework [132] provides further interesting opportunities for standardizing metabolic modelling. This new version extends the framework to also include extracellular enzymes and eco-evolutionary interactions. There is also potential to include other complex bacterial interactions, such as those through toxins and phages [276, 293]. Taken together, it should be possible to translate the model presented in this thesis fully into the new COMETS framework. Testable hypothesis for many of the open questions could potentially be generated with this approach. As-is, it is very difficult to compare the outcome of our studies with these other models, or the outcomes of these models with each other. As we have discussed, several models exist that broadly agree in their modelling approach with the model in this thesis: the gut is modelled with populations of bacteria on a lattice, and metabolism is modelled with dynamic FBA using population-level objective functions [123, 132, 122]. However, there are many small differences in model assumptions, such as how to treat the availability of rare nutrients, whether bacterial advection should be included, or whether the microbiota in the lumen should be considered to be in steady state. Our model is also unique in that it focuses on the infant gut microbiota, and the unique microbiota and physiology associated with infants. Nonetheless, to allow for the better comparison of models, it would be useful for the future development of the field of microbiota modelling if more studies used the same framework.

Newer frameworks have also become more complex in various other ways. For example, the spatial dynamic ACBM framework includes three-dimensional space, in contrast to the earlier models we discussed that all included only two dimensions [305]. This allows ACBM to make more quantitatively accurate predictions of growth rates and nutrient consumption in bioreactors with two-species communities [305]. The µbialsim framework used modern computing power to model the entirety of the

773 species then available from the AGORA database in a single simulation environment [304]. This allowed the framework to model a very complex community, with many different interactions between species [304]. Some newer models also use GEMs differently from the methods we have discussed previously. For example, MAMBO does not use existing genome-scale metabolic models, but takes the meta-genomics of fecal samples as input [301]. It dynamically reconstructs GEMs, and then combines these with bacterial abundances to predict the metabolism of a microbiota, including competitive interactions [301]. This may be particularly useful in the study of microbiotas that have not been well-characterised, as no existing genome-scale metabolic models are necessary. Similarly, the MICOM pipeline takes bacterial abundances as input, but MICOM combines these with existing genome-scale metabolic models to calculate what growth rates and dilutions can plausibly form the observed abundances [302]. MICOM then uses these growth rates as constraints when predicting bacterial metabolism. By working this way MICOM evades the problem of setting unknown uptake bounds, at the cost of requiring bacterial abundances and being unable to predict the development of a microbiota over time [302].

Overall, the modelling of the gut microbiota using genome-scale metabolic methods has given several useful insights. Firstly, it shows the usefulness of taking a multi-scale approach, by considering temporal or spatial variability. Models that incorporate this variability can model more complex systems, without including far-reaching assumptions about community cooperation. Secondly, it has shown that many different ecological interactions are plausible in the gut microbiota, from competition to mutualism, but also that assumptions greatly matter in what kind of interactions are predicted.

6.4 Future directions

We will now highlight how the model presented in this thesis, and its metabolic modelling approach in general, may be useful to the field in the future. Firstly, a closer integration with the eco-evolutionary methods we discussed previously may allow for evolutionary questions to be answered. For example, it is unclear how the mutualistic relationship between humans and *Bifidobacterium* came into existence [70]. In humans, milk oligosaccharides specifically stimulate *Bifidobacterium*, which are beneficial to the host, and this is often used as an example of co-evolution [70, 220]. Milk oligosaccharides were an early development in the evolution of mammals, and

are present in nearly all mammals, including marsupials, monotremes, and all orders of placental mammals [68]. However, *Bifidobacterium* are not typically dominant in the gut of non-human newborn mammals [310, 311, 312, 313, 314]. For example, in newborn Asian elephants (*Elephas maximus*) bacteria of the order Enterobacteriales are initially dominant, as in human infants, but these are then replaced by Lactobacillales and Bacteroidales [313], not *Bifidobacterium* as in human infants. Asian elephant milk has been characterised, and contains a similar composition of oligosaccharides as human milk [68], at a higher concentration [315]. It is unclear what occurs in the guts of these animals, but the milk oligosaccharides may be digested by other bacteria, such as *Bacteroides* or *Lactobacillus* species [313, 310]. It is possible that these bacteria have developed their own mutualistic relationships with non-human animals. Such relationships may be even more complex in marsupials and monotremes, whose milk contains little lactose, but whose milk oligosaccharides are also directly consumed by the newborn [69, 316]. Modelling using eco-evolutionary methods [295, 122, 294] may provide insights into the evolution of unique newborn gut microbiotas, and the evolution of the link between different bacteria and milk oligosaccharides. For example, van Dijk et al. already created a model that predicted that serial transfer causes populations of interacting microbes to either focus on a high yield or a high growth rate [295]. Such a serial transfer between hosts must be an essential part of an evolutionary model of microbiota development. Baijic et al. used a GEM of *E. coli* to examine in detail how its metabolism may evolve and how this may, in turn, influence the environment [294]. Such an examination could also be performed on the GEMS of other common gut bacteria, such as *Bacteroides*, *Bifidobacterium* and *Lactobacillus* species, to potentially uncover mechanisms of co-evolution.

Secondly, a further refinement of the modelling approach may allow for a better understanding of the observed compositions of the human infant gut microbiota. "Everything is everywhere, but the environment selects" is a common tenet in microbiological research [317]. We now know that in the infant gut not everything is always present, but it is clear that the infant gut is exposed to many more bacterial species than the few that become abundant [200, 54]. Thus, selection by the environment is crucial in determining the composition of the gut microbiota. However, for many aspects we do not yet know why the environment selects in the way it does. Traditionally, studies in this field have been mostly observational. These studies have consisted of analyses of fecal samples of infants born and raised under

different conditions. With metagenomic techniques it is now possible to analyse the bacterial and metabolic composition of many fecal samples from the same infant, collected over time. Using these techniques, work by Tsukuda *et al.* [31] and Rao *et al.* [200] has provided extensive descriptions of the large-scale patterns of microbial and metabolic succession, as well as of the variability in development between infants. These analyses can show how the distribution of bacteria differs between infants, and may correlate different abundances with different conditions. However, it cannot answer the more fundamental 'why' questions. For example, as we discussed in chapter 2, *E. coli* is typically more abundant than *Bifidobacterium* because of the initial presence of oxygen in the gut. But why are the strictly anaerobic *Bacteroides* spp. often abundant in the infant gut, while the related *Prevotella* spp. are not? And why is *E. coli* typically abundant in the early infant gut, but sometimes the similarly facultatively anaerobic *Klebsiella* spp. is abundant? Without this understanding it is difficult to predict how these patterns may be influenced to create a healthier infant gut microbiota. This is where we believe *in silico* modelling can play a crucial role, by taking the bacterial and metabolic data as input, and providing further testable mechanistic predictions, as we have shown in this thesis.

In vitro studies may allow for these predictions to be tested. *In vitro* studies have already allowed for some more insight into, for example, the metabolism of *Bifidobacterium* [90]. Next to studies of infant fecal samples, these *in vitro* studies have been the primary source of input into our computational model. However, even in *in vitro* studies it often remains unclear why certain ecological effects occurs - e.g. why *Cutibacterium avidum* outcompetes *Anaerobutyricum hallii* when fed with lactate [194]. It is difficult to gather the right data and do the right experiments to find a mechanistic explanation for such complex effects without existing hypothesis. The wide variety of experimental conditions used in different *in vitro* studies (e.g. [245, 83, 160, 188]) further complicates creating a complete view of the interactions in the microbiota. Some studies include *in vivo* experiments in which germ-free mice or pigs are inoculated with a human microbiota. This allows relevant physiological factors to be included, such as the production of mucin by the gut wall, the absorption of metabolites, and immune interactions [318]. However, this approach also introduces its own biases [318]. Many species establish at different abundances in germ-free mice and pigs, compared to humans, and some species do not establish at all [319]. A better approach may be the recently developed gut-on-a-chip models. These *in vitro* models incorporate a living portion of the gut wall, better imitating the infant

gut conditions than an *in vitro* model that only includes bacteria. A gut-on-a-chip setup can also include physiological interactions, and conditions can be much more tightly controlled than in a mouse or pig model [320, 321]. *In silico* modelling may help guide these models to test more concrete hypothesis, such as the hypotheses we developed in chapter 3 and 4 on cross-feeding relations in the infant gut microbiota.

In conclusion, we expect that the metabolic modelling approach described in this thesis will be refined much further. We also expect that this approach will help explain the composition of the infant gut microbiota, and how it can be shaped.