UNIVERSITY OF BIRMINGHAM

DEVELOPING AND VALIDATING A COMPUTATIONAL MODEL OF THE GUT MICROBIOTA–MUCOSA INTERACTIONS TO REPLACE AND REDUCE ANIMAL EXPERIMENTS

by

TIMOTHY ROGER FOSTER

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

September 2022

ABSTRACT

The human gut is teeming with microbial life. Within the most densely populated part of the gut - the colon - there are an estimated 1 trillion microbial cells. These microbes - the gut microbiota - have profound impacts on our health from the moment of birth, all the way through our lives. Some provide epithelial cells with nutrition, by breaking down dietary fibres, some may be essential in the development of the immune system, and some microbes can provide protection against the invasion of the gut by harmful pathogens. Understanding such a complex and diverse system calls for the development of models that can represent key features of the system, while simplifying it, to make generalisable conclusions. The most widely utilised model of the human gut is the mouse gut, which is used in a great many studies on the gut microbiota. In addition, various *in vitro* models have been developed, which typically represent the gut using chemostats or "gut-on-a-chip" systems that may include epithelial cells. Several population-level *in silico* models of the gut microbiota have also been developed, based on systems of differential equations. However, agent-based modelling - a modelling approach which allows researchers to investigate how local interactions between cells can influence population-level outcomes - has not been utilised extensively to investigate the gut microbiota.

In this thesis, I will present a new modelling platform - eGUT - an agent-based modelling platform for modelling the gut microbiota. This model focusses on the interactions between microbial cells and the host epithelium at the mucosal surface, by explicitly modelling the spatial detail of the mucosal region. This mucosal compartment can be connected to other regions, such as the gut lumen and the host's circulation system, in order to model the interactions between these regions, such as the exchange of chemicals and microbial cells and the flow of digesta through the gut. Partial differential equation solvers are used to model the reactions and diffusion of the various chemicals within the system, allowing interactions such as cross-feeding,

competition and other metabolic interactions to arise naturally, and to explore how niches may be created or partitioned, and how the various behaviours of gut microbes and the host epithelium affect the fate of particular gut microbes.

In order to ensure that the eGUT platform enables a physically and biologically realistic representation of reality, it has been subjected to various tests and validation experiments. These include test of numerical algorithms that ensure the models of solute diffusion and consumption work as expected, as well as more complex tests, such as benchmarking eGUT against other *in silico* models able to simulate biofilms, and conducting validation experiments using MIMic, a chemostat-based *in vitro* model of the gut microbiota.

eGUT now passes all numerical tests without issue, showing that its basic models of reaction and diffusion are functioning correctly. It is also in good agreement with a range of other in silico models when running the benchmark biofilm model, given the variation between models and their fundamentally different approaches. However, the results of a colonisation resistance study conducted both in MIMic and eGUT show different results. It is unclear whether this is the result of problems with the assumptions made in the modelling system, or imprecision in the values of estimated growth parameters for the species used in the model.

Although some issues remain to be further investigated, eGUT has the potential to be a valuable resource to gut scientists wanting to model the interactions between members of the microbiota and their host. eGUT may act as a convenient testing ground for hypotheses and questions regarding population dynamics in the gut, or for the development of probiotics and/or prebiotics. This may lead to the reduction or possibly the replacement of the use of animal models, including mice, in gut research laboratories.

ACKNOWLEDGEMENTS

This thesis is the result of the teaching, encouragement, guidance and support of a huge number of family, friends, teachers and mentors. I am indebted to my parents for their love and support throughout my life, for their encouragements to be inquisitive and pursue knowledge from a young age, and for supporting me through my undergraduate studies and beyond. I am also hugely grateful to my PhD supervisor, Jan-Ulrich Kreft, for his wisdom and advice, and for his belief in my abilities, coupled with a belief that there is always room for improvement. Most of all, Jan's ability to always make time for all of his students is deeply appreciated.

I am also grateful to my funders, the National Centre for the 3 Rs (NC3Rs) for their funding as well as their continuous support and interest in the project.

The eGUT was developed parallel with its sister project, iDynoMiCS 2.0. Much of the shared core of these modelling platforms was developed by Robert Clegg and Bastiaan Cockx. Bastiaan Cockx also provided me with a great deal of support, advice and assistance in the development of eGUT, and has been an invaluable friend and colleague throughout the course of my PhD.

Much of my work would not have been possible without the guidance and assistance of our collaborator, Richard Horniblow, and the rest of the Horniblow lab members, Sarah Corrigan, and Phuong Linh Ta. Thank you, Richard, for your expertise and help in the lab, as well as your advice. Much of the labwork was carried out with the

assistance of a number of Masters and undergraduate students, Karanjot Sandhu, Gabriella Cooper, Mousomi Chakravorty and Migle Savukynaite, whose efforts are greatly appreciated.

I would also like to thank the other members of the Kreft lab, both past and present, for their friendship, academic discussion, and advice - Eleni Christidi, Kim Summers, Roberto De La Cruz Moreno, Cansu Uluseker, Anjali Vasudevan and Sasikaladevi Rathinavelu. These, and the rest of the 3rd floor biosciences group have been a joy to share an office, and the occasional drink with.

When I arrived to start my studies in Birmingham, I had the good fortune to meet a wonderful group of fellow PhD students, who have remained a constant source of friendship, kindness and hilarity throughout my time in Birmingham, as well as introducing me to many other wonderful people. Thank you, Imogen Mansfield, Paris Lalousis, Fabian Burkhardt, Georgia Bird, Kaya Gromocki, Amoul Mangat, Ravi Punn and Marianna Hurst for making life in Birmingham such a joy.

Finally, I would like to thank my wonderful wife, Shakira, for her never-ending support and love throughout my studies, and for never failing to put a smile on my face when I come home. In addition, her efforts to keep me fed and watered whilst writing this thesis are commendable.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

1.1 The Microbiota

Close interactions and associations between unrelated organisms, termed symbiosis, from the Ancient Greek "συμβίωσις" (living together), is a widely observed phenomenon throughout the domains of life on Earth. While symbiotic relationships are often equated with co-operative interactions, in their broadest definition they are not necessarily beneficial to both members of the relationship, instead lying on a scale from mutualism, by which both members benefit, to parasitism, where one member benefits at the cost of the other (Dimijian, 2000). An example of mutualism would be the relationship between legumes and the nitrogen-fixing rhizobia, while pathogenic exploitation of eukaryotes by bacteria such as *Mycobacterium tuberculosis* are a clear example of parasitism (Steinert et al., 2000). In between these two extremes are commensal relationships, where one member derives a benefit and the other experiences no significant fitness effect, such as many of the relationships between epiphytes and trees (Blick and Burns, 2009).

Humans act as hosts to a huge number of microbial symbionts, spanning the whole range from mutualistic to parasitic. These are known as the microbiome or microbiota. These human-associated microbes can reside on the skin and on various epithelial surfaces, but the vast majority reside in the gastrointestinal tract – specifically the colon. The colon of a 100 kg "reference man" contains an estimated 3.8×10^{13} bacteria

while those in or on other organs comprise only around 10^{12} cells in total (Sender et al., 2016).

As there is some ambiguity regarding the meaning of both the microbiota and the microbiome, I will explicitly define them here. In this thesis, I will refer to the community of microbes living in and on an organism as its "microbiota", and to the microbes residing in the gastrointestinal tract specifically, as the "gut microbiota". The "microbiome", on the other hand, will refer to the union of the microbiota and the host (especially the host's gut epithelium) as well as the physical and chemical interactions between all of these organisms. This scheme is a slight modification of that proposed by Marchesi and Ravel (2015), with the main difference being my inclusion of the host in the definition of the microbiome, which I feel more fully encapsulates the entire "biome" that the microbiota inhabits.

1.2 - The Gut Microbiota in Health and Disease

Recent advances in research into the gut microbiome have revealed a multitude of links between the gut microbiota and human health. Differences in the makeup of the microbiota correlate with a range of diseases and dysbioses, including IBD (Gophna et al., 2006; Manichanh et al., 2006; Sokol et al., 2009, 2008; Willing et al., 2009, 2009), immune disorders (Atarashi et al., 2011; Stokes, 2017), obesity (Shen et al., 2013; Turnbaugh et al., 2006) and allergies (Panzer and Lynch, 2015). In addition, the makeup of the microbiota when the gut is challenged by a pathogen plays an important role in whether the pathogen can successfully colonise the gut. For example, the presence of a commensal bacterium in mouse guts – *Mucispirillum schaedleri* – reduces the ability for *Salmonella enterica* serovar Typhimurium to colonise the guts of mice and cause infection, at least partly due to competition for electron acceptors between the two species (Herp et al., 2019).

While the majority of food nutrients are broken down and absorbed by the host in the mouth, stomach and small intestine, many complex carbohydrates and other fibres pass through into the colon. The presence of these nutrients provides a niche for microbes which can continue the breakdown of "indigestible" fibres, providing both themselves and their hosts with nutrition. These microbes have a huge range of metabolic strategies and ferment most of the nutrients not fully digested by the host before reaching the colon, including non-starch polysaccharides, resistant starch, proteins, peptides and fats (Diether and Willing, 2019; Blaak et al., 2020). Many of the

primary products of these fermentation processes are short-chain fatty acids (SCFAs) including acetate, propionate and butyrate (Wilson, 2008). These SCFAs are the primary energy source for colonocytes – the epithelial cells lining the colon (Blaak et al., 2020; Litvak et al., 2018) and are preferred to other energy sources, such as ketone bodies, amino acids and glucose (Roediger, 1980). In particular, evidence from rat colonocytes indicates that butyrate is preferred over acetate and propionate as an energy source (Clausen and Mortensen, 1994).

Many of the typical microbes resident in the gut are either commensals or mutualists. Commensals derive benefits from an interaction with another organism, in this case taking advantage of the favourable environment provided by the host, while the host is unaffected. Other gut microbes are mutualists (synergists), with both the host and the microbe deriving a net benefit from the interaction (Wilson, 2008). However, the gut also provides a habitat for harmful pathogenic microbes, which exploit the host in order to increase their reproductive fitness. These include the diarrhoeal pathogens – *Cryptosporidium* spp., *Campylobacter* spp., enteropathogenic *Escherichia coli*, enterotoxigenic *E. coli*, *Salmonella* spp., *Shigella* spp., *Vibrio* spp. and many more. Together, these pathogens cause millions of deaths each year due to diarrhoea, especially in the developing world. In 2015, diarrhoea caused 1.3 million deaths in children under 5-years-old worldwide, making it the fourth leading cause of death in this age bracket (Troeger et al., 2017).

Other diseases of the gut include *Clostridioides difficile* infection, a colonic infection often causing recurrent diarrhoea and in severe cases, pseudomembranous colitis (Zhu et al., 2018). *C. difficile* infections are associated with antibiotic use and with a disrupted or abnormal gut microbiota (Cammarota et al., 2015; Zhu et al., 2018), which may be actively maintained by *C. difficile* (Fletcher et al., 2021). Antibiotic use typically alters the makeup of the microbiota, leading to chemical changes in the gut favourable to *C. difficile* and high levels of colonisation, including the presence of particular bile acids and carbohydrates, and the absence of key competitors (Theriot et al., 2014). In addition, this favourable environment may be actively maintained by *C. difficile* through toxin secretion (Fletcher et al., 2021).

Changes or abnormalities in the microbiota in early life can have a disproportionately large impact on lifetime health and gut function. For example, malnutrition in young children can have long-lasting effects on the microbiota, preventing the normal adult microbiota from developing, even when more nutritious food supplements are provided. These problems lead to long-term changes in metabolism, which can be difficult or even impossible to reverse and which in turn cause growth issues (Blanton et al., 2016; Subramanian et al., 2014). Disruptions of the microbiota in newborn infants, especially preterm infants, who are typically on antibiotic treatments, can often lead to poor health outcomes and conditions such as necrotising enterocolitis or sepsis (Ficara et al., 2020). The risk to these children of developing these diseases has been linked to the absence of certain key members of typical babies' microbiotas, such

as *Bifidobacterium bifidum* (Hall, 2019). Other factors, such as mode of delivery and breastfeeding vs. formula feeding can have significant impacts on the development of babies' microbiotas. Furthermore, the microbiota plays a fundamental role in the development of the human immune system from birth (Kalbermatter et al., 2021), and its maintenance and function throughout life (Salzman, 2011).

It is therefore clear that an improved understanding of the gut microbiome could lead to significant numbers of lives being improved or saved. Treating these problems will require a better understanding of their root causes, and in order to fully understand how the host and its microbiota interact to lead to disease and dysbiosis we need to understand the ecology and population dynamics of the gut microbiota.

1.3 - Colonisation Resistance

One of the key concepts in understanding the interactions between mutualists, commensals and pathogens in the setting of the gut is colonisation resistance. Colonisation resistance refers to an antagonistic interaction between a commensal or mutualist resident of the gut microbiota and an invading microbe, e.g., a pathogen, by which the resident microbe prevents the pathogen from invading (van der Waaij et al., 1971). This may be due to competition for a particular resource or niche, or to the production of toxins or other chemicals that reduce the growth of the invading pathogen (Fons, 2000). The early work of Rolf Freter and co-workers identified the combined competition for attachment sites and substrates as key requirements for preventing the colonization of mouse guts by invading *E. coli* that were identical to the resident *E. coli* (Freter et al., 1983). The theory of colonisation resistance provides the framework for the development of probiotics and prebiotics – new prophylactic treatments designed to prevent the invasion of the gut by pathogens. Probiotics are by definition beneficial gut microbes that should help to provide colonisation resistance against particular pathogens in the event that they are introduced into the gut. Prebiotics are nutrients that are designed to feed particular gut microbes which may then act as probiotics and prevent colonisation by pathogens (Callaway et al., 2008; Isolauri et al., 2004).

One of the earliest observed examples of colonisation resistance is the antagonism of Enterobacteriaceae infection in humans by *Escherichia coli* Nissle 1917 (*Ec*N). *Ec*N was

first cultured by Alfred Nissle, from the faeces of a German soldier who, unlike his comrades, had survived in an area heavily contaminated by *Shigella* without contracting diarrhoeal disease. *Ec*N was soon patented as an antagonist against enterobacterial pathogens and its production began under the name Mutaflor. Thus, *Ec*N can be considered one of the first probiotics ever marketed and manufactured as such (Sonnenborn, 2016).

The antagonism of enteropathogens by *Ec*N has been linked to a variety of competitive interactions. For example, competition for iron in the gut includes the production of toxic microcins which mimic iron-sequestering siderophores (Massip et al., 2019; Sassone-Corsi et al., 2016) and high-affinity, lipocalin-resistant uptake pathways possessed by *Ec*N (Deriu et al., 2013). *Ec*N also competes successfully for Zinc (Behnsen et al., 2017). Competition for other resources includes galactitol (Eberl et al., 2021) and oxygen (Litvak et al., 2019). The blocking of epithelial cell surface receptors that are used by invasive enterobacteria has also been put forward as a possible mechanism of colonisation resistance (Altenhoefer et al., 2004).

1.4 - Current Gut Modelling Approaches

The importance of the microbiota in human health and disease makes it a vital field of research, but despite the clear relationships between the microbiota and host health, there are still huge gaps in our understanding of how the microbiota impacts host metabolism, health and susceptibility to pathogens. The relationships between host health and the microbiota are subject to many confounding effects from the host genome, lifestyle and diet (Nicholson et al., 2012). The interplay of these many factors makes the system difficult to study and introduces noise that may obscure observable effects. In addition, there are inherent difficulties in studying the human microbiota, as gut microbiota research often requires invasive methods and perturbations that would not be suitable for human studies. As a result, most microbiota research carried out in humans is either metagenomic studies on cohorts of healthy subjects or clinical trials of microbiota-related health interventions (Kostic et al., 2013).

Much of today's microbiota research relies on animal models, including mice – the most widely used model species in scientific research (Robinson et al., 2019), but other model species include zebrafish, *Drosophila* and bobtail squid (Kostic et al., 2013). However, the guts of these various animal models differ from human guts in structure, length, cross-sectional area, transit time, and microbial composition.

More recently, success in the realm of probiotic treatments has been achieved by Panigrahi et al. (2017), in a study on newborn children in India. Here, a combination of a probiotic and a prebiotic (termed collectively a "synbiotic"), composed of the

bacterium *Lactobacillus plantarum* and a fructo-oligosaccharide prebiotic significantly reduced the occurrence of sepsis and death during the first 60 days of life from 9% in the placebo group to 5.4% in the treatment group. A keystone of their success was using a human isolate of *Lactobacillus* that can colonise the gut for a significant time period (Panigrahi et al., 2008).

More widely, trials of probiotics for the treatment of gastroenteritis in a variety of ages have seen many positive results. A meta-analysis of clinical trials of probiotic interventions for acute infectious diarrhoea found that diarrhoea duration in treatment groups was on average 24.76 hours shorter than controls and stool frequency on day 2 was on average 0.8 less. However, the variance in all outcomes was high, with some studies showing no clear difference between control and treatment groups. The grouping of studies using a wide variety of different microbial species may obscure some important differences in the efficacy of different microbes as probiotic interventions (Allen et al., 2010). However, in other areas of human health, such as obesity and body weight, the efficacy of probiotic and symbiotic treatments is far less clear. For example, in a recent meta-analysis of synbiotic trials for the treatment of obesity, there was no clear signal of a positive effect. The statistics for the combined results showed a small statistically significant reduction in weight (800g), but not in BMI, waist circumference or body fat (Hadi et al., 2020).

Another area in which gut microbiota research has yielded success is in the development of faecal microbiota transplantation (FMT). This is the transfer of

screened faecal matter from a donor to a patient's GI tract, with material delivered via a nasogastric tube, colonoscopy, retention enema or a capsule (Allegretti et al., 2019). The primary application for FMT is the treatment of recurrent *C. difficile* infections. Indeed, antibiotic treatment of *C. difficile* infection is so often associated with reinfection and serious complications that FMT has become a very promising alternative to antibiotics (Ianiro et al., 2018). In one clinical trial, a course of FMT appeared to be far more effective than a course of the typical antibiotic treatment, vancomycin, in resolving recurrent *C. difficile* infections (Cammarota et al., 2015).

A systematic review of FMT trials confirms the positive outlook for the use of FMT in the treatment of *C. difficile* infections. Analysing 132 studies, the authors found the mean final cure rate (the rate of cure after multiple FMTs) for *C. difficile* infections was 95.6%. However, the efficacy of FMT in treating other chronic gut diseases was far less promising. The mean final remission rates (remission rates after multiple FMTs) for Crohn's disease and ulcerative colitis were, respectively, 47.5% and 39.6% (Lai et al., 2019).

Current models of the gut microbiota include *in vivo*, *in vitro* and *in silico* models, but the most prevalent model in gut microbiota studies is the mouse. Mice are the most widely used animal model in scientific research, accounting for 73% of all animals used for research in 2018 (Robinson et al., 2019). They are small, docile, easy to breed, have short generation times, and have well-studied genomes with extensive production and study of gene knockouts. Furthermore, the ability to raise gnotobiotic

mice makes them ideal for gut microbiome studies (Guan et al., 2010; Nguyen et al., 2015; Rosenthal and Brown, 2007). However, there are important differences between the guts of mice and our own. Not least, the size – mouse guts are much shorter and thinner than our own – but also the structure – mouse guts have a large functional caecum, and no haustra. Mouse guts also have a much shorter retention time than human guts, not only because of the difference in size, but also because mice have far higher metabolic rates than us, forcing them to eat much greater volumes relative to their body weights than humans. Mice also engage in coprophagy, and have furrows that transport mucus and microbes in the opposite direction to the digesta (Hugenholtz and de Vos, 2018; Sakaguchi, 2003). It is no surprise then, that mouse guts harbour a functionally and phylogenetically different community of microbes from our own (Nguyen et al., 2015). Unfortunately, unlike various *in vitro* and *in vivo* models that will be explored in further depth later, these structural differences cannot be altered. We cannot "build" a mouse with a human-like gut.

There are a variety of *in vitro* models of the gut used by various laboratories across the world to investigate the microbiota. These range from arrays of chemostats, representing the various compartments of the gut (i.e. – the macro-scale structure) to microfluidic models which attempt to recapitulate the micro-scale structure and dynamics of the gut, culturing both microbial and epithelial cells together.

The macro-scale *in vitro* models include the M-SHIME model (Vermeiren et al., 2012) and the related MIMic model (see chapter 3 of this thesis). M-SHIME consists of up to

five chemostat-like vessels which represent the stomach, small intestine and the ascending, transverse and descending colon. The colon compartments contain mucin "microcosms", allowing mucus-associated bacteria to thrive by adhering to the mucus. This model is typically inoculated with faecal microbes, which are then allowed to multiply and colonise the model gut, as nutrients are added to the system, and fluid is pumped from one compartment to the next. This setup allows complex communities of microbes to be cultured over several weeks and the presence of mucus in the chemostats allows the maintenance of mucus-associated bacteria. This system can also be continuously sampled during experiments, unlike mice, which must be sacrificed in order to sample the various different sections of the gut. However, M-SHIME cannot simulate the interaction between the gut microbial community and the host, as no animal cells are cultured in the model, meaning nutrient absorption and metabolic activity – a crucial component of the system – cannot be modelled. Moreover, the immune system is absent.

In this respect, another *in vitro* model, HuMiX, represents a more complete system. This is a microfluidic model with separate compartments for medium, epithelial cells and microbes, connected by semi-permeable membranes (Shah et al., 2016). This model can also simulate the oxygen gradient in the gastrointestinal tract by delivering an aerobic growth medium in contact with the underside of the epithelial cells, whilst pumping nitrogen gas through another compartment connected to the top of the microbial compartment.

A drawback of both *in vivo* and *in vitro* models lies in measurement. Measurement of variables such as microbial population sizes relies on inexact or biased methods, such as plate counting or using cell counting machines. Both of these methods rely on taking small samples of a whole population and estimating cell numbers within those samples. Sources of uncertainty may include cell clumping, and cell death or loss during sampling, leading to uncertainty in results and analysis. Furthermore, the number of measurements that can be taken during an experiment is often limited, as taking measurements may be costly, have some impact on the system, or, as in the case of many *in vivo* models, require sacrificing the animal. In addition, only a limited number of variables can be feasibly measured. Thus, these models may provide only a limited amount of data.

1.5 - Mathematical Modelling

An alternative to *in vitro* and *in vivo* models are mathematical models – abstractions that use sets of equations and variables to model and explore physical systems. Mathematical models range from simple equations that describe a physical relationship such as Newton's 2nd Law, F = ma, to complex systems composed of many equations, such as the particle swarm model of Kim and Shin (2006). While some mathematical models can be evaluated using pen and paper, most complex models, especially those that include stochastic variables, are carried out *in silico*, that is, on a computer. *In silico*, like *in vitro* and *in vivo* models, are simplifications of the real systems they simulate. In order to model a complex biological system, simplifying assumptions must be made by the modeller and indeed this is the purpose of modelling. Modellers must determine what can be simplified or ignored without affecting the overall results. Differences between results from the simulation and those from the real system can then be used to assess the suitability of these assumptions and make changes. Thus, model development requires a cycle of improvement and refinement, as assumptions are tweaked in order to achieve better representation of the real system. For example, a model (Hart et al., 2019) of two cross-feeding microbial populations failed initially to recreate results observed in the laboratory. This prompted the researchers to take more specific measurements of the bacterial growth under different nutrient limitations, and to consider processes, such as evolution, that had previously been left out of the model, eventually leading to a model that emulated

the observations in the laboratory. This process of validation against experimental results is key in producing accurate and trustworthy models.

A key advantage of *in silico* models is in the amount of data they can produce. Unlike some *in vivo* or *in vitro* models, which only provide data at the endpoint of an experiment, *in silico* models can save and output their state at numerous intervals during an experiment. This data can be permanently saved, providing a complete time course of results to the modeller and allowing powerful statistical analyses and insights into temporal patterns and trends.

A variety of approaches have been taken over the years to the mathematical modelling of microbial populations. Different approaches have been used to answer different kinds of question. For example, Freter et al. (1983; 1984) modified the mathematics of continuous stirred tank reactor models to develop a differential-equation based model of microbial populations in the large intestine, incorporating competition for attachment sites in an effort to understand the relationship between resident and invading bacteria in the gut.

In silico modelling approaches for the gut are typically population- or community-level models which rely on differential equations describing the growth of different populations and the consumption or production of chemicals. For example, Muñoz-Tamayo et al. (2010) model a community of four microbial species in the colon carrying out fermentation of polysaccharides. The size of each population and of the concentrations of each chemical are governed by a system of differential equations.

Kettle et al. (2014) use a slightly more complex system of 10 bacterial functional groups, and a system of differential equations to model the population dynamics and fermentation of four different substrates in the colon.

One of the drawbacks of these differential equation-based models is that spatial heterogeneity within the gut microbiota is ignored. However, spatial heterogeneity on a small scale is likely to play an important role in the microbial populations of the gut. For example, the closer a microbial cell is to the epithelium, the more impact metabolites and antibodies released by epithelial cells will have on it. While microbes further away are more likely to be swept away into the lumen. The availability of attachment sites in the mucosa is, as shown by Freter et al. (1983), very important, making some spaces in the gut more "valuable" to microbes than others. Furthermore, due to limited mixing in the gut, as well as cell-cell and cell-mucus adhesion, microbial growth should lead to the development of clusters of related cells, creating the potential for complex mutualistic and antagonistic interactions between microbial species and strains.

Taking a different approach to these population-level, or top-down approaches to modelling are bottom-up approaches which seek to model microbial populations at the small scale in an attempt to understand how the behaviours, distributions and metabolism of microbes lead to larger scale phenomena. Some of the earliest models to take a bottom-up approach were biomass continuum models, which were used to model biofilms, and were able to provide insights into many of the spatial phenomena

observed in biofilms. For, while Wanner and Gujer (1986) used a 1-dimensional continuum model to investigate solute consumption, biomass distribution and interspecific interactions in mixed biofilms, while Dockery and Klapper (2001) were able to use a 2-dimensional continuum model to demonstrate how substrate limitation can lead to fingering instabilities in biofilms.

Another bottom-up approach is the cellular automaton model, which uses grid cells as a model for microbes in a biofilm. This has been used effectively in many biofilm modelling systems. For example, Pizarro et al. (2004) used a cellular automaton model to investigate the self-organisation of spatial structure by mixed biofilms.

1.6 – Agent-Based Modelling

Another group of bottom-up *in silico* models are the agent-based models. In an agent-based model, the individual members of a group are modelled discretely, each with a set of behaviours and properties, as opposed to the modelling large groups of individuals as single entities. Furthermore, rather than being simply biomass arrayed on a grid, individuals in agent-based models may have particular spatial characteristics, such as size, shape etc. and can move independently. These individuals are known as agents, and the outcomes of agent-based models arise simply from the actions of many such agents.

Agent-based models are particularly useful for modelling populations that are spatially heterogeneous, as opposed to well-mixed. In heterogeneous populations, members of the same species may experience different environments depending on local solute concentrations and the identity of neighbouring individuals, and thus behave in different ways, possibly adapting to their local conditions. Agent-based models are also particularly effective in situations where neighbouring cells may interact. For example, competing for a resource or exhibiting a cross-feeding relationship (Hellweger et al., 2016a). The use of agent-based modelling allows various phenomena, such as interactions between neighbouring groups or clusters of cells to be emergent, arising from more fundamental processes, rather than having to be specifically included in the definition of the model (Grimm, 1999; Kreft et al., 2001). For example, the interaction between two microbes, one with a high growth yield but

a low growth rate and the other with a low growth yield but a high growth rate is highly dependent on the spatial distribution of the microbes in relation to another, as demonstrated using agent-based modelling in Kreft (2004).

Agent-based models have been used successfully to study the population dynamics of nitrifying biofilms (Picioreanu et al., 2004) and the impact of cell morphology on spatial distribution in biofilms (Smith et al., 2017). Other microbiological systems, such as the *Anabaena* filament (Hellweger et al., 2016b), have also been the subject of agent-based modelling approaches.

An important drawback of agent-based modelling is also one of its key features. The focus on the small-scale detail at the level of the individual cell means that only small systems can be modelled at this full level of complexity. The size of a spatial system modelling individual microbial cells can typically only model spaces up to the scale of millimetres. In order to model larger systems, results must be extrapolated from the small-scale, or whole systems can be scaled down to sizes that fit the modelling capabilities (see section 2.4). Another solution to this problem is the use of superindividuals – agents which represent large numbers of individuals. Agent-based models can also be computationally demanding compared to differential equation-based models, and may require the use of high performance computers to aid in running simulations (see section 3.5).

Although the vigorous mixing and constant flow of material through the gut may suggest that the system is a well-mixed one, our current understanding suggests that

the mucosa – the layer of epithelial cells and secreted mucus lining the GI tract – harbours a distinct biofilm-like microbial population, which has been observed in mouse and human guts. However, as human gut biopsies are typically washed as part of their required preparation, there is still much that we do not know about the makeup and nature of the human mucosal biofilm (Duncan et al., 2021). Additionally, studies on the mammalian gut have shown clear spatial heterogeneity, both radially and longitudinally (Lu et al., 2014; Li et al., 2020). In addition, there is plenty of evidence of local interactions both between members of the microbiota, and between microbiota members and host epithelial cells. These include metabolic interactions as well as interactions between host and microbiota via the immune system (Sommer and Bäckhed, 2016).

Spatial heterogeneity and diversity of metabolic resources are thought to be important in maintaining diversity in microbial communities (Fierer and Lennon, 2011), and changes in the biogeography of the gut microbiota have been observed in the guts of inflammatory bowel disease and hepatic encephalopathy patients (Donaldson et al., 2016). However, the interspecific interactions that drive these changes are still not well understood.

The nature of the gut environment, and especially the mucosal biofilm therefore lend themselves naturally to agent-based modelling, which enables scientists to study the impact of small-scale interactions in a spatially heterogeneous environment.

Questions regarding how small-scale cell-cell interactions might drive population level phenomena in the gut, such as metabolic interactions, population dynamics and the maintenance of diversity are difficult to answer using animal models. Agent-based modelling can provide an alternative, that can be used to explore and understand how processes in a heterogeneous environment like the gut lead to the population-level phenomena we observe in humans and other animals.

In this thesis, I describe the development of a new modelling platform for the gut and gut microbiota, eGUT. eGUT utilises agent-based modelling, allowing the interactions between individual cells, as well as between mucosal and luminal microbial populations and the epithelium to be explored. In addition to a description of eGUT, I will detail the testing and validation of the platform and evaluate its potential as a modelling platform for gut research.

CHAPTER 2 - eGUT DESCRIPTION

This work was carried out in collaboration with Bastiaan Cockx, and benefitted from previous development by Robert Clegg, Kieran Alden, Stefan Lang and Sankalp Arya. All work, unless highlighted was carried out by myself, with testing, and feedback provided by B. Cockx, or was carried out in a joint manner. Where work was done primarily by B. Cockx, it is mentioned in the text.

2.1 Introduction

As explained in Chapter 1, a deeper understanding of the dynamics of microbial populations within the human gut and their relations to the health of host animals, especially humans, could enable scientists to better prevent and deal with diseases of the GIT, both acute and chronic. Currently, the majority of research that seeks a deeper understanding of the microbiota is carried out using *in vivo* models, particularly mice. With funding from the National Centre for Replacement, Reduction and Refinement (NC3Rs), my PhD project seeks to reduce and replace animal models in microbiome research by providing an alternative modelling platform for gut microbiome research.

In this chapter I will describe eGUT, an *in silico* agent-based, multi-compartment modelling platform for gut microbiome research. The development of eGUT has proceeded in tandem with the development of the biofilm focussed model iDynoMiCS 2.0 – an agent-based model that is designed to replace the original iDynoMiCS, which was developed in our research group from 2006 to 2011. iDynoMiCS 2.0 and eGUT share many functionalities, including the diffusion-reaction solver, the compartment system, the mechanical relaxation process, the information structure and the output functionalities. However, eGUT has several key capabilities in addition to those present in iDynoMiCS 2.0. Of these, the most important are a model of the epithelium and mucosa, including the agents that represent epithelial cells, mucus production, transfer reactions, which allow agent-mediated transfer across a surface and compartment-compartment solute transfer, which allows solute transfer between

compartments that are not connected by a well-mixed region. eGUT allows its users to simplify and generalise complex population dynamics, test hypotheses, and link the properties of individual microbial species to their effects in the gut.

For this description of the eGUT software, I will be utilising the "Overview, Design Concepts and Details" (ODD) protocol, developed by Grimm et al. (2006) as a guide for presenting the model in as clear and comprehensible a manner as possible, and updated in 2010 (Grimm et al., 2010) and 2020 (Grimm et al., 2020). I will also provide an example of a simulation in eGUT in order to demonstrate how its features work.

2.2 Development of eGUT

2.2.1 Overview

The eGUT software is a branch of iDynoMiCS 2.0. Most of the core functionality of eGUT is shared with iDynoMiCS 2.0, and changes within iDynoMiCS 2.0 are regularly merged into the eGUT code base. The eGUT software has been developed by building upon and adding to the code base of iDynoMiCS 2.0. In addition to the functionality of iDynoMiCS 2.0, eGUT adds new functions specific to modelling the gut, as well as a range of functionality for agent movement between compartments and new types of connections between compartments. The iDynoMiCS 2.0 code is therefore a subset of the eGUT code.

iDynoMiCS 2.0 was developed as a successor to the earlier iDynoMiCS (Lardon et al., 2011), taking inspiration from the structure of that modelling platform. The original iDynoMiCS model will henceforth be referred to as iDynoMiCS 1. The early development of iDynoMiCS 2.0 was chiefly carried out by Robert Clegg, Stefan Lang and Bastiaan Cockx. Since I began work on this project, iDynoMiCS 2.0 and eGUT. have been developed in tandem, and I have contributed to both platforms over the course of my PhD. As my work on iDynoMiCS 2.0 has been copied into the eGUT code base, I have presented this work as pertaining to eGUT in the interests of presenting the work as a coherent project. However, my work on iDynoMiCS 2.0 software contributes as much to that platform as it does to eGUT.

In this section, I will clarify my contributions to the iDynoMiCS 2.0 and eGUT platforms.

2.2.2 Work Common to iDynoMiCS 2.0 and eGUT

Much of my development work throughout my PhD project has been common to both iDynoMiCS 2.0 and eGUT. One of the key contributions I made was reworking the system by which agents move between compartments. Where the movement of agents had previously been coupled to boundary classes which also controlled solute behaviour, I developed a new system, where processes that move agents between compartments are controlled by two interfaces – arrival and departure. Any departure process can be coupled to any arrival process, making agent movement a modular and customisable system. I also decoupled the processes of arrival and departure from the boundaries controlling solute flow, which they had previously been a part of. Instead, agents moving from one compartment to another move to a special agent container called an arrivals lounge, which is completely separated from any processes of flow or diffusion.

To complement this new system, I also wrote a number of new processes for agent arrival and departure, including attachment of an agent to a biofilm by a random walk, arrival of an agent in a chemostat, departure of an agent moving through an

open spatial boundary, removal of agents a certain height above a solid surface, and removal of agents unconnected to a solid surface (connected volume filtration).

I also made several contributions to how eGUT and iDynoMiCS 2.0 handle agent biomass. There was an existing system of structured biomass, where agent biomass can be composed of multiple kinds of biomass, such as active and inert. However, there was no corresponding density system, by which users could define the density of the different biomass components. To introduce this, I wrote a new class for calculating volume from a user-defined list of densities for different mass types.

In addition, I updated the existing code that determined the shape of bacilli to ensure that EPS, when still part of a bacillus agent takes the shape of a capsule around the cell, rather than simply adding to the rod's length as other types of biomass do.

In a related piece of work, I updated the system for scaling density between 2D and 3D, so that the developments described above would continue to work in density-scaled agents.

Another key development in eGUT/iDynoMiCS-2 that I contributed to was the new agent spawning system. The need for this was spurred by the need to initialise a large confluent layer of epithelial cells for an epithelium, but the interface developed for spawning was quickly adapted for other uses. I contributed to the initial conceptualisation and development of this system, and its application for the epithelium.

I also contributed to the agent colouring system for model output, by which a colour palette of pre-defined colours is used to automatically assign colours to different agent types, and users can define quantitative variables to be encoded through colour.

In addition to developing the above features of the software, I also spent a lot of time carrying out bug fixing. For example, during my work on eGUT I found that there were issues with the chemostat ODE solver when multiple chemostats were connected. Investigating this, I found that the solver was not properly solving for inflowing solute from another compartment, which I was able to fix. I additionally fixed an EPS secretion system which was not functioning as designed.

Protocol testing and code review were also continuous processes, to ensure that new developments to the platform by myself and collaborators did not impact the existing functionality or accuracy of the platform.

A particularly extensive period of testing and debugging took place when we replaced the previous reaction-diffusion solver with the current FAS solver. In the early stages of implementing this algorithm in iDynoMiCS-2, there were unknown problems with convergence to a solution. In order to help tackle this, I developed a reporting system that would allow us to record solute concentrations at specified steps within the reaction-diffusion algorithm. This enabled us to plot the changing shape of the concentration gradient through time to assess the convergence properties of the solver. In order to visualise this data as well as possible, I also developed a Matlab script to plot the values in the solute concentration grid over time.

2.2.3 eGUT-Specific Code Development

The eGUT software incorporates all of the changes, updates and developments described above, but also includes work that pertains to eGUT alone.

The most central and important eGUT-specific development was the development of epithelial agents as a new agent type with properties relating to volume, mass, position and relation to a particular boundary of the spatial mucosa compartment. As mentioned above, this also required the development of a new spawning system to create a confluent layer of epithelial cells, as well a new system for determining the solute concentrations experienced by epithelial agents, which are often larger than other agent types and border multiple grid elements.

Another key part of developing the epithelium was developing a new subtype of reaction – the transfer reaction. These enable epithelial agents to interact metabolically with the compartment they border. This type of reaction can also have reactants and/or products that are in different compartments, or solely within epithelial agents. Rates of reaction are per unit surface area, rather than per unit volume as with other reactions.

Another key element of eGUT is the mucosal environment, of which a key component is the mucus. I modified the pre-existing EPS secretion system for microbial agents to the epithelium, and developed a new model of stochastic movement based on the

Stokes-Einstein equation in order to simulate random motion of agents, which can be used to simulate the fluid behaviour of mucus, as detailed in section 3.5.

2.2.4 Code Availability

The eGUT source code is available on Github at the following repository:

https://github.com/Secondus2/eGUT

All protocol files for the tests run in this thesis can be found in this repository.

The iDynoMiCS 2.0 source code can be found at the following repository:

https://github.com/kreft/iDynoMiCS-2

2.3 ODD

2.3.1 Overview

Purpose

eGUT enables specification of a large variety of different agent-based models of the gut. This provides a modelling platform for researchers interested in microbial population dynamics and microbiota-host interactions in the gut. It allows researchers to investigate how local interactions between neighbouring cells can affect population level dynamics in the gut.

State Variables and Scales

eGUT models typically contain two key types of entities: agents and solutes. Simulations centre around the behaviour of, and interactions between, agents and solutes. Agents can represent individual microbial cells, mucus particles and epithelial cells, while solutes represent the dissolved chemicals in the environment.

The setting in which agents and solutes interact is the compartment, and there can be several, connected compartments. Compartments define the space in which agents and solutes exist. They can be spatial, with a defined size and boundaries, or dimensionless, representing a simple well-mixed bulk compartment.

Agents can be characterised by their particular morphology, size, location, density, biomass composition and various behavioural characteristics. These characteristics are modular, independent and can thus be freely combined in many different ways. This is a fundamental difference to iDynoMiCS 1, where specific combinations of agent characteristics had to be constructed by a chain of subclasses, constraining possible combinations and not enabling changes at runtime. Thus, eGUT enables users to freely and more efficiently define their particular agent types. Agents can be placed in a compartment at the start of the simulation, with the user defining (potentially random) initial positions and characteristics of agents. However, new agents arise as microbial agents divide or produce EPS particle agents. Agents can also be removed from a compartment through any process that moves an agent into a different compartment or out of the simulation completely, such as death or detachment from the mucosal biofilm. One of the most important characteristics of agents is their metabolism. This is defined by a set of reactions which are the property of all agents belonging to a particular species. These reactions define how agents and solutes interact, specified by an equation for the rate of each reaction based on concentrations of solutes and/or biomass fractions, and providing stoichiometric yields of the reactants and products.

While the suite of agent characteristics available to users of eGUT is large, there are distinct differences between microbial cells, EPS particles and epithelial cells. For example, microbial cells and EPS particles can currently only have coccoid or rod

morphology, while epithelial cells can only have rectangular morphology. Furthermore, only epithelial cells, sitting at the boundary of two compartments, can carry out transfer reactions that transfer solutes from one compartment to another, but epithelial cells currently cannot participate in spatial growth (although their masses can change), movement or division, as they occupy fixed places in the epithelium.

Users define the characteristics and behaviour of solutes by giving initial concentration for each solute in each compartment and a diffusivity for spatial compartments which is used to model a solute's diffusion through space. Solutes can be replenished or removed at certain types of boundary, which can model a fixed solute concentration at a particular boundary, or the inflow/outflow, production or removal of solutes in the bulk liquid.

The simulated space in which agents and solutes reside and interact is the compartment. Compartments can be spatially defined, with either 2 or 3 spatial dimensions, or they can be dimensionless, in which case they simply represent a well-mixed volume. Spatial compartments are either rectangular or cuboidal, with side lengths determined by the user. Both spatial and dimensionless compartments can contain agents, although epithelial agents should only be used in spatial compartments. There are no defined limits on the size of spatial compartments, but the computational demand of simulating a spatial compartment scales with the number of simulated agents and thus with the compartment size. Thus, the practical limit for the side length of a spatial compartment is ~1 mm. Dimensionless

compartments, by comparison, are far less limited in size, as they do not require the computationally demanding simulation of cell collisions or the solving of PDEs to determine reaction rates. Thus, dimensionless compartments are constrained solely by the number of agents, and the memory requirements of storing and handling the properties of those agents. It is certainly quite possible to model dimensionless compartments containing millions of agents. In compartments with no or few agents, size is only limited by the limits of "doubles" in Java – that is, numbers that can be written using 64 bits of information. However, in interactions between compartments, such as the exchange of solutes through flow or diffusion, very large differences in scale may lead to the loss of precision in solutions.

A typical gut simulation consists of at least one spatial domain, representing the mucosa with its epithelial layer, connected to a simple well-mixed lumen compartment. The lumen is generally modelled as a dimensionless bulk compartment and is connected to the spatial mucosa compartment by a lumen boundary, allowing both solutes and agents to move between the two compartments.

Further compartments can be added to the simple two-compartment system. For example, if exchange between the body and the epithelium, or reactions within the body are of interest, a well-mixed compartment representing the dynamics of (some of) the body's organs can be connected to the epithelial face of the mucosal compartment via an epithelial boundary, allowing solute exchange via the epithelium. In addition, if the differences between gut sections along the axis of the GI tract are of

interest, further pairs of lumen and mucosa compartments can be added to the model, with directional flow from one lumen compartment to the next.

Process Overview and Scheduling

The processes that take place in an eGUT simulation are controlled by a timer. During a global timestep, various obligatory processes are carried out, including the spatial sorting of agents in spatial compartments and the creation of certain output files. However, the vast majority of processes that occur during a timestep are optional modules, which are declared by the user and may differ from compartment to compartment. The user also determines the length of a global timestep, and the length of timesteps for individual processes. These do not necessarily have to be the same, optional processes can have a longer or shorter timestep than the global one. User-defined processes within eGUT are carried out as follows:

1. Agent arrivals – agents that have been removed from one compartment in the previous timestep and are due to arrive in a new compartment are added to their new compartments

2. Other optional processes – Any other user-defined processes are carried out in the order defined by the user, which requires careful consideration. These include running the reaction-diffusion solver, mechanical relaxation of agents, production of summary files and output images etc.

3. Agent departures – Agents leaving compartments are removed and queued for entry into a new compartment in the next time step

These processes are carried out for each compartment, one by one. They are ordered in such a way that all agents in the entire multi-compartmental system are considered once and only once during one timestep, and agents moving from one compartment to another do not "miss" a timestep of growth, reaction, and other changes during their transfer to their destination compartment. Event-driven scheduling is not yet implemented.

2.3.2 Design Concepts

Agent-based simulations aim to provide insight into how the activities of individual agents lead to population-level outcomes, such as clustering, competition, altruism and plasmid dynamics (Hellweger et al., 2016a). eGUT seeks to use agent-based modelling of microbial communities and epithelial tissues to better understand and predict how population-level outcomes can arise from the activity of individual cells. In particular, differences between the metabolism and enzyme kinetics of different species tend to lead to interactions between bacterial groups, that may take the form of cross-feeding, competition or cheating and similar interactions occur with agents representing the epithelial tissue. Furthermore, parameters such as yield and efficiency in substrate utilisation can lead to unexpected outcomes in mixed populations, which would not be predicted by population-level models. For example,

under certain conditions, a microbial strain with a low growth rate can dominate a biofilm even whilst competing with another strain that has a higher growth rate, if the strain with the lower growth rate has a greater growth yield than its competitor. (Kreft, 2004).

A key addition for eGUT is that it allows users to explore these concepts in the setting of the gut microbiome, with agents comprising both the microbiota and the host cells interacting in the mucosal region. Other regions, such as the gut lumen (microbes only) and organs of the body (represented as mixed compartments via ODEs) connected to the mucosa can be incorporated into the model to assess how the interactions between gut microbes and their host lead to population-level outcomes, such as infection, dysbiosis and probiotic effects. The development of spatial structures, collectives or gradients within an eGUT simulation only derive from processes acting at the level of the individual and are not imposed on the system by top-down controls.

As eGUT models processes at the level of the individual cell, there are some processes at a finer level that are ignored. For example, the intracellular processes that determine the orientation of mother and daughter cells during a division event are not considered, and are instead assumed to be random. Various processes in eGUT are stochastic, and the Mersenne-Twister random number generator is used to determine how they are realised. eGUT users can provide a random number seed in a protocol file, which enables a simulation to be reproduced exactly.

Agents in eGUT have limited access to global information, just as microbes do in the real world. For example, agents can only interact with their neighbouring agents and their reaction rates are determined by their local solute concentrations, rather than any kind of global variable. Agent interactions are mechanical, chemical or informational. For mechanical interactions, agents are able to measure the degree of overlap with another agent that they have come into contact with, or a solid boundary and can produce repelling forces to reduce this overlap. They can also exhibit attractive forces that bring agents closer to each other, or to a solid boundary. For chemical interactions, agents interact with the solute grid via reactions which are typically modelled as Michaelis-Menten equations. The reaction rates that agents exhibit at the solute concentrations they experience are then used to solve the reaction-diffusion equations. For informational interactions, these may also be based on chemical signals where the above holds or consist of plasmid transfer.

A key assumption of this system is that agent growth and movement occur at a slower timescale to reactions and diffusion. This assumption allows eGUT to solve for the steady state of the reaction diffusion equation while assuming that agent positions and sizes are 'frozen' as they change more slowly and will have a negligible effect on that steady state. Growth and movement can then be applied as separate processes the effects of which will be considered in the next timestep of the steady state calculation. This is called timescale separation and is a powerful tool for simplifying

the modelling of a set of interacting processes that work at different timescales (Gunawardena, 2014).

In order to ask questions about a system as complex as the gut or even including other relevant organs of the body, simplifications should be made that reduce the complexity of the modelled system to a smaller number of variables. Thus typical eGUT simulations tend to feature fewer species, fewer solutes and fewer reactions than in the real gut, requiring less information to specify the system and validate output while allowing users to investigate the impact of changing particular parameters. For example, in the polysaccharide fermentation case study (Section 2.3), based on work by Muñoz-Tamayo et al. (2010), instead of modelling the full complexity of a polysaccharide fermenting community, key microbial guilds are modelled according to their metabolic behaviour and function. This allows a good deal of simplification of the system, while still representing all the key reactions that characterise the system.

eGUT's design makes it very easy to observe and collect information from simulations. Information in eGUT is in a tree structure and is input as an Extensible Markup Language (XML) file. The state of the simulation at every global timestep is output in the same format, meaning these output files can be reused as input files to restart a simulation from a particular time point. Thus there is a record of the simulation's state recorded at the end of every timestep. There are also various processes that can be specified by users to output further summaries, statistics and images to describe the

state of the simulation. These include POVray images, SVG images, bookkeeping spreadsheets and customisable summary statistics.

2.3.3 Details

Initialisation

An eGUT simulation is initialised using an XML file that contains information describing the compartments and the microbes that inhabit them, the global timestep, and length of the simulation, and any processes that should be carried out, optionally with their own timesteps. A random number seed can be specified to ensure reproducibility of simulations. Certain parameters, such as the base units used by eGUT, and the directory for results are stored in a separate config file containing defaults for these parameters that most users would not need to change.

Input

Typical inputs in eGUT will be the kinetics and stoichiometry of growth (reaction rates) in the form of ODEs derived from the literature or experimental work, cell shape and cell division threshold masses as well as rates of supply or removal of solutes or agents, and a variety of other physical and biochemical parameters dependent on the system and question studied. See section 2.4 for examples of such parameters.

Submodels

Compartments:

eGUT compartments come in one of two types. Dimensionless compartments are well-mixed and agents within them have no specified location. They are typically used as bulk compartments and connected to other compartments via boundaries. A bulk compartment may also be modelled just as a boundary condition. Spatial compartments can be two-dimensional or three-dimensional. They are always rectangular or cuboidal, with user-defined side lengths. As well as the sizes and shapes of compartments, users can define their spatial resolution, boundary conditions, connections to other compartments and diffusion layer characteristics. Figure 2.1 shows a typical combination of compartments that might be used in a gut model. The boundaries at the edges of a compartment have boundary conditions that define how solutes and agents interact with the edges of their compartment and/or with other compartments. In eGUT, spatial boundaries are simulated using a virtual layer of grid elements adjacent to each edge. The values of the concentrations in these virtual grid elements determine the behaviour of the diffusion-reaction solver, and are updated at each solver step where necessary to maintain the required effect. The types of solute boundaries used in eGUT's spatial compartments are as follows:

- Neumann boundaries – boundaries with a defined solute flux. In eGUT, these boundaries can be used to simulate a solid, impermeable barrier. For example, at the edge of a spatial domain. In this instance, these boundaries are called

"solid boundaries". Solid boundaries are modelled in eGUT by setting the concentrations in the virtual grid elements equal to their neighbouring real grid element, ensuring that no diffusion can occur between these grid elements.

- Dirichlet boundaries – boundaries with a fixed concentration. In eGUT, Dirichlet boundaries are used to simulate the diffusion of solutes from a well-mixed bulk volume into a diffusion-dominated region. In order for the diffusion-reaction to be soluble, at most one edge of a spatial domain can have a Dirichlet boundary. The solute concentration at the Dirichlet boundary could be fixed concentrations determined by the user to mimic a bulk compartment, or they could be defined by the concentration within a well-mixed bulk compartment linked to the spatial compartment. To model Dirichlet boundaries in eGUT, the solute concentrations of the virtual grid elements at the boundary are set to the bulk concentrations. The well-mixed grid elements within the spatial compartment will take on the concentration values at the boundary, while within the diffusion-dominated region, the diffusion-reaction PDE solver will determine solute concentrations.

- Periodic boundaries – Paired boundaries where the concentrations at either end of a dimension are treated as though they are adjacent, meaning that the dimension forms a virtual loop. In order to model this, the solute concentrations in the virtual grid elements at either end of the dimension are set to the values of the real grid elements at the opposite end of the dimension.

- Biological boundaries – Boundaries at which epithelial agents carry out transfer reactions. These reactions determine a rate of solute concentration change per unit surface area in the spatial compartment and within the epithelial agent. Unlike other boundary types, these do not utilise virtual grid elements to achieve the desired solute behaviour, but instead use transfer reactions to derive rates of concentration change in grid elements neighbouring the epithelium. The diffusion-reaction solver simultaneously solves the PDE treating the epithelium as a solid boundary.

In addition to the above rules on solute behaviour, boundaries also place rules and limits on agent behaviour – specifically on agent movement. The types of agent boundaries in spatially explicit eGUT compartments are as follows:

- Periodic boundaries – A given dimension is treated as a virtual loop, such that the dimension's two edges are considered to be equivalent positions. Any agent or part of an agent that moves beyond a dimension's minimum or maximum extent will be considered to have moved to the opposite edge of the dimension in terms of its co-ordinate position

- Solid boundaries – An edge of the domain is treated as a solid surface. Any agents moving past this surface experience forces that push them back as part of the agent relaxation process

- Open boundaries – Agents may pass through an edge of the domain, either moving into a different compartment or being removed from the simulation

These various boundary used in a mucosa compartment to model different physical structures. These are described in Table 2.1.

Table 2.1 – Types of boundaries in eGUT and how they combine solute boundary behaviour with agent boundary behaviour

Simulated structure	Boundary behaviour	
	Solutes	Agents
Lumen boundary	Dirichlet boundary with concentration determined by bulk solute concentrations in lumen	Open boundary allowing agent exchange between lumen and mucosa
Periodic boundary to simulate continuation of mucosa lengthways along gut wall	Periodic boundary	Periodic boundary
Epithelial surface	Biological boundary between epithelium and its mucosa	Solid boundary
Blood boundary	Biological boundary between epithelium and a separate compartment	Solid boundary

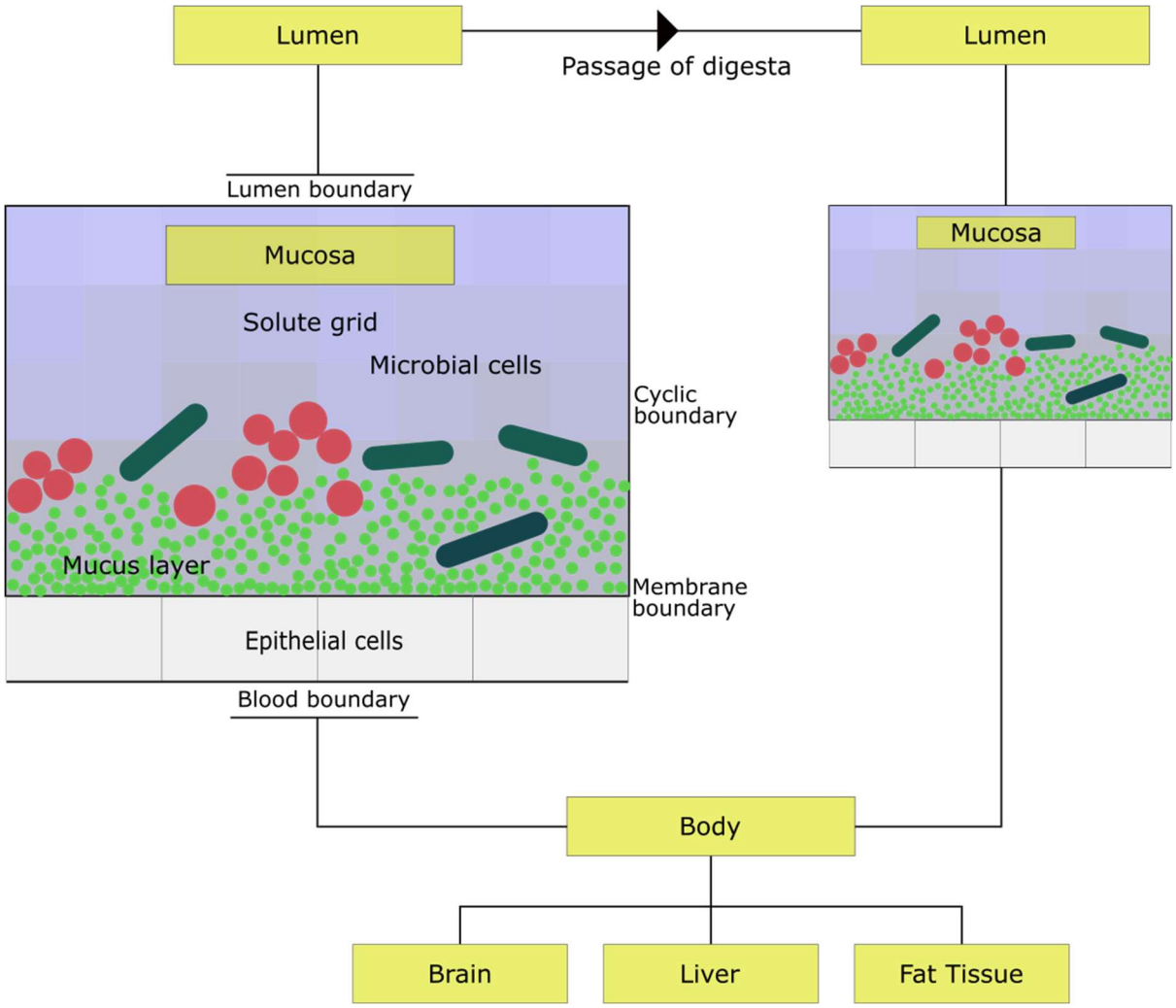

Figure 2.1 – Structure of compartments in a typical eGUT model. Spatial mucosa compartments connect to well-mixed lumen compartments, through which digesta flow from one gut section to the next. Mucosal compartments contain epithelia, which connect them to the well-mixed body compartment, which can have further connections to simple well-mixed compartments representing other organs.

Users can also define the thickness of diffusive boundary layers where well-mixed regions meet regions of space containing biomass, as well as the spatial resolution for the reaction-diffusion solver which will determine the size of the grid elements used by the solver.

Neighbour Searching (work primarily by B. Cockx):

Many of the agent interactions within eGUT require the program to search for neighbouring agents. For example, when searching for agents that may be colliding with one another, exert adherent forces on one another or close enough to exchange plasmids. A naïve neighbour searching algorithm, in which every agent is considered a possible neighbour of every other, would take a huge amount of time to run with large agent numbers, as for each agent, the distance to every other agent would have to be measured, meaning that the computational time would be proportional to $n(n-1)$, where n is the number of agents. In order to improve the efficiency of this spatial searching, agents can be sorted using a spatial registry which sorts agents into neighbourhood groups and tracks and reassigns them as they move.

There are two alternative modes of organising this spatial registry. One setup is the split tree. This behaves as a quadtree in 2D space, or as an octree in 3D space. The split tree splits a 2D or 3D compartment into four or eight, respectively, equal-sized parts by splitting the space in half in each dimension. Each of these regions can be recursively subdivided into the same number of smaller regions by performing the same operation again. The R-tree puts agents or groups of neighbouring agents into minimum bounding rectangles because it is quick to calculate intersection of rectangles. Several nearby agents (represented as rectangles) are recursively merged into larger bounding rectangles to form the hierarchical tree structure.

Diffusive and well-mixed regions

For modelling the mucosa, the typical setup is to model a spatial mucosa compartment, with an epithelium on one edge, and a lumen boundary connecting to a lumen compartment at the opposite edge. The other edges generally have periodic boundaries. The epithelium produces a solid boundary and a layer of mucus to which bacteria coming from the lumen may attach and grow, while the edge adjacent to the lumen is typically free of agents. The presence of a lumen boundary automatically triggers a spatial division within the mucosa compartment, which is divided into a well-mixed region and a diffusion-dominated (diffusive) region.

In the spatial grid, the diffusive region comprises of any grid elements containing agents, plus a diffusive boundary layer above the mucosal biofilm with a user-defined thickness. Diffusive boundary layers are regions of a medium or fluid adjacent to a surface in which the movement of solutes is dominated by diffusion, rather than by convection or turbulent flow. The thickness of such a region is dependent on the rate of flow above it, and the roughness of the adjacent surface (Jørgensen and Revsbech, 1985). In eGUT, this thickness value is determined by the user. The rest of the spatial domain is considered well-mixed and this region has the same solute concentrations as the lumen compartment connected to the lumen boundary. In the diffusive region, solute concentrations are determined by the diffusion-reaction steady state.

In order to determine which grid elements are part of the diffusive boundary layer, all grid elements that do not contain agents perform a neighbourhood search with a

search distance equal to the diffusive boundary layer thickness. If any agents are found within this search distance, the grid element becomes part of the diffusive boundary layer (see Figure 2.2).

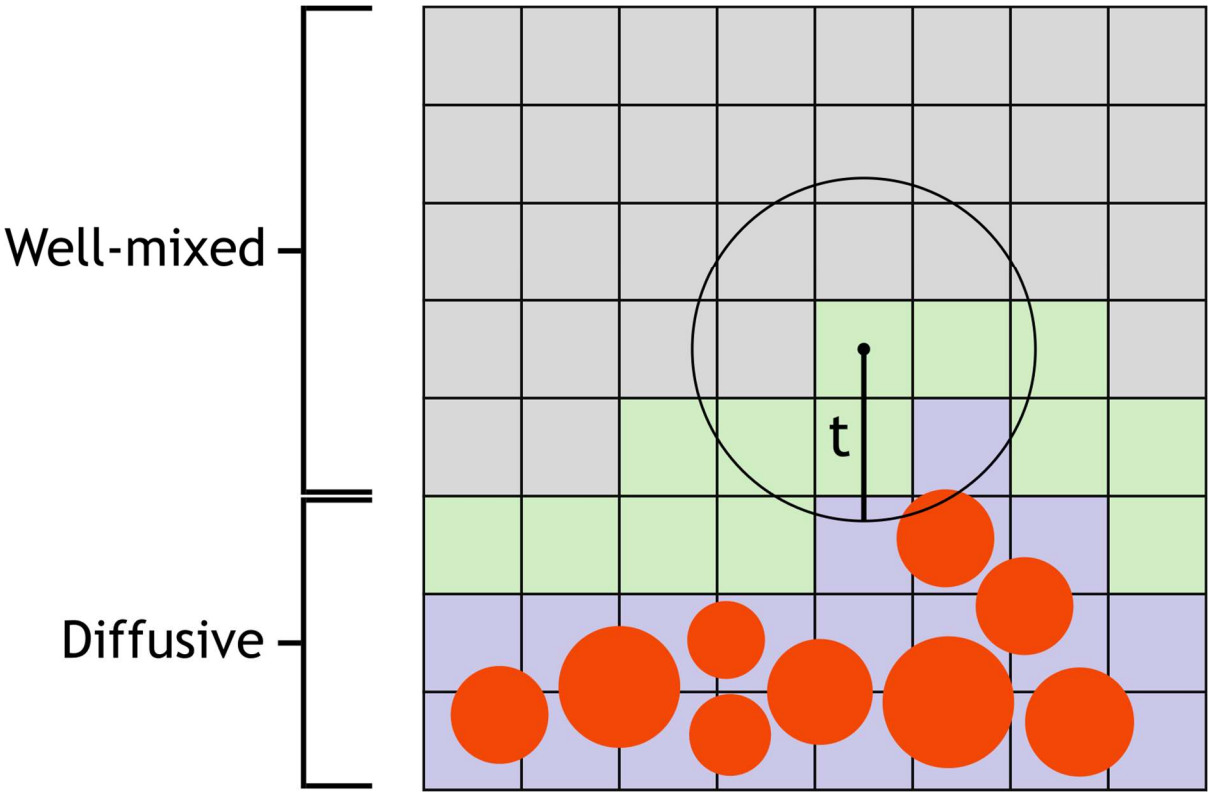

Figure 2.2 – Diagram showing how the behaviour of solutes in a spatial mucosal compartment is determined. The spatial compartment is split into a diffusive region, in which solutes move via diffusion, as determined by the reaction-diffusion solver, and the bulk, which is considered to be dominated by convective movement of solutes, in other words, being well-mixed. The concentration in the bulk region is set equal to the concentration in the connected well-mixed lumen compartment. Any regions in which agents are found is considered a part of the diffusive region. Specifically, any grid elements in which part of a cell is present, shown in blue in the diagram, are always considered part of the diffusive region. Above this is a diffusive boundary layer, with a thickness, t, which is set by the user. In order to determine which grid elements are considered part of the diffusive region, a rolling-ball

method is used. A circle (in 2D), or sphere (in 3D) is placed with its centre point at the centre of each grid element. If the circle or sphere collides with any agent, then that grid element is considered part of the diffusive boundary layer.

Reaction-Diffusion Solver (work primarily by B. Cockx):

The reaction-diffusion solver used in eGUT is a multigrid Full Approximation Storage (FAS) algorithm as described in Brandt (1977). The multigrid algorithm was chosen because computational effort scales linearly with the size of the system thus making 3D simulations feasible. The implementation of the algorithm is based on the implementation in Press et al. (1993) and Lardon et al. (2011). This solves the partial differential equation (PDE) for the steady state solute concentration fields that balance reaction with diffusion in a spatial compartment. It divides spatial compartments into grids of squares or cubes (depending on the number of dimensions). In each dimension, the grid element should have a side length that satisfies

$$g = \frac{l_x}{2^a} = \frac{l_y}{2^b} = \frac{l_z}{2^c}$$

Where g is the side-length of the grid element, l is the side-length of the compartment in the dimension given, and a, b, and c are integers. This ensures that the number of grid elements that fit into each dimension is an integer and that doubling the grid element side-length will either produce a grid element with the same properties, or one that spans the whole domain in at least one compartment. This is a useful property for the FAS algorithm which is solved at different resolutions differing from each

other by powers of 2. Users can set a spatial resolution for the finest grid of the reaction-diffusion solver that determines the scale at which agents interact with solutes. Note the resolution of the grid should be seen relative to the size of the agents, e.g., smaller agents require finer grids. Typically, for microbial agents, mass and reaction rates are distributed to grid elements based on the location of the agent's centre of mass. If more precision is required in distributing biomass to the grid, the "Collision mode" can be used. This uses the processes used for collision detection to identify which grid elements an agent overlaps with, and to obtain the magnitude of an agent's overlap with each grid element. From this, the proportion of the agent's biomass in each grid element can be calculated, and its contributions to reactions can be distributed according to those proportions.

Agents:

The agents in eGUT are split into two major groups. The microbial or EPS type agents are represented as solid spherical or rod-shaped particles. These agents must have a mass, density and morphology and when in a spatial compartment, each of their points, Px must have a location. Spherical cells have a single point, P, which defines their position, while rod cell have two points defining their position (Figure 2.1). The distance between these points is the rod 'length', l, and between these two points, the rod has a cylindrical or rectangular shape, in 3D and 2D, respectively. The shape of the rod at each of its ends is described by a semicircle or hemisphere around each of

its points, with a shared fixed radius, r. The radius of spherical cells, on the other hand, is calculated from their mass and density, while the same information in rod cells is used to define the rod's length (Figure 2.3). EPS particles are like spherical microbes but produced by microbes and lacking most of their characteristics.

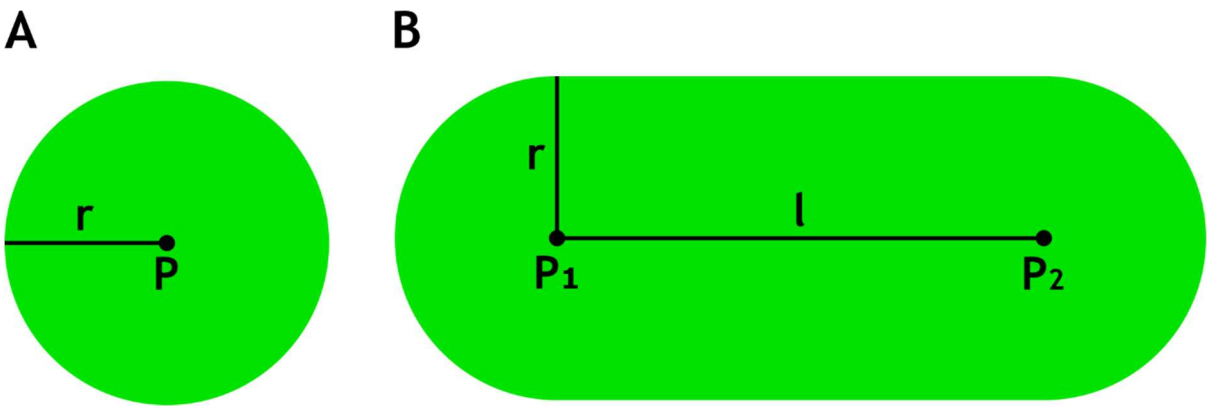

Figure 2.3 - The models for spherical and rod-shaped agents. (A) A spherical agent with a single point, P. The radius, r, is defined by the agent's mass and density. (B) A rod-shaped agent with two points, P1 and P2 and a fixed radius, r. The length, l, is defined by the agent's mass and density and grows with its mass.

Epithelial agents, on the other hand, have a fixed volume and position, being considered static at the boundary of the mucosal domain. Their positions are defined by two opposite corners at the edge of the mucosal compartment. In 2D, these corners form two ends of a line, while in 3D they form opposite corners of a rectangle (see Figure 2.4).

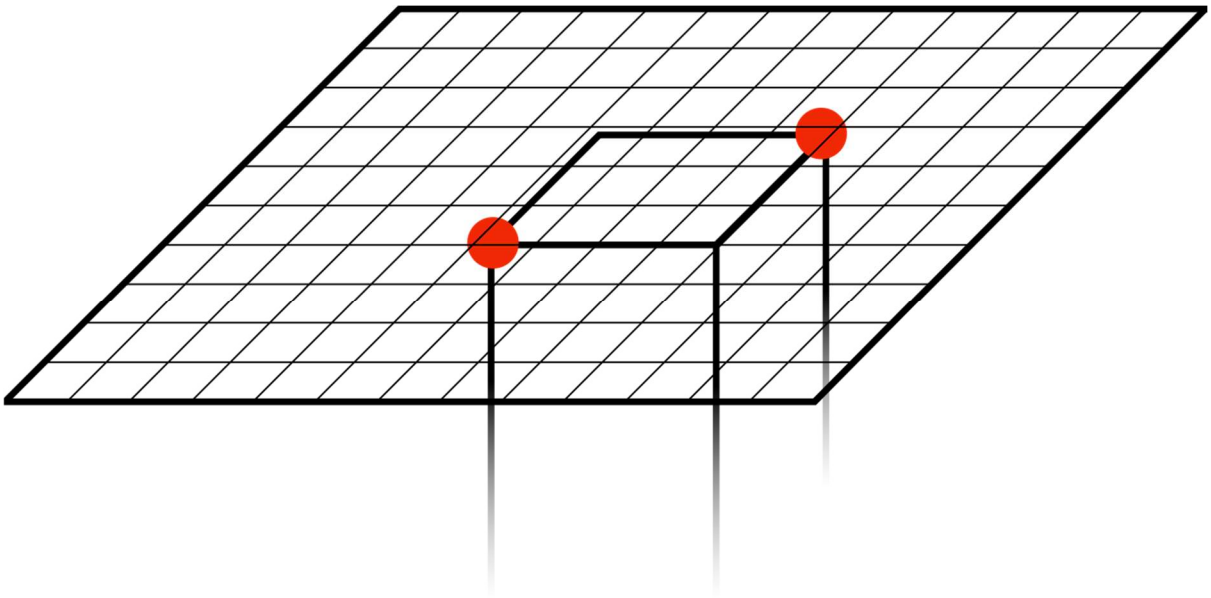

Figure 2.4 – A diagram showing how the shape of an epithelial agent is defined. An epithelial agent's position is defined by two corners (red circles) of the cell's apical face, that is, the face that forms the epithelial surface of the mucosa. All sides of the agent are considered to be parallel to the spatial grid of the mucosal compartment. Although epithelial agents do have mass and volume, their dimensions in the apical – basal plane are not explicitly modelled, as demonstrated by the fading lines.

An agent's biomass may be structured, meaning it can be composed of different kinds of biomass with different properties. For example, agents may contain both metabolically active and inert biomass, with reaction rates depending on the amount of metabolically active biomass, and inert biomass arising from the decay of active biomass. Other kinds of biomass , such as storage compounds, could also be modelled. Although the total mass of the epithelial cell is not explicitly linked to the agent's size, epithelial cells can store masses of particular compounds, in the same way that other agents store a structured biomass.

Agent Growth

As agents carry out reactions, they will typically gain or lose biomass. In response, their bodies require regular updating in order to reflect such changes. This updating works in slightly different ways depending on the cell shape.

For spherical agents (cocci), a change in total biomass entails a change in radius. The total biomass and the biomass density are used to calculate the volume and then the radius of a sphere in 3D or the cross-section area and then the radius of a cylinder in 2D.

In the case of rod-shaped agents (bacilli), a change in biomass entails a change in length, as radius (width) is fixed. Similarly, the total biomass and biomass density are used to calculate the volume for the rod, and the new length is then assigned to the agent.

Epithelial agents cannot change shape in the same way as other agents, but they can accumulate or loose biomass through internal reactions.

Scaling Agents Sizes in 2D

In order to simulate 2D models, a number of assumptions and adjustments have to be made. An implicit third dimension (z) is required to retain consistency for physical units such as volume or concentration, in eGUT this third dimension is 1 μm thick.

The 2D agent shapes are extruded into this virtual dimension, thus their pseudo 3D shapes have a uniform cross-sectional area and a thickness of 1 μm. This translation from 3D to 2D comes with several side effects such as a lower density of circle packing as compared to sphere packing (Clegg and Kreft, 2017). To mitigate this effect, Clegg et al. (2017) proposed a density scaling factor of 0.82 for 2D simulations with spherical agents. Appropriate scaling factors for other agent shapes or mixtures of agent shapes are unknown.

Another side effect of 2D simulations results from the constraint that the length of the virtual third dimension has to be identical for all agents. This can result in unwanted agent size effects for agents with very small or large radii. eGUT can scale the density of agents in order to retain consistent agent diameters and lengths between 3D and 2D simulations.

Due to the virtual third dimension of 1 μm in 2D simulations, cell radii and/or lengths can differ between 2D and 3D compartments. eGUT can scale agent densities, such that the dimensions of agents in 2D match what they would be in a 3D environment.

Users define an actual 3D density, ϱ_{3D}, which is then used to calculate a scaled density in the 2D compartment, ϱ_{2D}. The exact calculation depends on the shape of the agent in question.

For spherical (coccoid) agents, the radius of a sphere is calculated based on the agent's mass and actual density:

$$r_c = \sqrt[3]{\frac{3}{4\pi}\frac{m}{\rho_{3D}}}$$

Where r_c is the radius and m the agent's total mass. The scaled density is thus given by:

$$\rho_{2D} = \frac{m}{\pi {r_c}^2}$$

For agents or filament elements with a rod (bacillus) shape, it is the length, rather than the radius, that must be calculated. The 3D length of the line-segment connecting the agent's points is given by:

$$l = (\frac{m}{\rho_{3D}} - \frac{4}{3}\pi {r_c}^3)/\pi {r_c}^2$$

And the 2D scaled density is given by:

$$\rho_{2D} = m/(\pi {r_c}^2 + 2r_c l)$$

Agent Division

In order to determine when growing agents should divide, eGUT users can define a threshold mass for division. This is checked immediately after the application of agent growth. In the event of division of coccoid cells, a second coccoid cell is created with the same centre point as the mother, and a mass value is drawn from a normal distribution centred on half of the mother agent's mass, truncated at 2 standard deviations from the mean in either direction, with a coefficient of variance set to a

default of 0.05, but which can also be set by the user. This mass is transferred from the mother agent to the daughter agent and the two agents then have their bodies updated to their new size, a random direction vector is chosen and the two agents are shifted by half of the mother's new radius in opposite directions along that vector so that their overlap reduces to approximately 0 μm.

Agent Relaxation

In spatial domains, agents must push one another as they grow and divide in order to spread the growing biomass. For shrinking biomass, agents might pull each other together via attractive forces. In eGUT, this is achieved by applying forces to agents that are overlapping with other agents, such that the overlaps between agents are minimised using one of two methods. The first is the shoving method, which is a computationally fast method for moving overlapping agents away from each other. In this method, any overlapping agents are identified, and the direction and magnitude of their greatest overlap is calculated. These overlaps are summed across neighbouring agents and a force vector in the opposite direction of the summed overlap is calculated, with a magnitude equal to

$$\frac{2r(S-1)+L}{2}$$

Where r is the radius of the agent to which the force is being applied, S is the shove factor, a parameter that can be determined by the user and L is the shoving limit, the

maximum agent overlap allowed in the system. All the forces from the shoving method are summed before being applied to the agents.

The alternative method for agent relaxation is the mechanical Euler's method. The timestep used in this method is adaptive, but the total time spent on relaxation must not exceed the length of the global timestep. The adaptive timestep is based on the fastest moving object in the system, and is defined as the following:

$$T = \frac{\min\{\mathrm{G}\max\{\Delta x_1, \Delta x_2, \ldots\ldots, \Delta x_n\}, m\}}{\sqrt{|\max\{v_1, v_2, \ldots\ldots, v_3\}|} + \varepsilon}$$

Where T is the time step, Δx_A is the overlap of a given pair of agents, A, m is the maximum movement allowed in the relaxation process (a user-defined variable with a default of 0.01 μm), v_A is the velocity of agent A, and G is move granularity and ε is a safety value equal to 10^{-9} which ensures that very small denominators do not lead to extremely high timesteps, and to avoid a denominator of 0.

The force experienced by an agent is calculated as with the shove method by identifying any overlaps with other agents, and their direction and greatest magnitude normal to the agent surface. This overlap vector is then used to calculate a force, using either Hook's law for springs as the default 'push' function:

$$F = k\Delta x$$

Where F is the force, k is a scalar spring constant parameter set by the user, and Δx is the overlap distance.

As an alternative, Hertz's soft sphere model (1882) can be used:

$$F = \frac{4}{3}\sqrt{kr_E}\Delta x^{3/2}$$

Where k is a scalar parameter set by the user, r_E is the "effective radius" and Δx is the overlap distance.

The effective radius is equivalent to the radius of an agent if it is colliding with a plane, but in the case of collisions between agents, the effective radius is given as

$$r_E = \frac{r_1 r_2}{r_1 + r_2}$$

Where r_1 and r_2 are the radii of the two colliding agents.

All forces experienced by each agent are summed before being applied. Relaxation steps are applied until the greatest velocity of moving particles in the system has decreased to below a user-defined threshold.

Random Walk

The main process of random agent movement utilised in eGUT is a random walk based on the Stokes-Einstein equation of Brownian motion. The equation is

$$D = \frac{k_B T}{6\pi\eta r}$$

where D is the diffusion constant, k_B is Boltzmann's constant, $1.380649 \times 10^{-23}\ J\ K^{-1}$, T is the absolute temperature, η is the viscosity of the medium and r is the radius of a spherical particle moving in the medium. In the case of eGUT, the viscosity of the medium is assumed to be 8.9×10^{-4} Pa s, which users particularly interested in the effects and importance of viscosity may change. T is assumed to be 310 K, but can also be changed.

In order to simulate a random walk, the displacement of an agent over the course of a time-step is estimated using the equation derived by Einstein in 1905 (Ebeling, 2004) for the mean-square displacement of a particle undergoing a random walk:

$$\frac{\sum_{i=1}^{n} x^2}{n} = 2dD\Delta t$$

where x is the displacement of a particle, d is the number of dimensions, and Δt is the size of the time step.

A displacement vector is created with a magnitude (Euclidean norm) equal to the mean square displacement, but with a random direction. This movement is then applied during the agent relaxation.

Agent Transfer

Agents can move from one compartment to another in eGUT if the compartments are connected by a well-mixed boundary. For example, the boundary between the mucosa

and lumen compartments allows agents to freely attach to and detach from the mucosa, with unattached agents residing in the lumen compartment. The arrival and departure of agents into and out of spatial and non-spatial compartments is controlled by a set of process managers. One process of agent attachment for spatial compartments, for example, assigns an attachment likelihood to different cells in the spatial grid based on how close they are to the mucosal biofilm. Arriving agents are then assigned to different grid elements based on these probabilities. The arrival process manager for non-spatial compartments is simply registering the agent's presence in the compartment.

Departure processes can also be more complex and varied in spatial domains than non-spatial ones. For example, one option for departure from a spatial domain is the agent "scraper" which simply removes agents with any mass points above a set height in the domain. Another departure process is one that removes "floating" agents or groups of agents that are not connected to the substrate or base of a spatial compartment. This works by iterating through all agents and searching for neighbours within a user-defined search distance. This allows networks of neighbouring agents to be built up and combined such that all agents with a continuous chain of neighbours between them will be in the same network. Any networks that do not contain any agents in contact with the base or substrate of a spatial compartment are removed from the compartment if this departure process is on. A minimal option for agent departures is a process which removes any agents that have moved outside of the

compartment's computational domain, and provides the user with a warning, as this should not generally happen in a correctly set up simulation. This behaviour is always included by default within any other departure processes.

The overarching structure for all arrival and departure processes involves special agent lists called arrival and departure lounges. Every departure process requires at least one Destination. If a Destination is not specified by a user, it is assumed that agents removed by the process should be completely removed from the simulation. On the other hand, users can actually specify more than one Destination, allowing agents leaving one compartment to arrive in a randomly chosen eligible Destination. In this case, users must provide the Destinations in a list, along with the proportions of agents that should go to each Destination.

Compartments receiving new agents from other compartments must create arrivals lounges in which to store new arrivals. These arrivals lounges are tagged with the identity of the compartment which the arriving agents have come from. Thus, the transfer of agents between compartments can be easily tracked by users, as the arrivals lounges present at the end of each timestep will show exactly which agents are moving from which compartment to which.

Agent Spawning

Instead of defining each agent and its location at initialization, users may wish to use one of eGUT's built-in spawning systems, which allows the placement of a given number of agents within a given domain. Users can define one or more template agents along with a number representing how many copies of this template should be added to the compartment. Agents can be spawned in a random location within a given spatial domain drawn from a uniform distribution, or in a regularly spaced pattern. This is particularly useful for epithelial layers, which must not have gaps in order for the simulation to run, and therefore need cells that are regularly distributed.

Reactions

Reactions in eGUT can be agent-controlled, environmental or transfer reactions. Agent-controlled reactions take place in an agent, and typically involve some part of the agent's biomass, which may catalyse a reaction, take part in a reaction, or be produced by a reaction. Environmental reactions, on the other hand, take place in the solute grid and do not require the presence of an agent. These reactions only ever involve environmental solutes. Transfer reactions are carried out by epithelial agents and involve the movement of a solute across an epithelial boundary in a spatial compartment. Transfer reactions may include other processes, such as the conversion of a solute to a different form or to a type of biomass as it crosses this epithelial boundary.

In order to define a reaction in eGUT, users must provide an equation for the reaction rate, values for any constants in the equation and stoichiometries for all reactants and products. Typically the reaction rate will be in the form of an ordinary differential equation such as a Michaelis-Menten equation or a Haldane equation. These equations give a relationship between solute concentrations and reaction rates, while the stoichiometries determine the relationships between the products and reactants and the yield of the reaction. Reaction rates for regular reactions (agent-internal and environmental) should return a rate in terms of mass units per unit time per unit volume, while transfer reactions should return a rate in terms of mass units per unit time per unit surface area.

The reactions and resultant changes in biomass and substrate concentrations will be iteratively solved by the reaction-diffusion solver until the reaction rates and concentrations within the spatial grid reach a steady state. The reaction-diffusion obeys the principles of conservation of mass.

2.4 Case Study – Polysaccharide Fermentation in the Gut

In order to demonstrate the capabilities of eGUT, an example gut model was simulated, based on the mathematical model of polysaccharide fermentation in the human gut by Muñoz-Tamayo et al. (2010), which in turn was based on Anaerobic Digestion Model 1 (Batstone et al., 2002).

Undigestible polysaccharides are typically the main carbohydrate source for microbes residing in the human colon. These can include resistant starch, cellulose, pectins, xylan, inulin, and hemicelluloses. In addition, various polysaccharides and oligosaccharides may be released by the degradation of mucins by microbial activity in the mucosa (Wilson, 2008).

The breakdown via hydrolysis and subsequent fermentation of these molecules is carried out by a community of anaerobic gut microbes performing a variety of interdependent metabolic activities, driving both fierce competition for resources, and many cross-feeding dependencies (Cockburn and Koropatkin, 2016).

One approach to modelling the fermentation of colonic polysaccharides, as developed by Muñoz-Tamayo et al. (2010), is a cut down version of the Anaerobic Digestion Model 1 (ADM1), a generalised model of anaerobic digestion in bioreactors for resource recovery from waste (Batstone et al., 2002). The model lumps the many functions and processes in polysaccharide fermentation into the following four trophic guilds or "species":

1. Glucose-utilising bacteria, which can hydrolyse polysaccharides into glucose, and ferment glucose, producing lactate, hydrogen, CO_2 and the short-chain fatty acids acetate, propionate and butyrate.

2. Lactate-utilising bacteria, which ferment lactate, producing acetate, propionate, butyrate, hydrogen and CO_2.

3. Acetogens (homoacetogens), which form acetate by using hydrogen to reduce CO_2.

4. Methanogens, which form methane from hydrogen and CO_2. Acetoclastic methanogens are not included in this model as they grow too slowly compared to the residence time in the colon and therefore do not usually occur.

As this model is based on the ADM1, it assumes that the environment is anaerobic and most members of the four trophic groups are obligate anaerobes (Ragsdale and Pierce, 2008). Hydrogen is a key intermediate, produced by the first two groups and consumed by the last two groups as too high concentrations inhibit the hydrogen producers for thermodynamic reasons and too low concentrations reduce the growth rate of the hydrogen consumers, also for thermodynamic reasons (Schink, 1997).

In addition to the original model's four microbial species, an epithelial layer composed of colonocytes and goblet cells is modelled. Epithelial maintenance is modelled by simulating the consumption of butyrate by the epithelium. Colonocytes typically derive most of their energy from aerobic respiration of short-chain fatty acids. Specifically, butyrate has been shown to be the preferred substrate in rat colonocytes

(Clausen and Mortensen, 1994) and is considered similarly central to metabolism in human colonocytes (De Preter et al., 2011). Here, epithelial cells consume butyrate and, in the case of goblet cells, this butyrate consumption fuels mucus production. Epithelial cell mass is assumed to be constant, with butyrate consumption representing cell maintenance and replacement. In goblet cells, half of the biomass/energy yield from butyrate is assumed to be directed towards mucus production. This assumes implicitly that any lacking or surplus resources are exchanged with the blood stream.

2.4.1 Model Setup

In order to simulate a region of the gut that was computationally tractable, a 32 μm × 32 μm epithelial surface was simulated for each of the three regions of the colon (proximal, transverse and distal), along with a spatial domain 256 μm high. A "scraper" process was implemented, that removed biomass above 150 μm, simulating the abrasion from the mucosa caused by the shear forces of digesta moving through the lumen. This set the effective thickness of the biofilm region to 150 μm, a typical thickness for the combined inner and outer mucus layers of the colon inhabited by mucus-associated bacteria (Johansson et al., 2011). Connected to each of these spatial compartments was a dimensionless compartment, which provided a bulk source of solutes, and which agents could move into and out of in exchange with the mucosa. The proximal colon lumen receives inflow at a steady rate, containing polysaccharides

at a concentration of 16.667 g l^{-1}, representing the "average-fibre diet" of the original model. Volume flows from the proximal colon lumen to the transverse colon lumen at an equal rate, and likewise from the transverse colon lumen to the distal colon lumen and out of the distal colon lumen (representing excretion). Agents in lumen compartments are transported along with the volume flow, at a proportion equal to the proportion of volume leaving the compartment each time step.

Agent exchange between each lumen compartment and its corresponding mucosa compartment is governed by three processes. Floating agent removal removes agents in the mucosa compartment that do not have a neighbouring agent within a 2 μm radius via which a chain of neighbours can be established that connects the agent to the epithelial surface. Removed agents move into the lumen compartment. The scraper process removes any agents within the mucosa compartment with a central point that is more than 150 μm above the epithelial surface. Once again, these agents are moved into the lumen. Finally, an attachment process leads to agents leaving the lumen and attaching to the mucosal biofilm at a rate of 0.01 min^{-1}.

Table 2.2 – Physical parameters of the gut model. Characteristics of lumen compartments and mucosal compartments apply to all three, which are physically identical

Physical property	Value	Source
Compartments		
Mucosal surface area	1,024 μm^2	Sadahiro et al. (1992)
Lumen volume	1.037×10^7 μm^3	
Flow rate	1.545×10^7 μm^3 day^{-1}	Muñoz-Tamayo et al. (2010)
Maximum height	150 μm	Johansson et al. (2011)
Diffusive boundary layer thickness	6 μm	
Microbial agents		
Density (dry mass)	0.1 pg μm^{-3}	
Division mass (dry mass)	1 pg	
Shape	Coccoid	
Mucus agents		
Density (dry mass)	0.01 pg μm^{-3}	Hill et al. (2022)
Average mass at excretion	0.05 pg	
Shape	Coccoid	
Epithelial agents		
Density (dry mass)	0.210 pg μm^{-3}	Brozek et al. (1963)
Dimensions	8 μm × 8 μm × 16 μm	Ovalle and Nahirney (2021)

In order to simulate the *in vivo* ratio between lumen volume and mucosal surface area, the total surface area of the colon was calculated based on the measurements made by Sadahiro et al. (1992), with the simplifying assumption of a constant cross-sectional radius of 20.25 mm through the entire lumen. This is the average radius of the lumen, weighted by the lengths of the different sections according to Sadahiro et al. (1992). This assumption was made because eGUT cannot simulate water loss from the digesta and the resultant changes in volumes and flow rates from one compartment to the next. Instead, a constant radius and constant flow rate are simulated. The flow rate was calculated by taking the proportional flow rate in the original model, considering the entire volume of the system (0.4967 day^{-1}), and applying the proportional flow rate to the total volume represented by the sum of eGUT compartments.

In order to simulate the degradation of mucins and the subsequent release of polysaccharides and oligosaccharides, a decay reaction was added to mucus particles, by which mucus degrades into an equal mass of polysaccharides, which is then available to microbes. In order to simplify the model, water and the short-chain fatty acids propionate and acetate, were removed from the reaction matrix, given that they are not reactants for any other reactions, but simply fermentation products. Another simplification was the lumping of polysaccharide hydrolysis with glucose utilisation. Both of these are carried out by the same species in the original model and keeping the reactions separate leads to instability in the reaction-diffusion solver.

The simulation was run for a total of 5 simulated days, with time steps of 1 hour. The chemostat solver and the spatial diffusion-reaction solver were run twice per global time step, with a time step of 30 minutes. An initial population of 60 of each microbial species, with initial cell masses of 10pg were placed in each mucosa compartment to begin the simulation. The epithelium, which was composed of 16 cells, each with an 8μm × 8μm surface, was allocated 11 colonocytes, and 5 goblet cells with a random distribution.

The kinetics and stoichiometry for the various microbial metabolic reactions were taken from the original model and values were adapted to masses in grams, rather than moles. Additional reactions for epithelial maintenance, goblet cell mucus production and mucus decay were also added to the model, based on values from experiments on both human and rat colonocytes metabolising radioactively labelled butyrate in an aerobic environment (Clausen and Mortensen, 1994; De Preter et al., 2011). A full breakdown of the reactions is shown in Table 2.3.

Table 2.3 – Petersen matrix (kinetic and stoichiometric matrix) for all reactions in the gut model, adapted from Muñoz-Tamayo et al. (2010). z represents the concentration of polysaccharides and s_a represents the concentration of solute a. $k_{m,b}$ values represent the maximum specific growth rate for growth on substrate b, while k_d and k_{md} represent decay constants of bacterial mass and mucus, respectively. x_c represents the biomass of species c. The same subscripts are used for both a particular substrate and the microbe that utilises that substrate. For example, s_{la} is the concentration of lactate, while x_{la} is the biomass of lactate-utilising bacteria. As methanogens and acetogens both consume hydrogen, the subscripts H_2m and H_2a are used for these two species, respectively. The IpH value represents pH dependent inhibition of methanogenesis. Meanings of subscripts are as follows: la – lactate, H_2 – hydrogen, bu – butyrate, CH_4 – methane, CO_2 – carbon dioxide, z – polysaccharide, H_2a – acetogen, H_2m – methanogen, mu – mucus, e – epithelium. For values of the parameters used here, see Table 2.4.

	Stoichiometry											**Kinetic Rate**
Process	**z**	**s_{la}**	**s_{H2}**	**s_{bu}**	**s_{CH4}**	**s_{CO2}**	**x_z**	**x_{la}**	**x_{H2a}**	**x_{H2m}**	**x_{mu}**	
Polysaccharide hydrolysis & glucose utilisation	-1	$Y_{la,z}$	$Y_{H2,z}$	$Y_{bu,z}$		$Y_{CO2,z}$	Y_z					$k_{m,z}\frac{zx_z}{K_{s,z}+z}$
Lactate utilisation		-1	$Y_{H2,la}$	$Y_{bu,la}$		$Y_{CO2,la}$		Y_{la}				$k_{m,la}\frac{s_{la}x_{la}}{K_{s,la}+s_{la}}$
Homoacetogenesis			-1			$-Y_{CO2,H2a}$			Y_{H2a}			$k_{m,H2a}\frac{s_{H2a}x_{H2a}}{K_{s,H2a}+s_{H2a}}$
Methanogenesis			-1		$Y_{CH4,H2m}$	$-Y_{CO2,H2m}$				Y_{H2m}		$k_{m,H2m}\frac{s_{H2m}x_{H2m}}{K_{s,H2m}+s_{H2m}}I_{pH}$
Epithelial maintenance				$-Y_{bu}$		1						$k_{m,bu}\frac{s_{bu}x_e}{K_{s,bu}+s_{bu}}$
Epithelial mucus production				$-Y_{bu}$		1					$Y_{mu,bu}$	$k_{m,bu}\frac{s_{bu}x_e}{K_{s,bu}+s_{bu}}$
Decay of microbial agents							-1	-1	-1	-1		$k_d x$
Decay of mucus agents											-1	$k_{md}x_{mu}$

Table 2.4 – Values of the parameters used in the gut model. Values taken from Muñoz-Tamayo et al. (2010) are converted from a molar system to an SI system.

Parameter	Value g g^{-1}	Source
$Y_{la,z}$	0.2247498	Muñoz-Tamayo et al. (2010)
$Y_{H2,z}$	0.0145152	
$Y_{bu,z}$	0.118935	
$Y_{CO2,z}$	0.242055	
Y_{z}	0.067797	
Y_{la}	0.150533	
$Y_{H2,la}$	0.008952	
$Y_{bu,la}$	0.195604	
$Y_{CO2,la}$	0.26895	
$Y_{CO2,H2a}$	10.91518	
Y_{H2a}	2.410218	
$Y_{CH4,H2m}$	0.755853	
$Y_{CO2,H2m}$	9.823661	
Y_{H2m}	3.475198	
Y_{bu}	0.50054	(Allan et al., 1996)
$Y_{mu,bu}$	0.03267	(Kilburn et al., 1969)
$k_{m,z}$	10.61947 day^{-1}	Muñoz-Tamayo et al. (2010)
$k_{m,la}$	82.10832 day^{-1}	
$k_{m,H2a}$	1.941729 day^{-1}	
$k_{m,H2m}$	0.40286 day^{-1}	

$k_{m,bu}$	0.0648 day^{-1}	(Clausen and Mortensen, 1994; De Preter et al., 2011)
k_d	0.01 day^{-1}	Muñoz-Tamayo et al. (2010)
k_{md}	0.02 day^{-1}	
$K_{s,z}$	0.01 g l^{-1}	Muñoz-Tamayo et al. (2010)
$K_{s,la}$	0.59687 g l^{-1}	
$K_{s,H2a}$	3.4×10^{-4} g l^{-1}	
$K_{s,H2m}$	3.13×10^{-6} g l^{-1}	
$K_{s,bu}$	2.87×10^{-4} g l^{-1}	(Clausen and Mortensen, 1994; De Preter et al., 2011)
I_{pH} (proximal colon)	0.00483	Muñoz-Tamayo et al. (2010)
I_{pH} (transverse colon)	0.39616	
I_{pH} (distal colon)	1.0	

2.4.2 Case Study - Results

After five simulated days, there are clear trends in the results, such as the differences in the mass of different species in different compartments. The methanogens are far less abundant than any of the other species (Figures 2.5, 2.6), which matches the results from the original model, and while glucose-utilising bacteria are the most abundant species in the proximal colon, acetogens are the most abundant in the transverse and distal colons (Figure 2.5). Furthermore, the total biomass in the proximal colon is far lower than the total biomass in either of the other colon sections (Figure 2.6). This is likely due to the fact that products produced by the glucose-utilising bacteria, including hydrogen, the limiting substrate in acetogenesis, are carried by the flow of digesta from the proximal colon lumen to the other compartments, while the proximal colon receives no supply of hydrogen. Indeed, the concentrations of hydrogen in the transverse and distal colons are an order of magnitude higher than the concentration in the proximal colon (Figure 2.7). Another factor that may drive the differences in species composition between the compartments is the polysaccharide concentration, which differs dramatically between the three compartments (Figure 2.8).

One of the most striking outcomes is the amount of mucus in the mucosa. This is very low, and does not appear to fully cover the epithelial layer. This may be because of the fact that the model considers butyrate metabolism to be the only source of energy for the goblet cells, and thus may neglect energy sources provided by the bloodstream which would contribute towards mucus production. Considering supply by the

bloodstream may lead to a better match between the simulated results, and those observed in the gut, where mucus constitutes the bulk of the 150 μm mucus layer (Johansson et al., 2011).

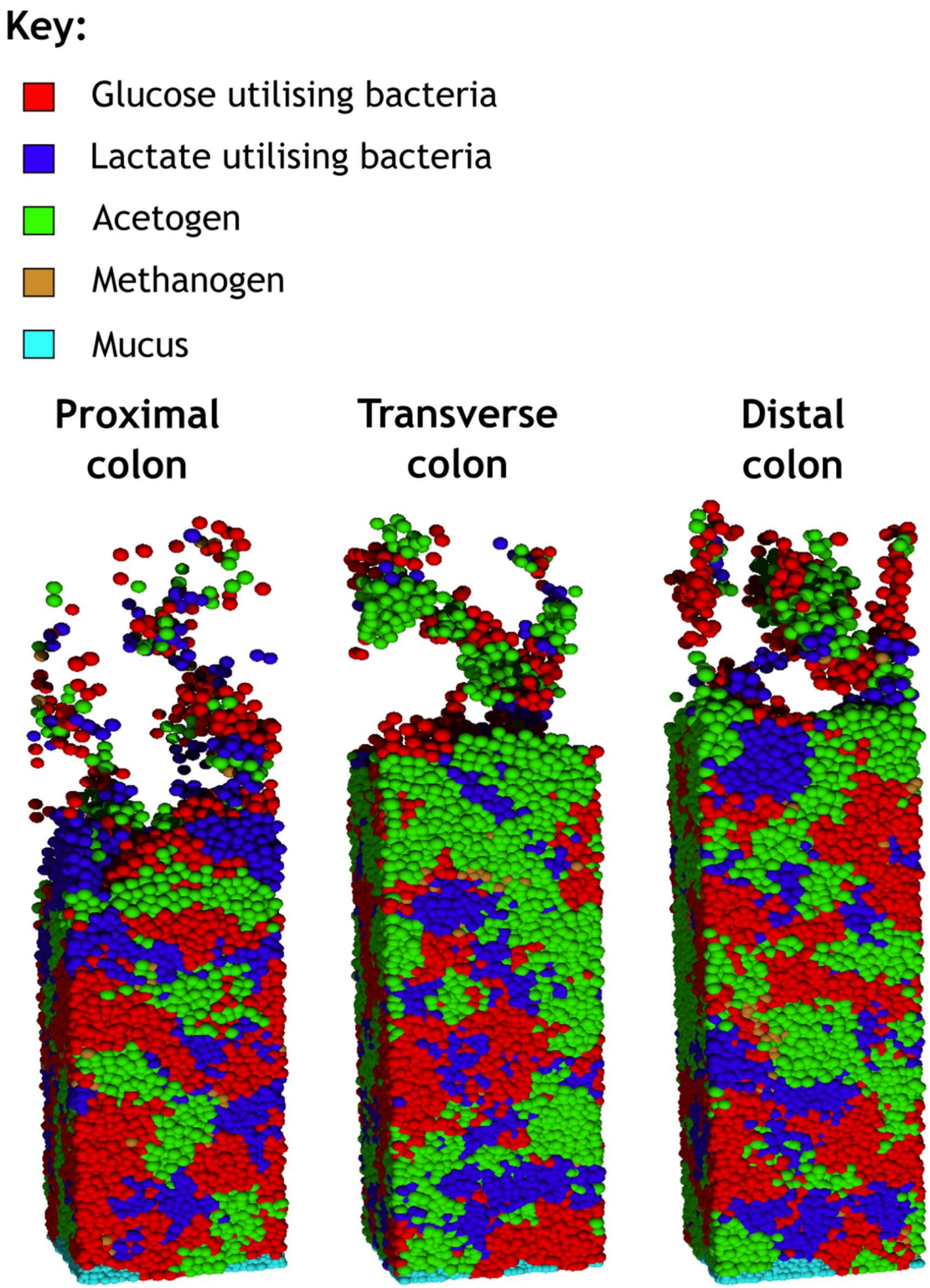

Figure 2.5 – Visual output from the gut model showing the mucosa compartments for each of the three colon sections modelled.

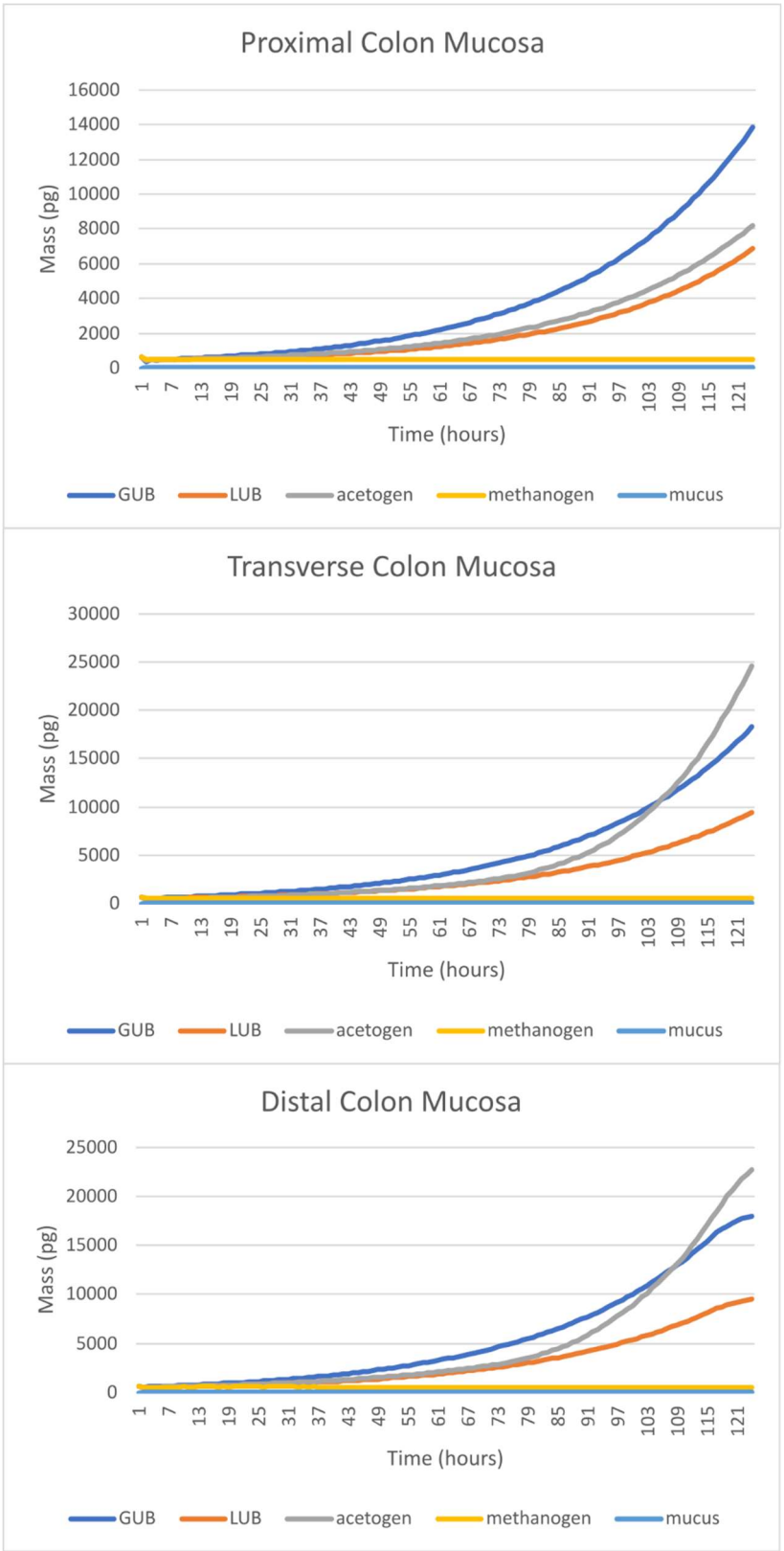

Figure 2.6 - Biomass of the different microbial species and the mucus in the mucosa compartments over time.

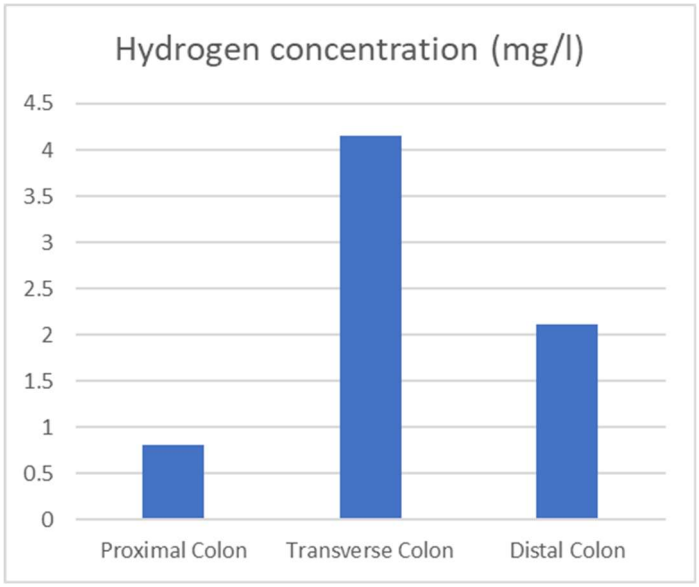

Figure 2.7 - Bulk (luminal) hydrogen concentrations in the proximal, transverse and distal colon after five simulated days.

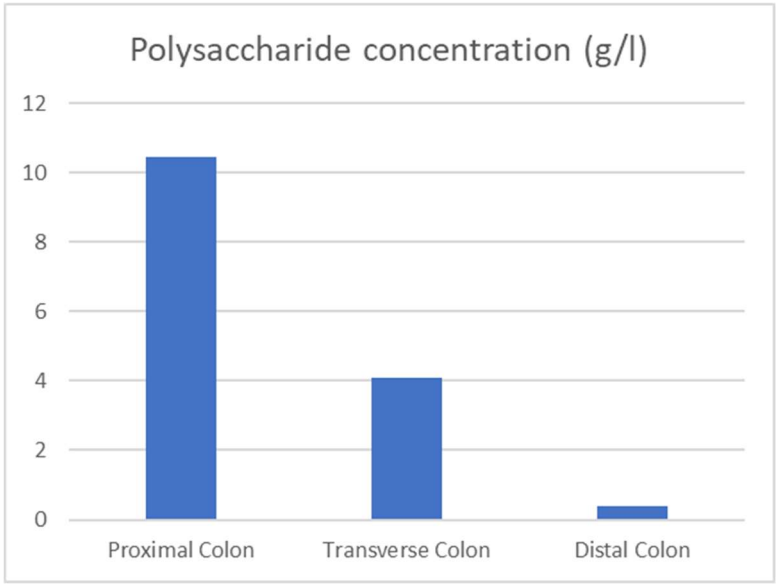

Figure 2.8 - Bulk (luminal) polysaccharide concentrations in the proximal, transverse and distal colon after five simulated days.

2.4.3 Case Study - Discussion

This case study shows that eGUT is able to successfully simulate a complex, multi compartmental model of the gut, providing users with comprehensive time courses, solute concentration profiles and high-quality image outputs. The patterns that emerge from the model provide avenues for further research, and data that can be compared to other models and measurements. Further work on this particular case study would be valuable, providing a more rigorous comparison with the original model by Muñoz-Tamayo et al. (2010) as well as empirical data on biomass and substrate concentrations. Moreover, eGUT provides a framework for investigating the amount of goblet cell metabolism required to produce a mucus layer as thick as is expected in a gut model.

2.5 Related Agent-based Models or Platforms

There are a wide variety of existing agent-based modelling platforms already used in microbial research. The simplest of these are general agent-based modelling platforms such as NetLogo (Wilensky, 2001), which provides a simple 2 dimensional world, divided into a grid of "patches" within which agents (known as "turtles") can move around, behave and interact. This simple modelling platform is applicable to a wide range of potential applications but cannot be used to simulate something with a complex environment such as the gut microbiome without significant additional work. One model that does attempt to utilise NetLogo to model the gut is GutLogo (Lin et al., 2018), a simple 2 dimensional gut model that simulates the ileum surface with microbial agents growing and dispersing along the surface, and a constant flow field representing the flow of digesta through the ileum and carrying metabolites through the domain.

Another key group of agent-based microbial models are the various synthetic biology models and modelling platforms, which seek to discover how particularly interesting or useful community behaviours or phenomena can be achieved through the creation of synthetic microbial communities. These include models such as BSim 2.0 (Matyjaszkiewicz et al., 2017), gro (Gutiérrez et al., 2017) and CellModeller (Rudge et al., 2012). All of these models simulate gene regulatory networks and diffusible signal molecules in order to explore and/or design synthetic microbial communities.

The other major group of microbial agent-based models are primarily biofilm models, which seek to explore population dynamics and ecological and evolutionary concepts in biofilms. These include iDynoMiCS 1 (Lardon et al., 2011), NUFEB (Li et al., 2019), Simbiotics (Naylor et al., 2017) and BacArena (Bauer et al., 2017). These models tend to focus on microbial metabolism and diffusion through the biofilm in order to assess how different growth strategies may benefit species growing in a biofilm.

2.5.1 Agents

The complexity of agents can vary substantially between these different agent-based models. For example, in BacArena and GutLogo, agents are 'points' on a grid so do not have a physical size or shape. In BacArena, agent reproduction causes daughter cells to be placed in neighbouring grid elements, which can in turn displace agents residing there, which are then moved again to a neighbouring grid element, and so on. In GutLogo, daughter cells are placed in the same grid element as the mother cell, but are liable to be carried downstream by the flow field. In the other models, agents are represented with a physical model and increase in size as they grow, pushing on other cells through mechanical interactions. In the cases of iDynoMiCS 1 and NUFEB, microbial agents are modelled as spheres, while in the case of gro, BSim 2.0 and CellModeller, microbial agents are modelled as rods. Meanwhile, Simbiotics and eGUT have both rods and sphere models available. Adhesive forces can also be modelled in addition to pushing forces in eGUT, NUFEB and Simbiotics.

Another way in which these models differ from one another is in the complexity and application of growth kinetics. One of the simplest models in this regard is GutLogo, which models agents with an "energy" which ranges from 0 to 100. Agents with energy below 80 are classed as "hungry" and are able to consume metabolites. Agents with an energy greater than 50 have a chance of reproducing, in which case the energy will be evenly divided between the mother and daughter cells.

Other models, such as CellModeller, allow users to set a fixed relative growth rate for each species in the model, leading to exponential growth. BSim 2.0 uses a growth model based on population growth models, leading to growth that slows as an agent approaches its maximum length, but which leads to exponential growth for a whole population.

In other models, including eGUT, iDynoMiCS 1, gro, NUFEB, Simbiotics and BacArena, users can write ordinary differential equations that define a relationship between substrate concentrations and a species' growth rate, such as Monod kinetics. Fixed relative growth rates are of course still possible in these models. NUFEB also has an energy-based growth model based on modelling the Gibbs energy of catabolism, and BacArena has a flux balance analysis (FBA)-based growth model, which enables the modelling of a metabolic network, with rates determined by a given objective function (Orth et al., 2010).

In most models, growth simply means an increase in biomass. However, some models have structured biomass, meaning that different kinds of biomass can be produced.

For example, eGUT, Simbiotics and NUFEB all have agents that can produce and secrete extracellular polymeric substances as small particles. eGUT can also have other types of biomass within an agent. For example, decay processes can cause active biomass to become inert biomass.

2.5.2 Simulated Domains

Most of the models discussed here have the ability to model domains from micrometre scales up to millimetres in some cases (e.g. NUFEB). In some models, these simulated spaces can only be 2D (gro, BacArena, GutLogo). In others, they must be 3D (NUFEB, BSim 2.0) while in others, both options are available (eGUT, iDynoMiCS 1, Simbiotics, CellModeller). In most models, diffusion happens within the domain, but in eGUT, iDynoMiCS 1 and NUFEB, the connection of a spatial compartment to a well-mixed bulk can be explicitly modelled. In this case, a diffusive boundary layer is created. This is a region above the biofilm region in which diffusion dominates the movement of solutes. Above this is a well-mixed region which shares its solute concentrations with the bulk. NUFEB can also model fluid dynamics and advection, where fluid flow can affect mass transfer and biofilm structure but biofilm structure not fluid flow.

2.5.3 Chemistry

In all of these models, chemicals are represented as concentration fields on discretised grids. These grid elements are cuboidal, apart from BSim 2.0, where chemical grid elements can be triangular. In almost all models, diffusion and reactions can be modelled using a finite volume method partial differential equation solver. In GutLogo however, it is unclear whether diffusion is modelled. In NUFEB, a further level of detail can be added by modelling fluid flow and the resultant advection. CellModeller can also implement a simple advection model, which applies a linear bulk flow in a given direction. Furthermore, NUFEB is the only one of these models that can simulate pH dynamics and that can model gas and liquid phases, allowing gas-liquid transfer to be simulated.

2.5.4 Summary

Agent-based models can be applied to a wide range of research questions. Within the field of microbiology, agent-based models and modelling platforms help scientists understand the dynamics of non-homogeneous populations and how small-scale spatial phenomena can lead to higher level emergent phenomena. For scientists studying the potential applications of engineering bacterial signalling pathways and behaviours, a variety of suitable models exist, including BSim 2.0, gro and CellModeller. For scientists interested in the population dynamics of biofilms, several

platforms are available, including iDynoMiCS 1, NUFEB, Simbiotics and BacArena. However, the application of agent-based modelling to the setting of the gut is something few have so far tackled. The GutLogo model is the first to attempt using agent-based modelling in the gut setting, but the NetLogo system it is based on does not allow a good representation of the physical environment of the gut, and seriously limits the processes that can be modelled. eGUT provides a platform tailored specifically for gut modelling, with a better representation of the physical characteristics and processes of the gut, and representing both the biofilm-like surface-associated mucosal microbes and the more well-mixed system of the lumen contents.

Still, some features lacking in eGUT are available in other modelling platforms. For example, BacArena allows users to model metabolism using flux-balance analysis, and NUFEB includes additional physical processes, with users able to simulate fluid dynamics and pH dynamics. The selection of an appropriate model should be based primarily on the research question and how the features of a modelling platform can be applied to simulate the processes and features of interest to the modeller. A summary of the comparison of all the models mentioned here is given in Table 2.5.

Table 2.5 – Summary of the features of the various agent-based models mentioned in the comparison above

Feature	eGUT	iDynoMiCS 1	NUFEB	Simbiotics	gro	BSim 2.0	BacArena	CellModeller	GutLogo
Can model microbial growth using ordinary differential equations	✓	✓	✓	✓	✓		✓		
Rod-shaped agents	✓			✓	✓	✓		✓	
Spherical agents	✓	✓	✓	✓					
2-dimensional spatial domains	✓	✓		✓	✓		✓	✓	✓
3-dimensional spatial domains	✓	✓	✓	✓		✓		✓	
Can model extracellular polymeric substances	✓	✓	✓	✓					
Can simulate gene regulatory networks				✓	✓	✓		✓	
Can model diffusive boundary layer	✓	✓	✓						
Can model growth using FBA							✓		
Can model pH dynamics			✓						
Can model fluid dynamics			✓					✓	

CHAPTER 3 - NUMERICAL TESTING AND *IN SILICO* VALIDATION

3.1 Introduction

One of the biggest barriers to the acceptance of predictions from *in silico* models is the lack of confidence in the validity of those predictions (Hewitt et al., 2015). This lack of confidence is understandable given the level of abstraction at which *in silico* models operate. However, one achievable way to alleviate doubts for potential users of a modelling platform is a rigorous and transparent process of testing.

While testing is typically an ongoing process within software development, it can become easy to forget that users deserve just as much clarity regarding testing procedures and certainty that a model does "what it says on the tin" as developers do, perhaps even more.

The testing of eGUT has taken two different forms during the development. There is continuous testing of the basic functions of the platform. This involves checking that the program is able to run simulations and produce output without obvious observable errors. This simply requires regular running of new models that test new functionality, and of previous models to ensure that all previous functionality is maintained. More involved testing includes the numerical testing, verification and validation described in this chapter, where the results produced by eGUT models are tested against mathematical predictions, and against other models.

As eGUT's functions are centred around the interactions between agents and solutes, the performance of the chemostat solver and the diffusion-reaction solver are key to the working of the platform. Unlike more visually observable properties of the platform, such as agent relaxation, it is not obvious when there is a problem with the ODE and/or PDE solver. In order to ensure that these algorithms, which control solute diffusion and microbial metabolism, are numerically correct, extensive inspection and verification of the software is required.

In order to test eGUTs ODE and PDE solvers, outcomes from eGUT are compared against a known analytical solution, that can be derived from the mathematical principles that the software implements. This type of testing will henceforth be referred to as "numerical testing". Such testing allows us to verify the accuracy of the reaction and diffusion of solutes within eGUT simulations, both in spatial and non-spatial compartments.

For testing other aspects of the platform, verification against analytical solutions may not be possible. An alternative form of verification is testing against other models, ideally those that have been tested and used widely. This involves modelling the same system using two different models, one being the validating model(s) – a model or group of models that make trusted predictions, and a model of interest, whose performance is compared against the validating model(s). The parameters and assumptions used in both models should match as closely as possible in order to ensure that they are attempting to model the exact same system, within their

individual limitations. Similarity of results suggest that the modelling platform is able to simulate the process of interest at least as well as the validating model(s), or that the modelling platform is overfitted due to the modellers making assumptions about the modelled system based on the outcomes that are being predicted.

Here, eGUT undergoes validation by means of a standard test case, known as Benchmark 3, which has been designed to compare various biofilm models, each with different designs, structures, assumptions and simplifications.

Finally, when trying to model particular outcomes on one scale using behaviours or interactions on another scale, an approach known as pattern-oriented modelling can be of great value. This approach requires modellers to identify and define key characteristics of the real-life system they are trying to replicate, and then observe the effects of a wide range of parameters operating at a lower scale on the outcome characteristics. This can be used to calibrate a series of variables in order to achieve a particular higher-level outcome or phenomenon (Grimm and Railsback, 2012). Here, pattern-oriented modelling is used to identify key physical characteristics of mucus in the gut mucosa, and then to calibrate a series of physical parameters in order to simulate mucus that behaves as it does in real life.

3.2 Methods

3.2.1 eGUT and Java

For the numerical tests and the Benchmark, the latest version of the eGUT modelling platform was used, with the FAS diffusion-reaction solver. Simulations were run using a Java 8 virtual machine.

For the Benchmark 3 simulations, a timestep of 12 minutes was used for all processes. Within this timestep, various core processes are simulated in a set order, while other processes (specifically, data reporting processes) occur less regularly than the global timestep. The order of processes in the spatial domain is as follows:

1. Agent removal – Agents with centers higher than 500 μm above the base of the biofilm are removed

2. Mechanical relaxation – Either shoving or Force-based Mechanical relaxation to minimize agent overlaps

3. Reaction-diffusion – Agents determine their reaction rates, based on solute concentrations and biomass amounts. Active agents also grow and divide. Solute concentration grids are updated according to reaction rates, and the boundary with the bulk compartment is updated

4. Reporting (only every 2 simulated hours) – Biomass density grids and totals of different biomasses are written to files

In the bulk compartment, there is a simpler series of processes as follows:

1. Solute concentrations are updated according to inflows, outflows and diffusion into the biofilm (as determined by the boundary between the two compartments)

2. A file recording solute concentrations is updated.

These two sets of processes are carried out separately within each timestep, with the bulk compartment carrying out its processes before the biofilm compartment.

3.2.2 Mucus tests

Mucus tests in eGUT were run using a Java 8 Virtual Machine, on the University of Birmingham's BlueBEAR HPC for 10 hours. Simulations comprised of a spatial domain 50 μm wide by 30 μm high. An epithelium was placed at the bottom edge of the domain, and was made up of 10 colonocytes and 10 goblet cells, each with a width of 2.5 μm and a height of 5 μm. These simulations were run in September 2019 on a slightly older version of the eGUT platform, with a different reaction-diffusion system to the one currently in use, and with epithelial cells modelled with a true volume within the spatial domain. This version is saved as a separate branch in the eGUT Github under the name "Sept_2019_mucus_tests". This should not be relevant to the results described here, which depend solely on the mechanical modelling of agents,

rather than the modelling of diffusion and reaction. Results were downloaded from BlueBEAR and the number of completed timesteps was recorded for each simulation. Visual scoring of the output images was performed by the author. Scoring is described in detail in Section 3.5.1. The full dataset of simulation results are available from the author on request.

3.2.3 Statistics

All statistics and data visualisation were carried out in R. Hotelling's T^2 test was carried out using the "ICSNP" package. Linear discriminant analysis was carried out using the "MASS" package.

3.3 Results

3.3.1 Numerical Testing of ODE solver

An important numerical solver implemented in eGUT is the Chemostat Solver process. This is a generic solver for ODE systems which calculates rates of change of solute concentrations, based on flow rates of matter into and out of the chemostat, and consumption of solutes by agents within the chemostat. In order to confirm that this system was producing numerically accurate results, two chemostat simulations were developed with known steady state solutions. Within these simulations, agents consumed substrate at a rate proportional to their biomass and to the substrate concentration. In one simulation (non-growing catalyst), a single agent was simulated which consumed substrate at a rate proportional to its biomass and to the substrate concentration in the chemostat. Biomass was held at a fixed amount, being unaffected by metabolism, decay or medium outflow, meaning the agent behaved more like a fixed catalyst inside the chemostat than a microbe. In the more complex test case (growing population), the agents behaved like microbes, consuming substrate at a rate determined by Monod kinetics, and growing at a rate equal to their consumption of the substrate. These agents multiplied as they grew but the population was subject to dilution by being removed with outflowing medium.

3.3.2 Non-growing catalyst agent

This test comprised of a single agent in a chemostat compartment, which used up the inflowing solute at a rate proportional to its concentration. The inflowing medium contained the solute at a set concentration, while outflow carried medium out of the chemostat. The rates of inflow and outflow were equal. This is a very simple setup which tested the basic functionality of the ODE solver used to model well-mixed systems such as chemostats.

This system can be described by the following differential equation

$$\frac{dS}{dt} = \frac{FS_0}{V} - \frac{FS}{V} - \frac{mkS}{V}$$

where S is the solute concentration in the chemostat, S_0 is the solute concentration in the inflowing medium, F is the flow rate with dimension volume per time, V is the volume of the chemostat, t is time, k is a constant governing the rate of solute consumption by the agent and m is the mass of the agent.

The steady-state solution for this differential equation is

$$S^* = \frac{FS_0}{F + mk}$$

A chemostat simulation was set up in eGUT with a non-growing catalyst agent, parameters set according to Table 3.1. The model was simulated in time steps of 100 minutes, for a total of 2000 timesteps (138.9 simulated days). The steady state

predicted for the parameters in this case was 4.0×10^5 pg μm^{-3} and the concentration converged to this concentration exactly (Figure 3.1).

Table 3.1 – Parameters used in the numerical tests for the chemostat solver. Mass is not defined in the growing population simulation, as this was a simulation of a growing population with mass changing over time.

Parameter	Catalyst in chemostat	Growing population in chemostat
S_0	$2.0 \times 10^6\ pg\ \mu m^{-3}$	$2.0\ pg\ \mu m^{-3}$
F	$1.0 \times 10^6\ \mu m^3\ min^{-1}$	$1.0\ \mu m^3 min^{-1}$
V	$1.0 \times 10^9\ \mu m^3$	$1.0 \times 10^3 \mu m^3$
k	$1 \times 10^5\ \mu m^3 min^{-1} pg^{-1}$	$1.0\ \mu m^3 min^{-1} pg^{-1}$
m	$40\ pg$	
μmax		0.1
Y		0.5

3.3.3 Growing Population in a Chemostat

In this simulation, a growing population of agents consumed the substrate and converted it to biomass. Agents were removed from the chemostat population at a proportion equal to the proportion of the chemostat's volume removed with the outflow with each timestep. This setup allows the ODE solver to be tested with a fluctuating biomass in the chemostat. It also provides another variable to test, given

that both the steady state solute concentration and the steady state biomass concentration can be tested against analytical solutions. Additionally, a Monod model of substrate consumption was implemented, such that

$$\mu = \frac{\mu_{max} S}{K_S + S}$$

Where μ is the specific growth rate, μ_{max} is the maximum growth rate, S is the substrate concentration and K_S is the half-saturation constant, the value of S at which $\mu = \mu_{max}/2$.

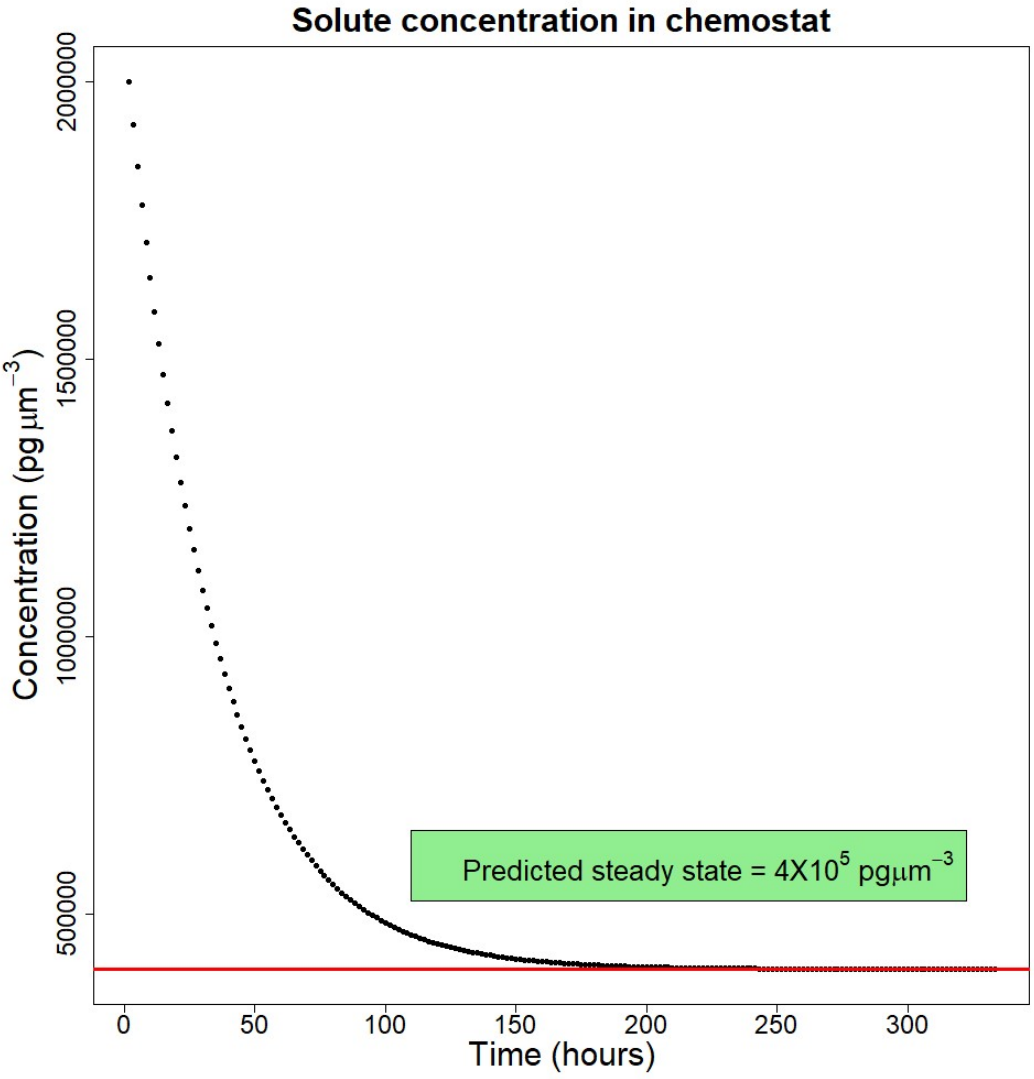

Figure 3.1 – Solute concentration over time in the chemostat test case with a non-growing catalyst agent. The concentration converges asymptotically to the expected stable steady state indicated by the red line.

Here, the rate of change of substrate concentration is given by

$$\frac{dS}{dt} = \frac{FS_0}{V} - \frac{FS}{V} - Y^{-1}\mu(S)P$$

where Y is the biomass yield from the substrate and P is the biomass concentration in the chemostat. This can be solved to find a steady-state for both P and S, ignoring the washout steady state of $P = 0,\ S^* = S_0$ (Kreft, 2009):

$$S^* = \frac{{}^{F}/_{V}\, K_S}{\mu_{max} - {}^{F}/_{V}}$$

$$P^* = Y(S_0 - S^*)$$

With the parameter values in Table 3.1, we obtain the following steady state predictions:

$$S^* = 0.010101\ pg\ \mu m^{-3}$$

$$P^* = 0.994949\ pg\ \mu m^{-3}$$

Running the simulations in eGUT yields the expected stable steady-state (Figure 3.2).

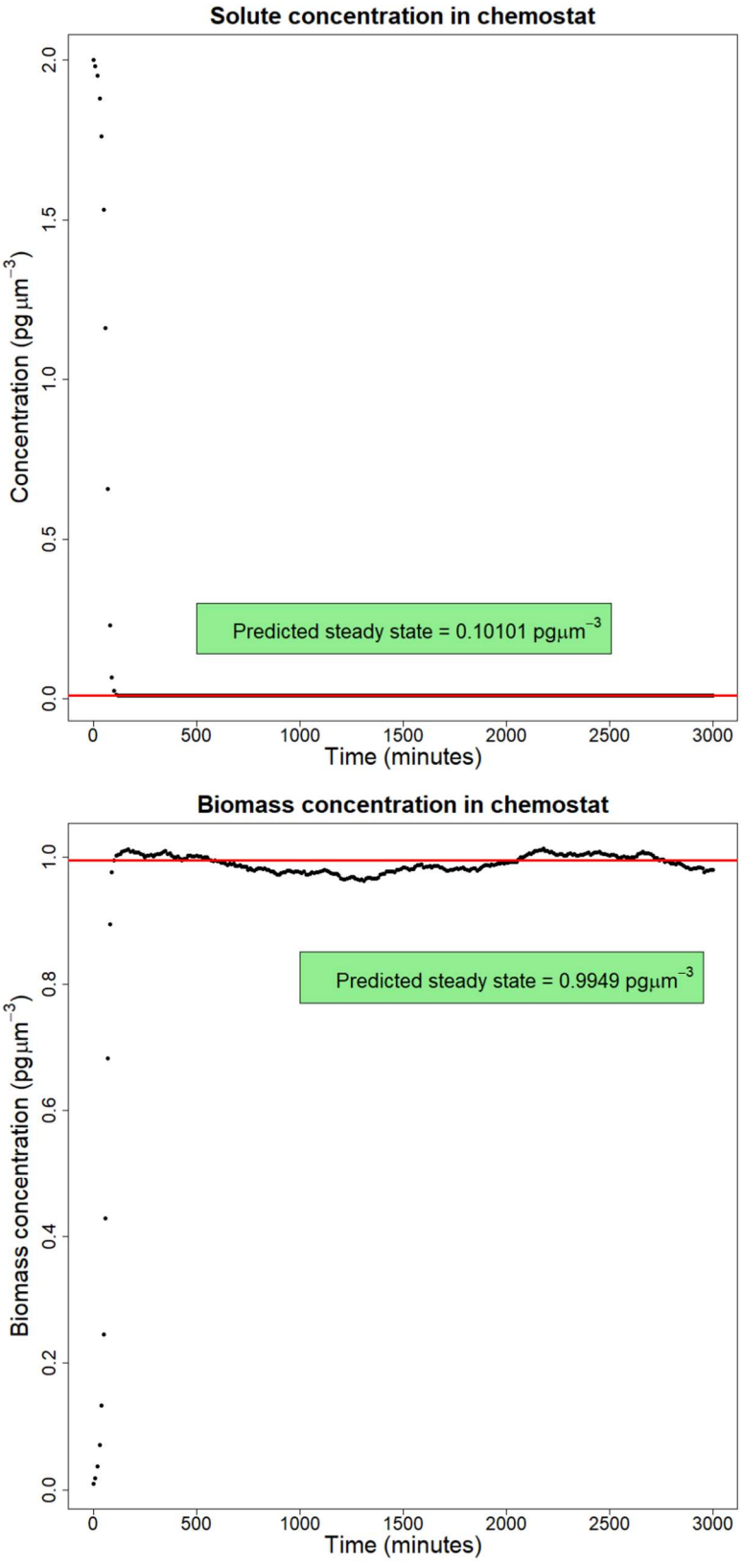

Figure 3.2 – The solute and biomass concentrations in the chemostat test with a growing population. The solute concentration converges quickly to the expected

stable steady state of 0.010101 pg μm^{-3} and the population quickly converges to the stable steady state value of 0.9949 pg μm^{-3}. Predicted steady states are indicated by red lines. Note the stochastic removal of agents from the chemostat results in biomass fluctuations around the steady state.

3.3.4 Spatial Domain Numerical Tests

The next set of numerical tests were carried out on the reaction-diffusion solver for eGUT's spatial domains. This is the PDE solver, utilising the FAS algorithm. During the initial implementation of this algorithm, we experienced various problems, and I spent significant time developing a reporter system to track the solution over the course of the solver steps, as well as Matlab scripts to visualise the solving process. After these initial problems had been eliminated, a test of the solver against analytical solutions was an important final step. This allows us to verify that the solver continues to work correctly without having to scrutinise the individual variables and calculations in the FAS algorithm.

In this setting, agents are once again treated like catalysts, able to consume the solute, but not to grow. The 2D spatial domain is connected at its top edge to a chemostat with a large volume containing solute at a set concentration. The solute diffuses through the domain from the well-mixed region at the top through the diffusion dominated region below. The speed of this diffusion is determined by its diffusivity and the concentration gradient. In the first test case, that of a thin layer of cells at the

bottom edge of the domain, an analytical solution for the flux, J can be calculated in this case. A second test case, with a 16 μm thick biofilm was also simulated. I am not aware of any known analytical solution for this case, but the relationship between height and solute concentration should have a characteristic shape which the simulation results can be compared to.

Thin non-growing cell layer

In this test, a thin non-growing layer of cells was simulated at the bottom of a spatial compartment, with a diffusive boundary layer above the cells, and a well-mixed region above that with a concentration of substrate at concentration S_0. The layer comprised a single layer of cells all within the bottom layer of reaction-diffusion grid elements. This setup allows the diffusion of solutes through the diffusive boundary layer to be tested against an analytical solution derived from Fick's law.

By Fick's first law for one-dimensional flux,

$$J = D\frac{dS}{dx}$$

where J is the areal flux density through the diffusive region, D is the diffusivity of the solute and x is distance in the dimension perpendicular to the direction of flux, here perpendicular to the plane on which the cells are placed (let us call this the x-axis) through which the solute must diffuse.

Given that at steady state, flux must be constant along the x-axis in the region where the substrate is not consumed, and is driven by consumption by the cells at one end, we can substitute J for the consumption rate at the cell-layer surface. Modelling a simple consumption rate proportional to biomass and substrate concentration, we get

$$\frac{mkS^*}{A} = D\frac{S_0 - S^*}{\Delta x}$$

where S* is the steady-state concentration at the biofilm surface, A is the surface area of the biofilm and Δx is the depth of the diffusive region. This can be rearranged to

$$S^* = \frac{DAS_0}{\Delta xmk + DA}$$

Setting the parameters as shown in Table 3.2, the predicted steady state concentration at the cell layer surface is $S^* = 1.8 \times 10^{-6} pg\ \mu m^{-3}$. In the simulation, the intercept is very close to the predicted value of S*, at $1.788 \times 10^{-6} pg\ \mu m^{-3}$ (Figure 3.3), with a slight deviation below the expected steady state of ~0.7%. This may be due to the discrete nature of the solute grid, which is not taken into account in the mathematical model. The value at a height of x = 0 is directly below the cell-layer, while the value at height x = 1 is directly above the cell-layer. This means that in eGUT, the solute concentrations at x = 0 and x = 1 are both affected by the metabolism of the cells, while the mathematical model only considers solute consumption at a boundary.

Table 3.2 – Values of the parameters used in the numerical tests for eGUT's spatial domain. Values were equal for all parameters except for the total biomass

Parameter	Thin cell layer	Biofilm
S_0	$2.0 \times 10^{-6}\ pg\ \mu m^{-3}$	$2.0 \times 10^{-6}\ pg\ \mu m^{-3}$
D	$36{,}000\ \mu m^2 min^{-1}$	$36{,}000\ \mu m^2 min^{-1}$
k	$100\ \mu m^3 min^{-1} pg^{-1}$	$100\ \mu m^3 min^{-1} pg^{-1}$
m	$128\ pg$	$1200\ pg$
Δx	$10\ \mu m$	$10\ \mu m$
A	$32\ \mu m^2$	$32\ \mu m^2$

3.3.5 Biofilm in spatial domain

For a biofilm simulation with a thicker layer of cells, no analytical solution is available for solute concentration at the surface of the biomass. However, the nature of the solid boundary at the bottom of the domain provides another testable feature. Boundary conditions at "solid" boundaries in eGUT are Neumann boundaries that have a no-flux condition. This means the concentration gradients at this boundary must be 0, while the concentration gradient at the surface of the biofilm is non-zero, as solute is diffusing into the biofilm. Thus within the biofilm, we expect a concentration gradient that is 0 at the bottom of the domain, and increases towards the surface of the biofilm, and a fixed concentration gradient within the diffusion dominated region above the

biofilm, as with the thin cell layer test. Testing this allows us to ensure that the multigrid solver works correctly when solving a system with reactions distributed through the vertical dimension over multiple layers of grid cells. As in the previous test, there is a well-mixed region above the diffusive region, in which the concentration remains equal to the bulk. The results of the test replicate the predicted features of the concentration-height curve well (Figure 3.4), suggesting that the boundary conditions imposed by the solid boundaries within eGUT are functioning as expected.

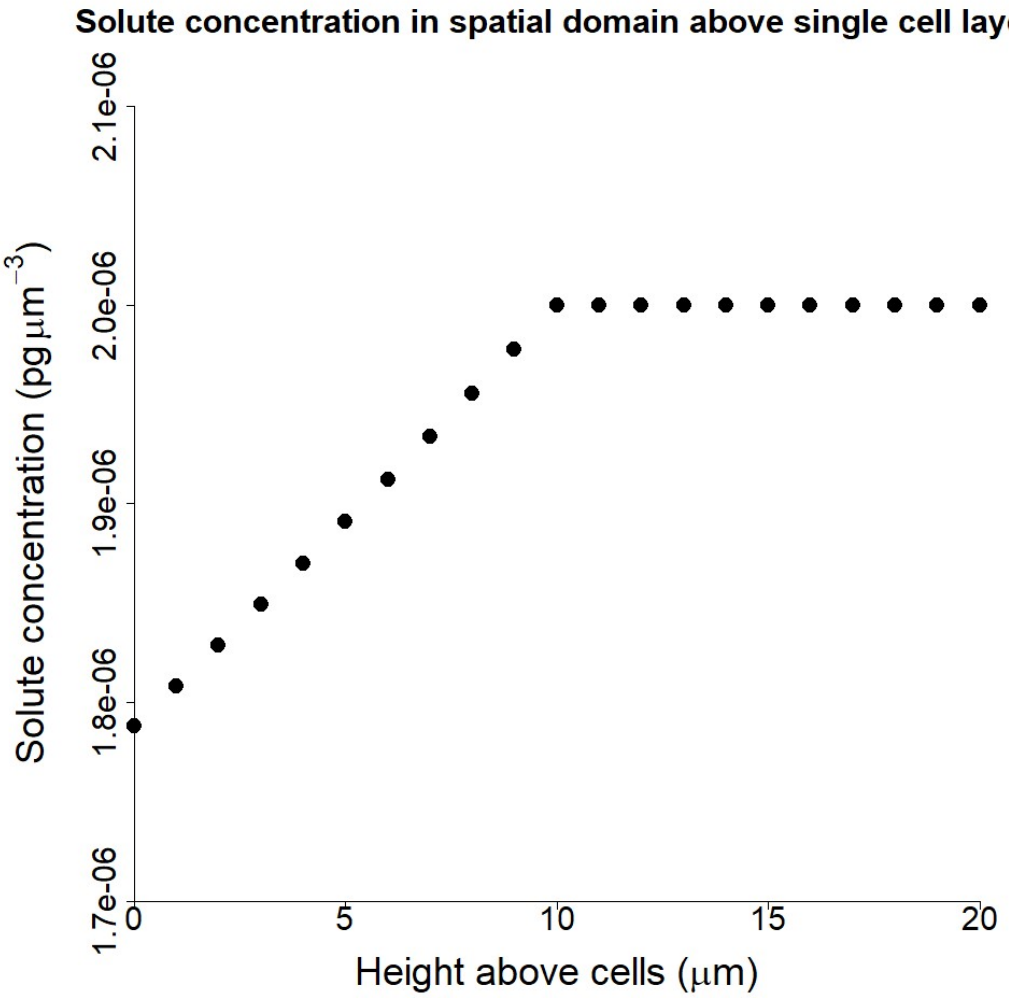

Figure 3.3 – The relationship between height and concentration in a spatial domain with a single layer of cells at the bottom of the domain. The gradient between S0 and S* in the diffusive boundary layer is linear as expected.

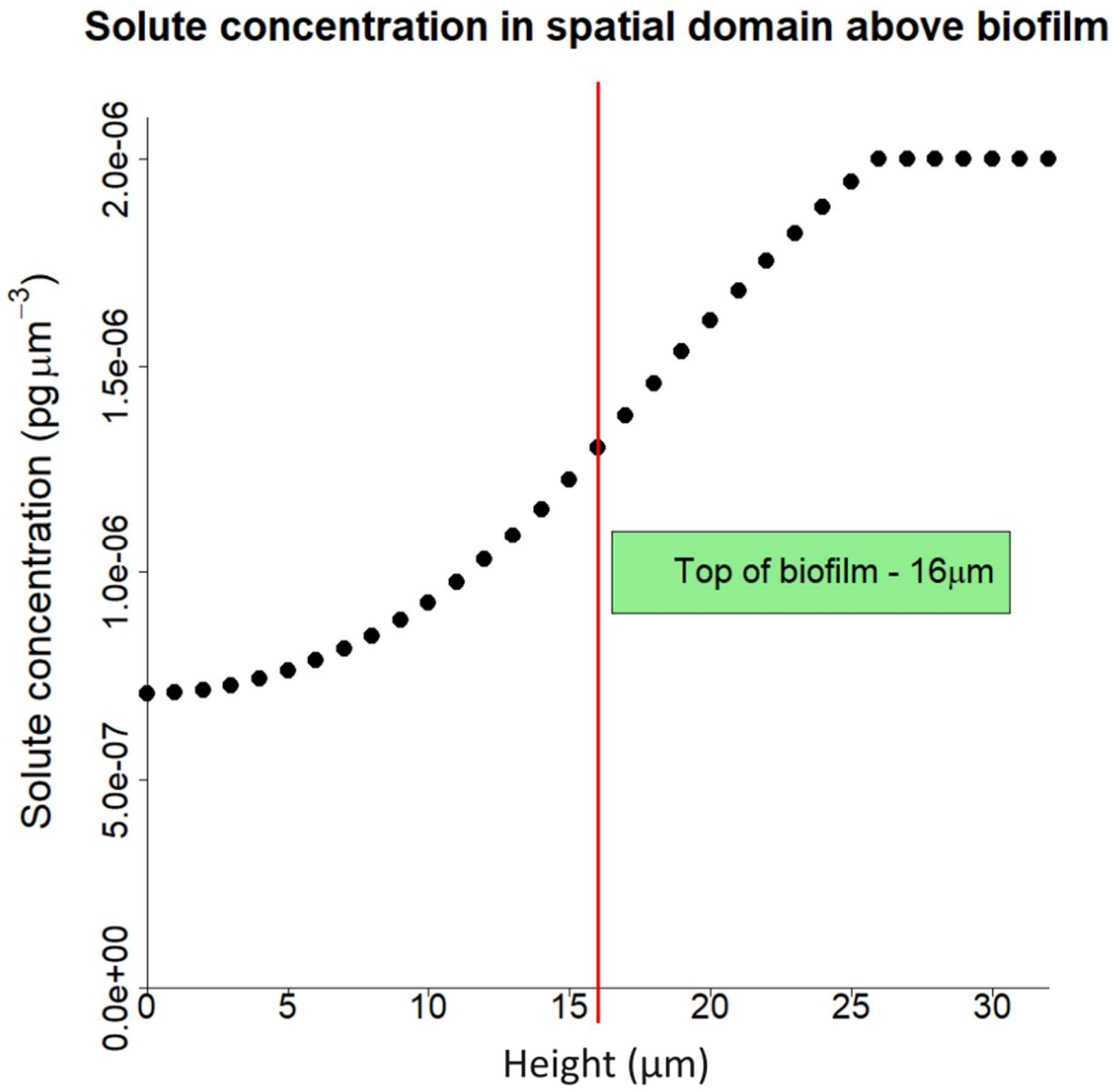

Figure 3.4 – Solute concentration plotted against height in a spatial domain with a biofilm in the bottom 16 µm. The concentration gradient is expected to be linear in the diffusive boundary layer above the biofilm and monotone decreasing to zero inside the biofilm, consistent with the simulation results.

3.4 Benchmark 3

3.4.1 Introduction

One of the most important applications of modelling microbial growth is in modelling the treatment of wastewater. The International Water Association (IWA) set up a Biofilm Modelling Task group to focus on modelling microbes in water treatment systems which published a series of Benchmark models. These were typical systems that were designed to act as a point of comparison between biofilm models and thus help to establish the effects of different model designs and simplifying assumptions on simulation outputs (Wanner et al., 2006). Benchmark 3 was designed to simulate microbial competition in a biofilm, with a source of carbon and energy represented by chemical oxygen demand (COD) being oxidized by a population of heterotrophs and a population of autotrophs oxidizing ammonia to nitrate (nitrification by complete ammonia oxidation).

The IWA Biofilm Modeling Task Group ran BM3 simulations on a wide range of modelling platforms, with a variety of different approaches to modelling biofilms. Later, BM3 was also used for model validation in the development of iDynoMiCS 1 (2011) and NUFEB (Li et al., 2019). Here, eGUT is compared against a selection of five models from the original IWA task group, as well as NUFEB and iDynoMiCS 1. A summary of the different models and their approaches to BM3 follows:

- CP – a two-dimensional agent-based model, with biomass spreading via shoving, developed by Cristian Picioreanu and colleagues (Picioreanu et al., 2004)

- DN – a two-dimensional cellular automaton model developed by Daniel Noguera and colleagues (Noguera et al., 2004)

- W – a one-dimensional continuum biomass model run on the AQUASIM software (Reichert, 1994) and developed by Peter Reichert and Oskar Wanner (Reichert and Wanner, 1997; Wanner and Reichert, 1996)

- M1 – a variant of the W model with a fixed boundary-layer thickness by Eberhard Morgenroth et al. (Morgenroth and Wilderer, 2000)

- NUFEB - A three-dimensional agent-based model that uses a platform derived from a molecular dynamics simulator by Li et al. (2019)

- iD – A two-dimensional agent-based model by Lardon et al. (2011). This platform is the precursor to the one described in this paper, and the implementation of BM3 is very similar

As this set of modelling platforms represents a variety of different modelling approaches, they provide a valuable set of results against which to compare eGUT. The BM3 scenario has previously been used by Lardon et al. (2011) and Li et al. (2019) to benchmark iDynoMiCS 1 and NUFEB, respectively. A description of the

implementation of the BM3 model in eGUT follows, henceforth referred to by the abbreviation BM3-eGUT.

3.4.2 Model Description

This model simulates multi-species biofilms growing in an aqueous environment as commonly found both in nature and in treatment systems for wastewater and drinking water. The biofilm is composed of two species representing microbial functional groups – an aerobic heterotroph and an aerobic autotrophic nitrifier. Both of these species undergo inactivation processes which transform an agent's active biomass to inert biomass, meaning that there are three types of biomass present in the biofilm: heterotrophic, autotrophic and inert. The two microbial species compete for oxygen and for space in the biofilm and are transformed into the same inert biomass, leading to vertical stratification of the three different types of biomass through the biofilm.

The biofilm is connected to a well-mixed, aerated bulk, from which these nutrients and oxygen diffuse. Both species grow according to Monod kinetics. There is no diffusive boundary layer, meaning the diffusive region only corresponds to any regions of the spatial domain containing biomass and any biomass-free grid elements are treated as well-mixed by the reaction-diffusion solver. Biomass above a certain height at the top of the biofilm is removed. Physical and chemical characteristics of

the environment are given in Table 3.3, and growth stoichiometry parameters are given in Table 3.4.

Several variants of the Benchmark 3 model exist. Here, the standard case was simulated, as were two special cases: a high influent ammonium case, and a low influent ammonium case. Three replicates of each case were run with different random number seeds to ensure that the results were reproducible and not contingent on the starting conditions. Simulations were run for 120 simulated days, but solute concentrations typically reached steady state much sooner than that.

In order to achieve the biofilm density of 10 g L^{-1} given in the original model description, a suitable agent density had to be found. After a series of preliminary BM3 simulations, a suitable agent density of 12.5 g L^{-1} was found to produce biofilms with a density of ~10 g L^{-1}.

Table 3.3 – Physical parameters used in the Benchmark 3 simulations.

Parameter	Value		
	Standard case	High ammonium	Low ammonium
Ammonium influent concentration	$6\ g\ m^{-3}$	$30\ g\ m^{-3}$	$1.5\ g\ m^{-3}$
Dilution rate	$0.0111\ min^{-1}$		
Volume of bulk liquid	$4.0\ \times 10^{6} \mu m^{3}$		
COD influent concentration	$30\ g\ m^{-3}$		
Carrier surface area	$320\ \mu m^{2}$		
Biofilm thickness	$500\ \mu m$		
Constant oxygen concentration in bulk liquid	$10\ g\ m^{-3}$		
Biofilm density	$10\ g\ L^{-1}$		
Agent density	$12.5\ g\ L^{-1}$		
Agent division dry mass	$4\ pg$		
Boundary layer thickness	$0\ \mu m$		
Shove Factor	1.05		

Table 3.4 – Petersen (stoichiometric) matrix for reactions in the Benchmark 3 simulations. Table and values adapted from Rittmann (2004) and Lardon et al. (2011). Biomasses are denoted with X. Specifically, X_H = heterotroph active biomass, X_N = nitrifier (autotroph) active biomass. Substrate concentrations are denoted with S, S_S for the organic substrate COD, S_N for ammonium and S_{O2} for oxygen. For the values of the parameters used in these equations, see Table 3.5.

	Biomass type		Substrate			Kinetic expression
	Active	Inert	S_S	S_N	S_{O2}	
Heterotroph growth	1		$\frac{-1}{Y_H}$		$\frac{-(1-Y_H)}{Y_H}$	$\mu_{max,H}\frac{S_S}{K_S+S_S}\frac{S_{O2}}{K_{O2,H}+S_{O2}}X_H$
Heterotroph decay	-1	1				$b_{ina,H}X_H$
Heterotroph maintenance	-1				-1	$b_{res,H}X_H\frac{S_{O2}}{K_{O2,H}+S_{O2}}$
Autotroph growth	1			$\frac{-1}{Y_H}$	$\frac{-(4.57-Y_N)}{Y_N}$	$\mu_{max,N}\frac{S_N}{K_N+S_N}\frac{S_{O2}}{K_{O2,N}+S_{O2}}X_N$
Autotroph decay	-1	1				$b_{ina,N}X_N$
Autotroph maintenance	-1				-1	$b_{res,N}X_N\frac{S_{O2}}{K_{O2,N}+S_{O2}}$

Table 3.5 - Kinetic parameters in the Benchmark 3 model

Parameter	Value
$u_{max,H}$	$5.9976\ day^{-1}$
K_S	$4\ g\ m^{-3}$
Y_H	0.63
$K_{O2,H}$	$0.2\ g\ m^{-3}$
$b_{res,H}$	$0.32\ day^{-1}$
$b_{ina,H}$	$0.08\ day^{-1}$
$u_{max,N}$	$0.1386\ day^{-1}$
K_N	$1.5\ g\ m^{-3}$
Y_N	0.063
$b_{res,N}$	$0.12\ day^{-1}$
$b_{ina,N}$	$0.03\ day^{-1}$
$K_{O2,N}$	$0.5\ g\ m^{-3}$

3.4.3 Results

Benchmark 3 simulations have been run on a wide variety of other modelling platforms ranging from 1 dimensional continuum models to 2 dimensional and 3 dimensional agent-based models like eGUT and NUFEB. In particular, iDynoMiCS 1, previously developed by our research group, has a similar architecture to eGUT, making similar assumptions and modelling biofilms in a similar way (Lardon et al.,

2011). Thus, a close match between the results of iDynoMiCS 1 and eGUT should be expected if eGUT is functioning as it should.

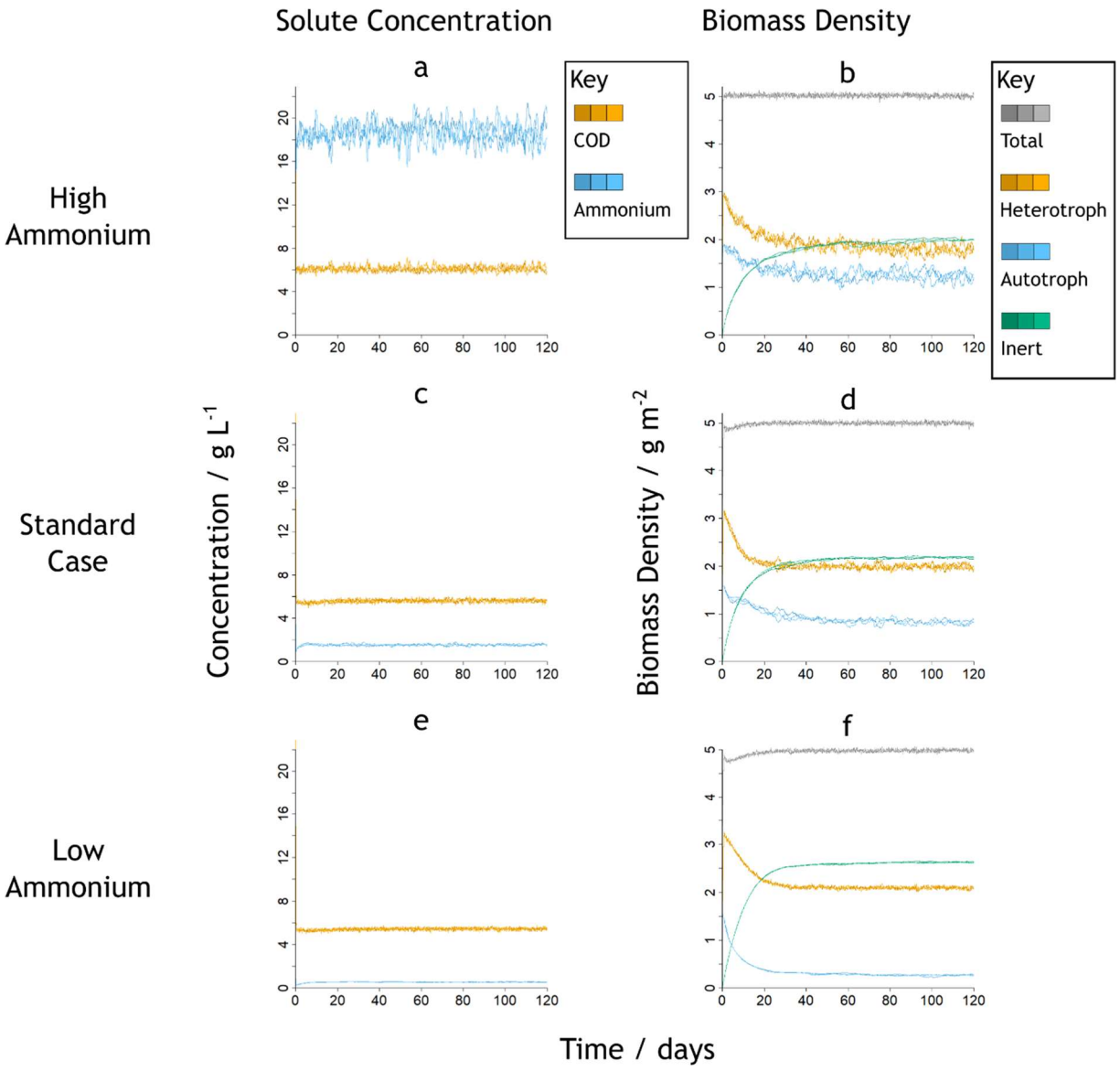

Figure 3.5 – Solute and biomass concentrations from the three Benchmark 3 scenarios simulated. Solute concentrations are those in the bulk liquid and biomass concentrations are biofilm totals. Oxygen is not included in the results, as this has a fixed concentration in the bulk. The three colours represent the three replicates.

The solute concentrations and biomass densities reached steady state within the 120 simulated days (Figure 3.5). The steady state concentrations from the BM3-eGUT simulations were compared with the results from other models using the multivariate version of the t-test, Hotelling's 1-sample T^2 test (Figure 3.6, Table 3.6). The results from BM3-eGUT are generally close to results from the other models, but with near-significant p-values, suggesting there may be significant differences between BM3-eGUT and the rest of the models. Notably, the BM3-eGUT results match those from iDynoMiCS 1 very closely, which is the expected result given that both models use the same reaction-diffusion solver and are similar modelling platforms with similar designs. The main thing that distinguishes BM3-eGUT and iDynoMiCS 1 from the other models is that BM3-eGUT and iDynoMiCS 1 tend to have slightly higher steady-state COD concentrations than the other models, suggesting that the biofilms in these models may have more autotrophic biomass and less heterotrophic biomass than other models.

Steady state biomass densities were also monitored in the eGUT simulations, and these were also compared to biomass densities found in the other models. Specifically, the areal biomass density was compared between the different models. This is the amount of a particular type of biomass per unit surface area at the base of the biofilm (Figure 3.7). As shown in Table 3.7, BM3-eGUT did not differ significantly from the other models in overall areal biomass densities.

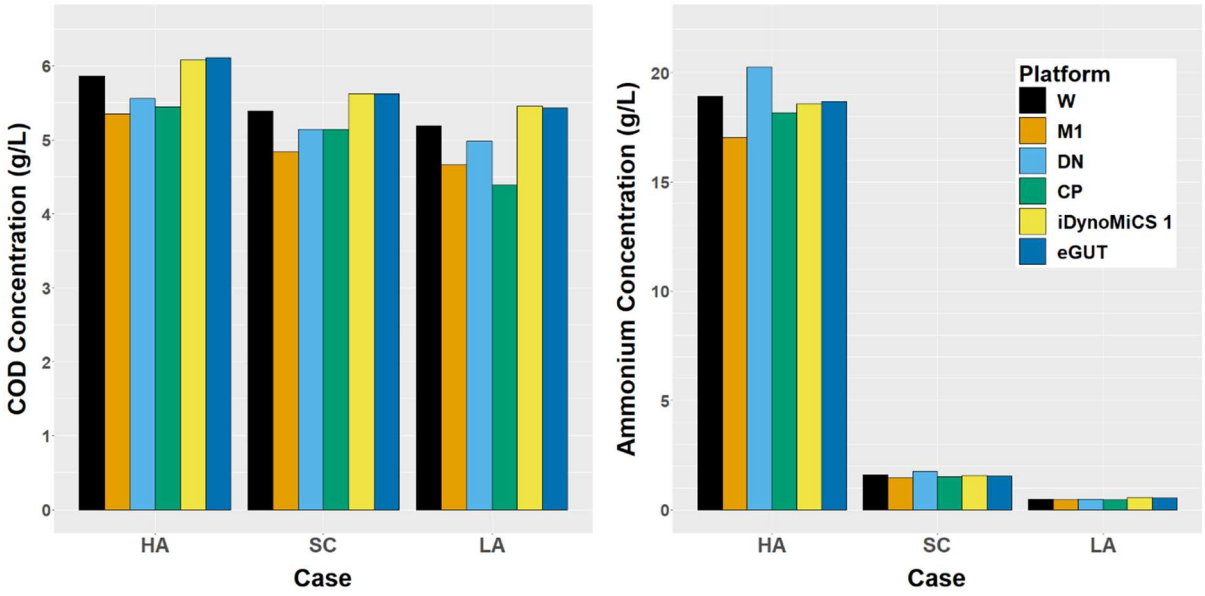

Figure 3.6 – Steady state concentrations of COD and ammonium in BM3-eGUT and the other BM3 models. HA = high ammonium, SC = standard case and LA = low ammonium

As well as the overall areal densities, the vertical distribution of the different biomass types through the biofilm were also investigated. These are plotted in Figure 3.8. These show a qualitatively similar pattern to the CP model (Noguera and Picioreanu, 2004), with fast-growing heterotrophs dominating the top of the biofilm, while autotrophs grow more slowly and are at their most abundant in the middle or bottom of the biofilm. Autotrophs vary widely in abundance between the different cases, being at very low numbers in the low ammonium case. This is to be expected, given that their energy source is at a low concentration. In all three cases, the bottom of the biofilm is dominated by inert biomass due to the lower substrate concentrations at the bottom of the biofilm reducing growth relative to maintenance and inactivation. This is most pronounced in the low ammonium case, which has the highest proportion of inert biomass of the three cases.

Table 3.6 – Steady state bulk liquid solute concentration results from Benchmark 3. All concentrations are in g m^{-3}. A 1-sample Hotelling's T2 test was performed comparing the set of comparison models to the results from eGUT.

		High ammonium		Standard case		Low ammonium	
		COD	Amm.	COD	Amm.	COD	Amm.
IWA models	CP	5.45	18.15	5.14	1.50	4.39	0.44
	DN	5.56	20.26	5.14	1.74	4.98	0.48
	W	5.86	18.93	5.39	1.59	5.19	0.48
	M1	5.35	17.03	4.84	1.45	4.66	0.45
	NUFEB	5.74	18.42	5.21	1.72	5.18	0.53
	iDynoMiCS 1	6.08	18.58	5.63	1.55	5.45	0.55
	BM3-eGUT (± standard deviation)	6.11 ± 0.05	18.68±0.02	5.62±0.01	1.54±0.01	5.43±0.01	0.54±0.01
	Hotelling's T^2 Test p-value	0.0548		0.0523		0.1306	

Given the differences in modelling approaches of the various IWA task group models, one might expect eGUT and iDynoMiCS 1 to produce results closer to the CP and NUFEB particle-based models than to any of the other IWA task group models. However, this was not observed. Instead, the results from the W platform were the closest match in steady state solute concentrations, while those from the M1 platform were the closest match for overall biomass densities. This is particularly interesting

given that these are both 1-dimensional platforms utilizing the AQUASIM software rather than agent-based. It is possible that the similarity derives from a closer match in biomass distribution than with the other models due to less stochastic mixing of the biomass. However, as the biomass distributions were not published for these models, this is difficult to determine.

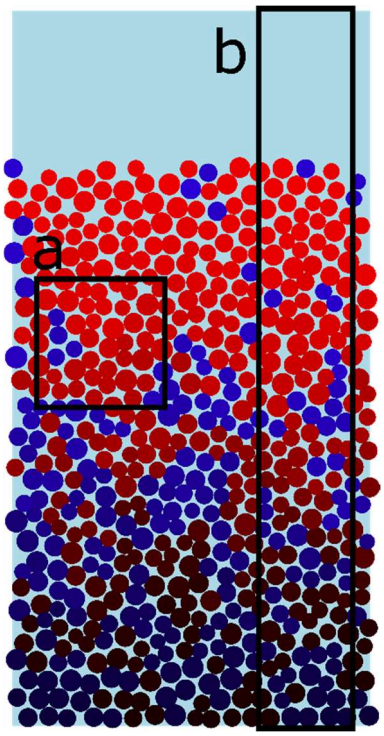

Figure 3.7 – A diagram to illustrate the difference between density and areal density as used in this chapter. Density, represented by box a, is the amount of biomass per unit volume, while areal density is represented by box b, and is the amount of biomass per unit area at the base of the biofilm. Note, BM3 has a virtual 3rd dimension with a thickness of 1 μm, which is why the boxes shown are considered volumes, rather than areas.

Table 3.7 - Steady state areal biomass density (mass per unit surface area) of different types of biomass in the biofilm. The eGUT results are from simulations using the Shoving algorithm. Hotelling's T^2 tests were performed to compare the results from eGUT to those from the IWA models.

		Areal biomass density (g/m^2)		
		Heterotroph	**Autotroph**	**Inert**
High ammonium	CP	1.71	1.07	2.42
	DN	2.92	1.10	0.98
	W	1.73	1.07	2.20
	M1	1.83	1.24	1.93
	BM3-eGUT	**1.76**	**1.24**	**2.00**
	Hotelling's T^2 Test p-value	0.6016		
Standard Case	CP	1.81	0.72	2.60
	DN	2.88	0.68	1.44
	W	1.88	0.79	2.33
	M1	2.02	0.83	2.15
	BM3-eGUT	**1.94**	**0.86**	**2.17**
	Hotelling's T^2 Test p-value	0.5475		
Low ammonium	CP	2.11	0.23	2.73
	DN	2.96	0.13	1.91
	W	2.00	0.21	2.80
	M1	2.14	0.21	2.65
	BM3-eGUT	**2.08**	**0.26**	**2.62**
	Hotelling's T^2 Test p-value	0.1038		

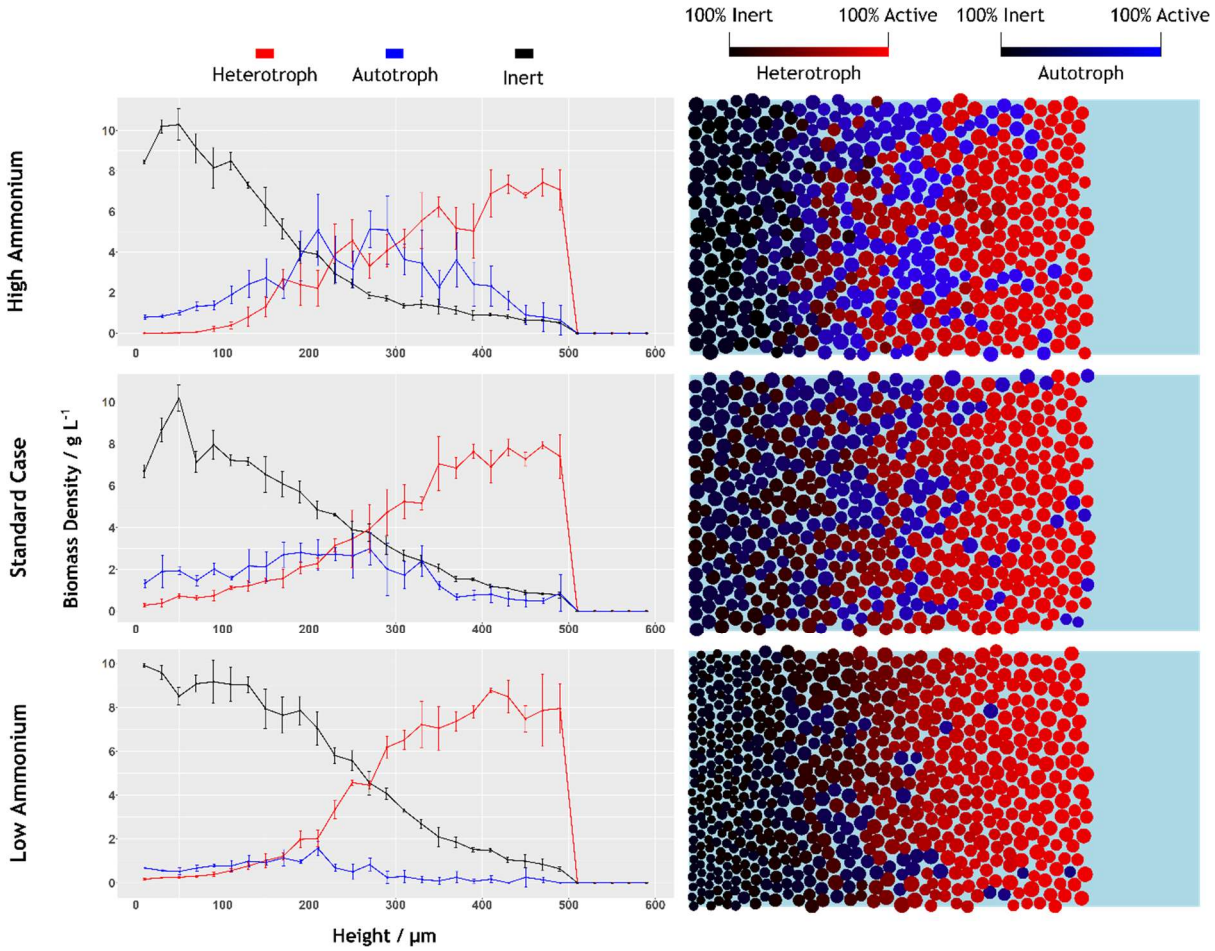

Figure 3.8 – Distributions of autotrophic, heterotrophic and inert biomass in the BM3-eGUT simulations

3.5 Mucus Tests

The gut mucosa is characterised chiefly by its association with the mucus produced by the goblet cells of the gut's epithelial lining (Duncan et al., 2021). The physical properties of this mucus determine how microbes interact with it, with some able to penetrate and reside within the mucus, while for the most part, the constant production of mucus creates a sterile region adjacent to the epithelium. This means having a realistic and reliable model for the physical behaviour and properties of mucus will be an important requirement in many eGUT simulations.

In order to establish a way to simulate the behaviour of the mucus layer that is produced by the epithelial mucosa, a pattern-based modelling approach was taken, focussing on several key attributes of the physical behaviour of mucus. Firstly, mucus is produced by goblet cells. These are fairly numerous, with studies in rats finding goblet cells to make up around 20% of cells in the epithelial surfaces of the colons of healthy animals (Lan et al., 2015; Miller and Nawa, 1979). Goblet cells produce mucus, which then spreads out across the entire epithelium, forming an adherent gel that provides continuous cover for the epithelial surface (Allen, 1989; Atuma et al., 2001). These key elements, adherence to the epithelium and spreading into a continuous layer from relatively frequent 'point' sources, were considered patterns that eGUT simulations should match. Especially the ability to spread across the epithelial layer, providing continuous cover is an important outcome in eGUT, considering the

importance of the barrier function provided by the mucus in protecting the epithelium against pathogens (Allen, 1989; Atuma et al., 2001).

Another key feature of gastrointestinal mucus is its physical properties as a 'gel'. Gastrointestinal mucus has both viscosity (resistance to flow) and elasticity (stiffness). As its elastic modulus is greater than its viscous modulus, it is classified as a viscoelastic solid (Lai et al., 2009). However, the mucus is not so solid that it completely excludes bacteria, as bacteria can and do penetrate the mucus layer, especially the outer layer of mucus, which is typically composed of looser, more soluble, and less viscoelastic mucus, often described as "sloppy" (Johansson et al., 2011). In order to reproduce this loose, open structure, especially on the luminal side of the mucus layer, was a key pattern that simulations should obtain.

Mucus is composed of a network of highly glycosylated proteins called mucins. They are composed of a protein backbone, with brush-like polysaccharide branches. These mucins swell upon secretion by the goblet cell, absorbing water, and forming a hydrogel. Modelling mucus as a fluid, or individual mucins is not feasible in eGUT, due to the structure of the platform, and the computational demands of fluid mechanics modelling. Instead, mucus is modelled as small particles released by the goblet cells.

An attraction function governed the behaviour of attractive forces between agents in the simulation. This was used to model both the attractive forces between mucus particles (simulating cohesion and elasticity) and the attractive forces between mucus

particles and the epithelial cells (simulating adhesion to the cell surface). The attraction force between two agents is defined as the following:

$$F = F_{max}\,(d - s)\frac{e^{1-\frac{(s-d)}{d_{sat}}}}{d_{sat}}$$

Where F is the attractive force between two agents, F_{max} is the maximum force allowed, d is the distance between the surfaces of the two agents, d_{sat} is the saturation distance at which the force is equal to F_{max}, and s is a small safety distance equal to 1 nm which ensures agents directly in contact do not experience attractive forces.

In order to achieve a range of physical behaviours in the eGUT simulated mucus, the following physical parameters, which govern the characteristics and behaviour of these particles, were varied in an optimization process:

Mucus particle mass – the mass of a mucus particle released by the goblet cells (varies inversely with the number of particles). This was varied between 4.0 $\times 10^{-4}$ and 4.0 pg.

F_{max} – The maximum attractive force in pg μm min^{-2}. This was varied between 1.0 × 10^{12} pg μm min^{-2} and 1.0 × 10^{16} pg μm min^{-2}.

d_{sat} – The saturation distance in μm, this was varied between 10 nm and 10 μm.

Temperature – This parameter is used to control the displacement of stochastic movement in the simulation. It is based on the Stokes-Einstein equation describing Brownian motion and requires a temperature value. It should be noted that this

temperature should be seen as a tuning parameter and it does not necessarily have to match the true temperature in the gut. This value was varied between 3 K and 300 K.

In order to investigate the effects of all of these parameters simultaneously, Latin hypercube sampling was used to split each parameter into 1000 small ranges, and sample from each of these smaller ranges just once within the global sample set. A random collection of these ranges was selected for each simulation, yielding 1000 different simulations representing variation across the whole range of each variable. As very little is known about all the variables investigated here, they were varied across multiple orders of magnitude. Thus, values were represented on a log scale for sampling purposes.

3.5.1 Scoring

After running the 1000 simulations, the visual appearance of the mucus was assessed and various different qualitative outcomes were identified. The results were then visually classified into different "mucus structures". These are shown in Figure 3.10.

Many of the mucus structures that arose from these simulations were clearly far removed from any realistic representation of the mucus layer in the colon. The large blob and small blob structures exhibit the opposite behaviour to real mucus, in that instead of adhering to the epithelial layer, the mucus adhered to itself, forming tight blobs. The geometric clusters and distinct mounds also failed to spread across the epithelial surface by adhesion, although they did not exhibit cohesion quite as extreme

as with the blob structures. The tight spread and loose spread structures display adhesion to the epithelial surface, reflecting the pattern observed in real mucus. However, the tight spread structure shows little fluidity or openness, suggesting a very high viscoelasticity, while the loose spread structure is more fluid and open, providing a structure that could be inhabited and possibly moved through by microbial agents. The “Not enough mucus” structure was common, and generally represented cases with very few mucus particles produced, due to the fact that these simulations did not proceed beyond one timestep.

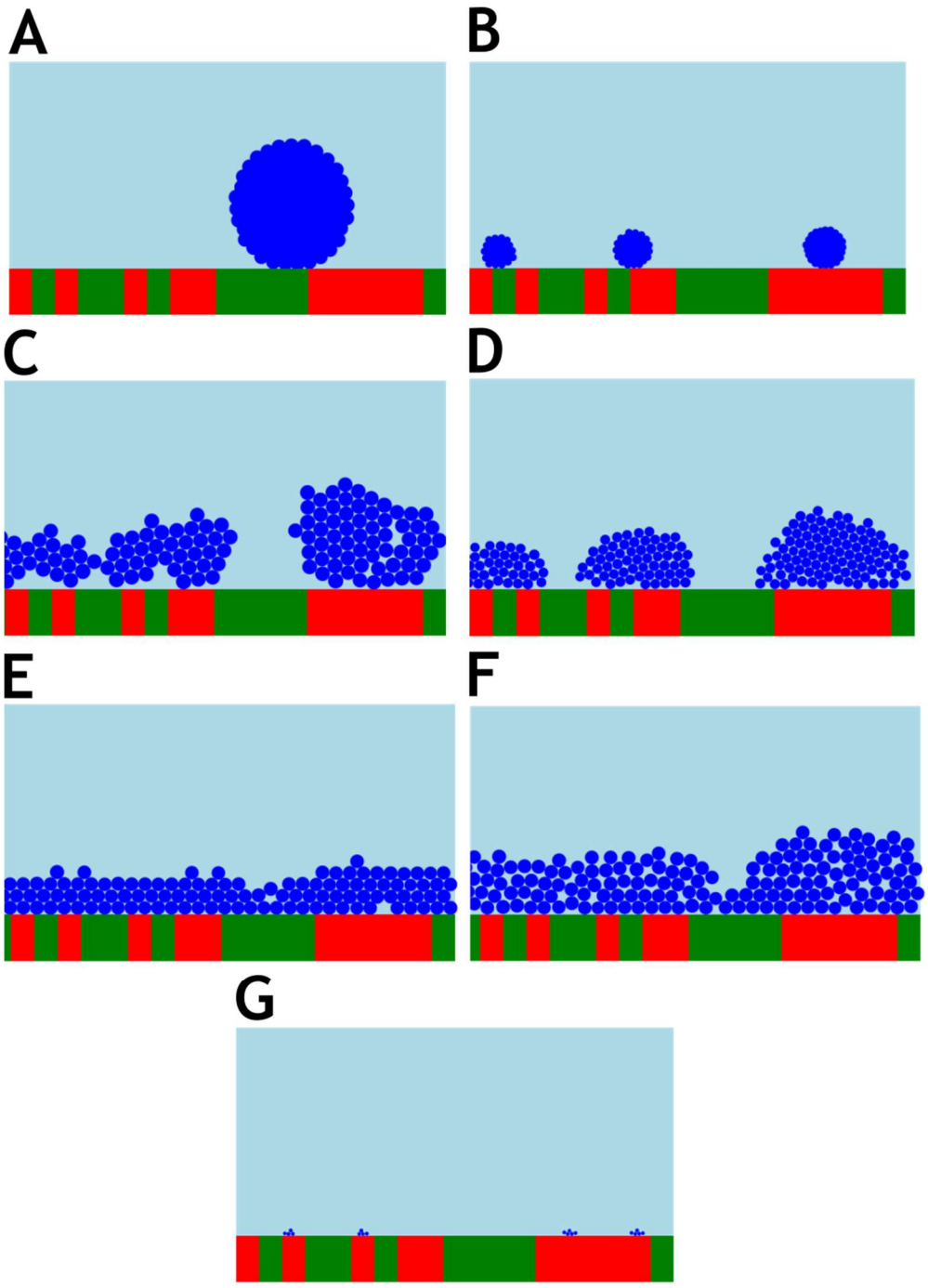

Figure 3.10 – The classes used for visual scoring of simulation results. Red cuboids represent goblet cells producing mucus, while green cuboids represent epithelial cells. Blue particles are mucus particles. Forms are A – large blobs, these simulations produced one or two large blobs of mucus bound together due to attractive forces. B - small blobs, these simulations yielded 1 to several smaller blobs of mucus particles also bound together by attractive forces. C – Geometric clusters, these simulations yielded aggregates of mucus particles that formed into crystal-like shapes formed

from tightly packed particles. D – Distinct mounds – Mucus clusters above goblet cells that failed to spread across the epithelium. E – Tight spread – Mucus spreads across the epithelium but particles are tightly packed together. F – Loose spread – Mucus spreads across the epithelium with a loose structure. G – Not enough mucus - Not enough mucus to determine the form.

Given the fact the "loose spread" structure fits the pattern provided by real mucus, further work was motivated by the attempt to understand what parameters could be used to achieve "loose spread" mucus behaviour in future simulations.

3.5.2 Effect of mucus particle size

There was a clear correlation between mucus particle mass and the number of timesteps completed in a simulation. This effect is almost certainly explained by the computational demand of simulating attraction forces for a large number of agents. The lower the mucus particle mass, the more mucus agents are produced by goblet cells in each time step, and the more time it takes for the simulation step to end.

There was also a very clear effect of the number of completed timesteps on the visual form of the mucosa, which tended to be "Not enough mucus" (Figure 3.11). All the simulations which were scored "Not enough mucus" did not complete more than one timestep. This was such a strong relationship that the effect of mucus particle mass would likely dominate any models of the effects of the various variables on simulation

outcomes, so the results from simulations which completed less than two timesteps were removed from the dataset. All further analysis was performed only on data from simulations with at least two completed time steps.

3.5.3 Linear Discriminant Analysis

The data from the remaining simulations were log transformed and then linearly rescaled to values between 0 and 1, before a linear discriminant analysis was performed, using the visual form as the predicted variable. This statistic creates a linear model combining the various predictor variables to explain the variation in the outcome – in this case, the mucus structure. This yields a number of linear discriminants – linear combinations of predictor variables that provide the best separation between different outcomes. The Linear Discriminant Analysis returned five linear discriminants, of which the first explained 81.6% of the variance in the outcome, the second explained 10.94% of the variance and the third explained 5.8% of the variance. This leaves only 1.65% of the variance explained by the last two linear discriminants put together.

As all predictor variables were rescaled to between 0 and 1, any large coefficients within the linear discriminants should represent a truly large effect of that parameter on the outcome. The largest coefficient in the first linear discriminant was the coefficient for the d_{sat} value for attraction between mucus particles, with a coefficient

of -5.357, while the second-largest coefficient was the F_{max} for the same attraction force, with a coefficient of -3.454. This suggests that low values for the saturation distance and low values for the maximum force of the attraction between mucus particles were needed in order to obtain high values of LD1, and thus a high likelihood of the mucus structure being loose spread or tight spread, which dominated the positive side of the LD1 axis. In other words, a weak, short-ranged interaction between mucus particles increased the chances of realistic mucus behaviour.

The coefficients and intercepts of the first three linear discriminants are given in Table 3.8 and the simulation results are plotted using the linear discriminants in Figures 3.12 and 3.13. These graphs show that Loose Spread and Tight Spread tend to cluster together, with high values of LD1. On the opposite end of the LD1 scale are Large Blobs and Small Blobs, with low values of LD1, and in the middle are Distinct Mounds and Geometric Clusters. Although Tight Spread and Loose Spread were generally close together, Loose Spread could be distinguished from Tight Spread in that it generally had lower values of LD2 and higher values of LD3.

Table 3.8 – The coefficients and intercepts of the first three linear discriminants of the analysis. Note that these values are the predictors of the rescaled logs of the original parameter values

	Mucus particle cohesive force		Mucus particle-epithelium adhesive force		Temperature	Mucus particle mass
	F_{max}	d_{sat}	F_{max}	d_{sat}		
LD1 coefficient	-3.454	-5.357	0.9000	0.003707	0.1224	1.087
LD1 intercept	3.400					
LD2 coefficient	-1.023	1.282	3.351	1.350	0.2262	-0.1126
LD2 intercept	-2.612					
LD3 coefficient	-2.536	1.946	-1.175	-0.2167	-0.02795	1.781
LD3 intercept	0.02119					

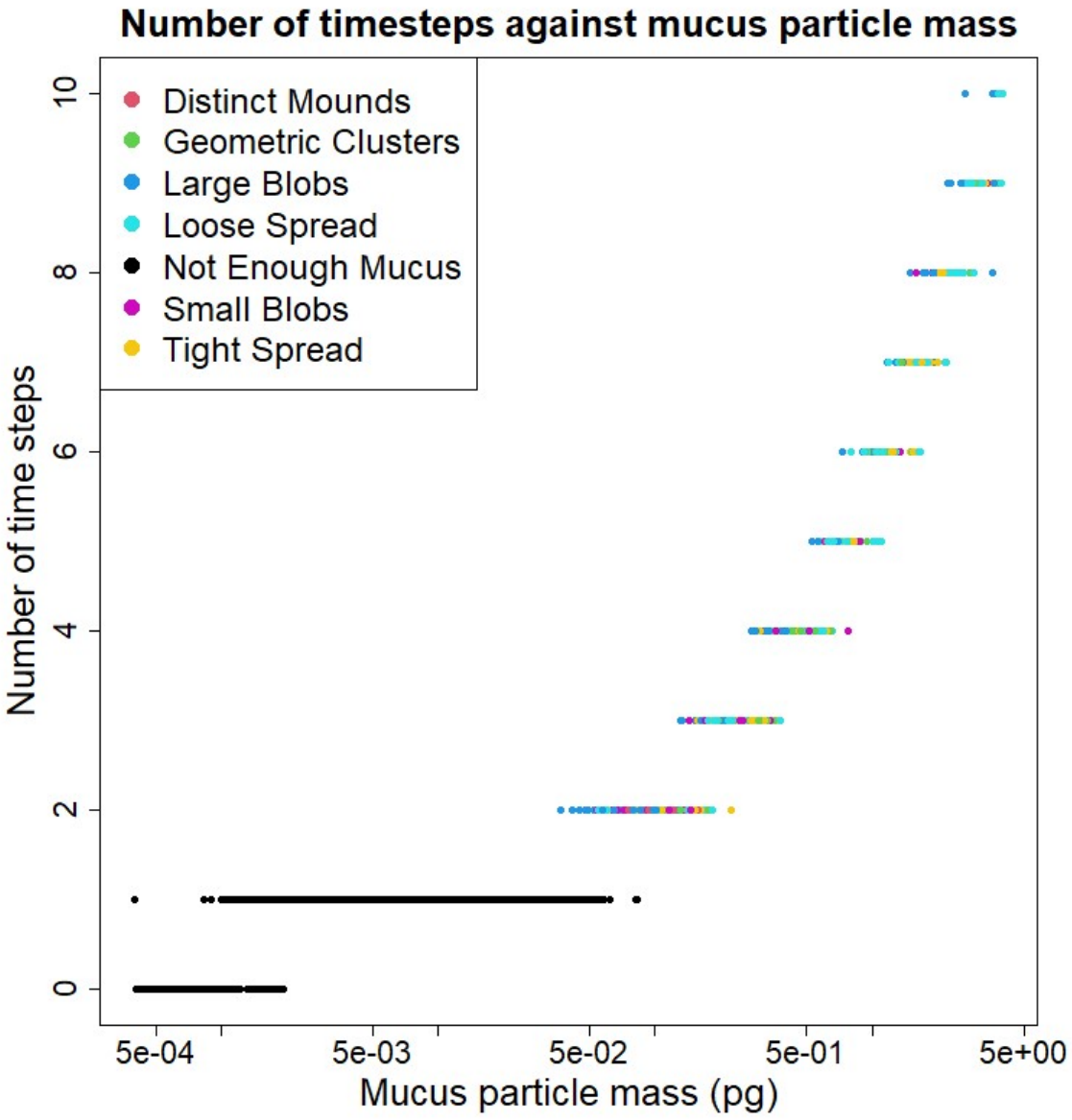

Figure 3.11 – Plot of the number of completed timesteps of the simulation, against the mucus particle mass. Colours show the mucus structure produced by the simulation.

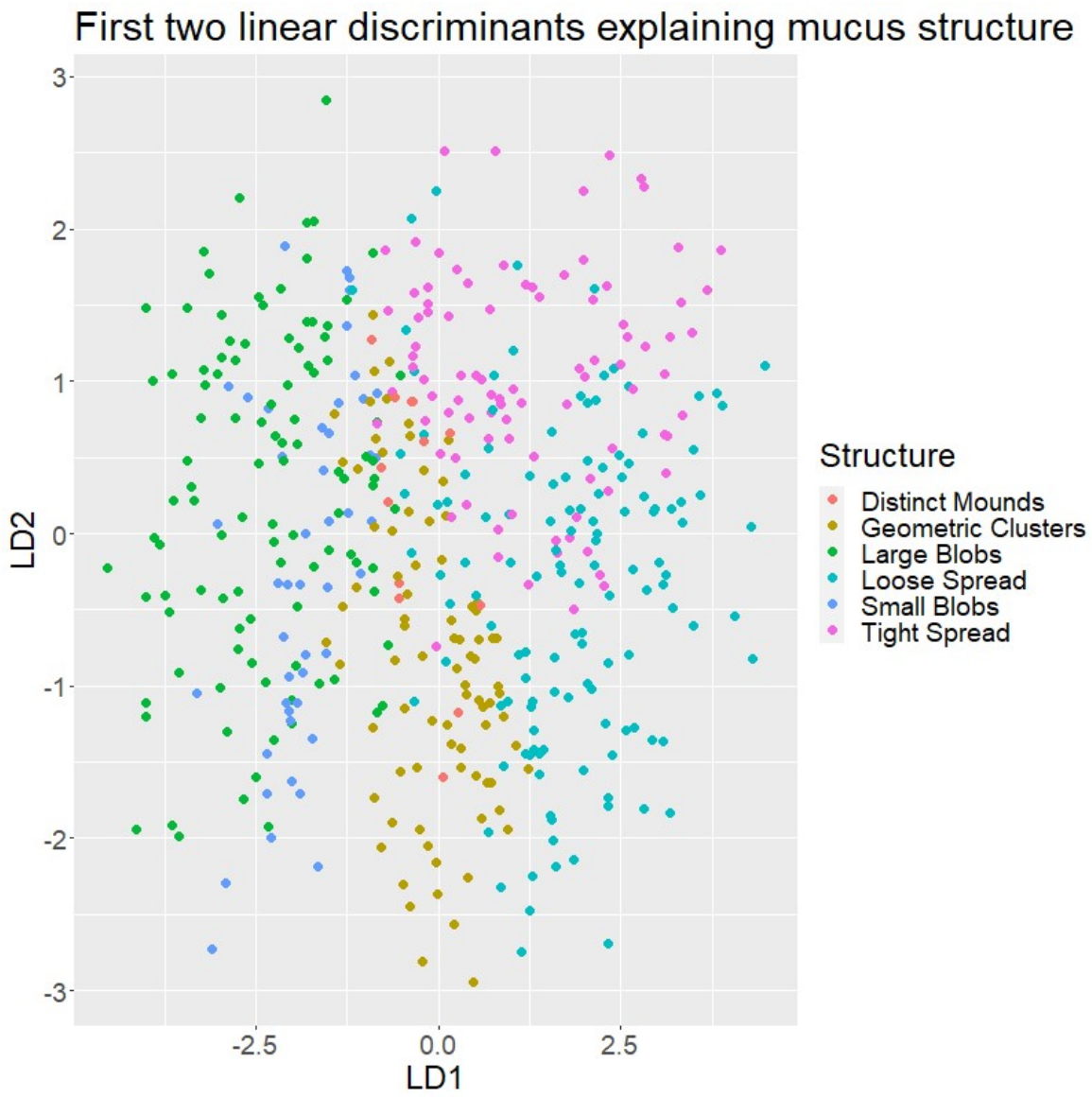

Figure 3.12 – First two linear discriminants from the linear discriminant analysis of the mucus simulations data. Groups could be clearly distinguished by these two linear discriminants alone.

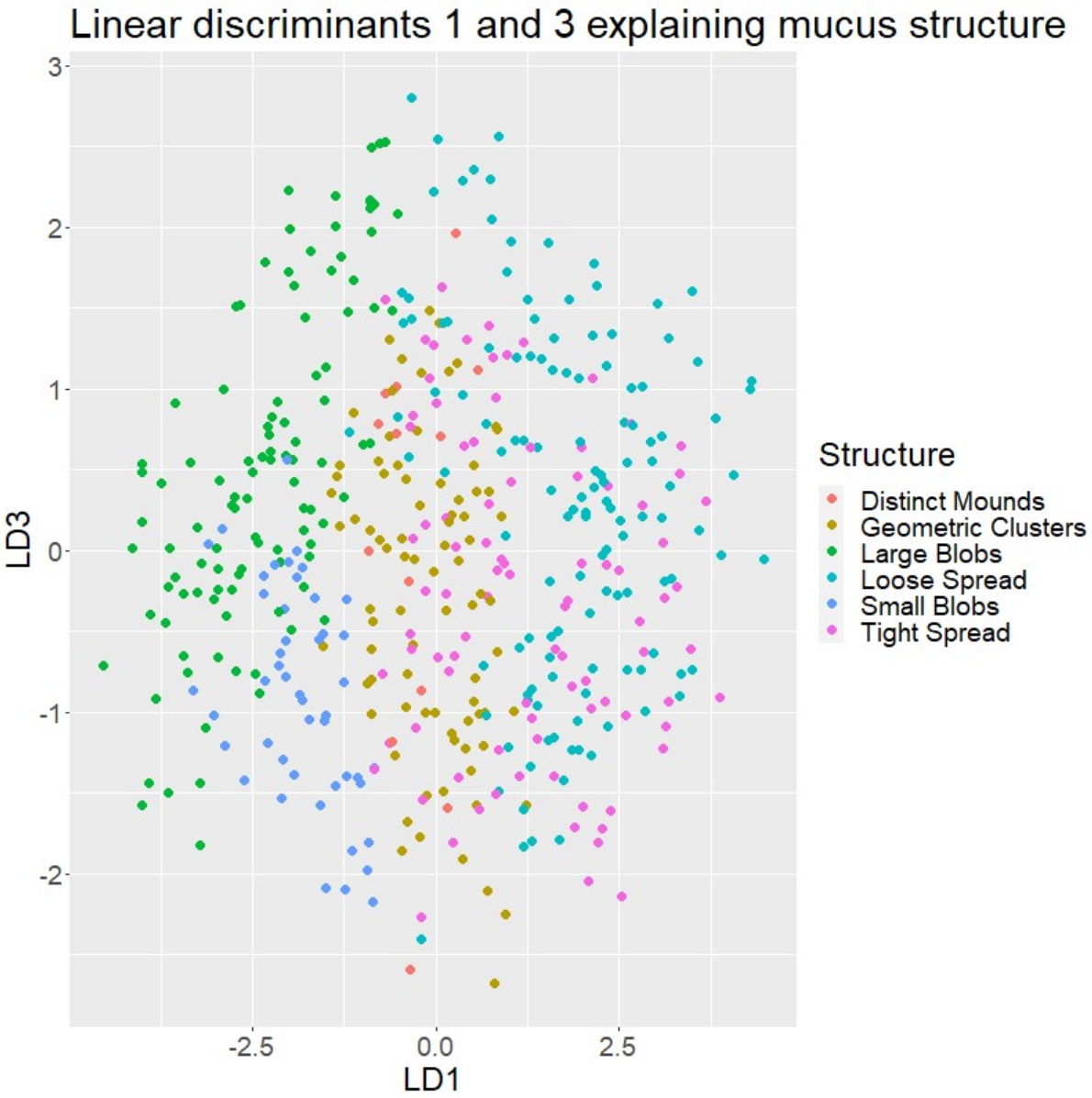

Figure 3.13 –First and third linear discriminants from the linear discriminant analysis of the mucus simulations data.

3.6 Discussion

The numerical tests performed in eGUT have demonstrated the platform's various numerical solvers and systems are sound and produce results that match analytical solutions. This has required a large amount of iterative testing, bug fixing, and fine-tuning of the solvers, and even a complete replacement of the reaction-diffusion solver system. This analytical testing is an essential and ongoing process in the development of a modelling platform, as having properly functioning and well-tested numerical components for the platform is essential for building trust between developers and users. The confidence that the basics of the model are working well is also important for carrying out further, more complex tests, as developers can be confident that any unexpected results from more complex results are not due to simple numerical errors in the platform's numerical algorithms.

The Benchmark 3 simulations further scrutinised outcomes from eGUT simulations of more complex biofilm systems, assessing the platform's ability to model microbial competition in a wastewater treatment setting. The steady state concentration results demonstrated that eGUT results were in fairly good agreement with other microbial modelling platforms with various different designs, from one dimensional continuum models to other multi-dimensional agent-based models. However, it is clear that the closest match to eGUT was iDynoMiCS 1, which is very similar to eGUT, sharing the same solver, same physical representation of the biofilm, and similar platform architecture. This is valuable to know, as it provides assurance that the systems within

eGUT work similarly to their counterparts in iDynoMiCS 1. However, removing iDynoMiCS 2.0 from the comparison of steady state concentrations to other models revealed a significant difference between eGUT and the other models in the low influent ammonium case. In addition, when looking at biomass, eGUT differs significantly from the other models in most BM3 variants, and clearly has a different total biomass in the biofilm. This may offer some explanation as to where the difference in steady state concentrations comes from. It is possible that as eGUT can hold a larger amount of total biomass than the other models, a larger proportion of inert biomass can build up in the eGUT simulations without altering the activity of the active biomass fractions. Further work will be necessary to adjust the agent density in eGUT to match that found in other models and analyse whether this reduces the differences observed here. However, the fact that eGUT's concentration results are so close to the other models despite differences in biomass density provide an early validation of eGUT's ability to model microbial competition in biofilm-like settings while the question of biomass density differences in the models remains unsolved. This is in part due to the fact that the density of the biofilm in the discrete biomass models is an emergent property rather than a settable parameter as in the continuum models. In the discrete models, the maximum density is set, while the mean density and density gradients are only constrained by this setting.

The BM3 work highlighted incomplete model and parameter descriptions and lacking availability of timeseries data and spatially resolved data to calculate statistics beyond

means or vertical gradients of biomass distribution in previous publications (Noguera and Picioreanu, 2004; Wanner et al., 2006) including our own (Lardon et al., 2011). For example, it was not explicitly stated that the oxygen concentration in the bulk was set to a fixed value, but keeping oxygen concentration constant was the only way to approximate results to previous models. The lack of vertical distributions of inert and active biomass made it very difficult to understand the differences in biomass densities between different models.

An independent and indirect verification of eGUT through comparison of NUFEB and iDynoMiCS 1 simulations, which we here compare with eGUT simulations, have demonstrated that NUFEB and iDynoMiCS 1 simulations of growing biofilms are matching fairly well. This gives further confidence in the results obtained with eGUT.

In order to develop a reproducible and reliable way to simulate mucus with particles, a pattern-based modelling approach was taken, as described by Grimm and Railsback (2012). Key characteristics of the desired outcome were defined, and a set of all the parameters that might influence the likelihood of the desired outcome were identified and varied across multiple orders of magnitude in a Latin Hypercube sampling of parameter space. This enabled the identification of some of the most important parameters in producing realistic mucus, and also revealed other important considerations, such as keeping mucus particles above a certain mass to avoid the simulations running extremely slowly.

The simulations that utilise attractive forces and stochastic movement are still very slow, taking longer to run than other simulations by orders of magnitude. This makes sense, given that attraction forces are recalculated with every step of the agent relaxation process, and unlike collision forces, attraction forces bring agents closer together, potentially increasing the number of interactions, neighbours and potential new collisions in the next relaxation step. This may lead to loops of collision and attraction whereby agents move towards and then away from one another repeatedly. Furthermore, when stochastic movement is also used, each step of resolving collision and attraction forces is followed by random stochastic movement, potentially bringing new pairs of agents close enough to incur collision or attraction forces. Further refinement to the process management of collision resolution, attraction and random movement is likely necessary to make this feature more usable. It may be valuable to decouple the collision force calculations from the attraction calculations and Brownian motion. Users can still model systems with mucus production by goblet cells without using attraction forces and stochastic movement, but may experience issues with mucus failing to spread across the epithelium to form a continuous cover.

CHAPTER 4 - *IN VITRO* VALIDATION USING THE MIMIC MODEL

This work was carried out with the assistance of several students who helped perform the experimental lab work. Preliminary lab work not included in the results presented here was carried out by Karanjot Sandhu and Gabriella Cooper. All practical work presented here was carried out jointly by myself, Mousomi Chakravorty and Migle Savukynaite. Initial conceptualisation of the work was done by me, Dr. Jan-Ulrich Kreft and Dr. Richard Horniblow. I planned and oversaw all practical work presented here, as well as carrying out all analysis and evaluation of the results. All modelling work in eGUT was carried out by me. Students were supervised by me, Dr. Jan-Ulrich Kreft and Dr. Richard Horniblow. Where data for figures and tables was not solely collected by me, a note is given in the legend specifying who did the work.

4.1 Introduction

4.1.1 Validation

As an *in silico* model, eGUT cannot model every complexity of the gut microbiome. This would be both practically impossible given our limited understanding of the gut and computational limitations, but also would be impossible to draw conclusions from, as one would not be able to isolate causes and effects. The eGUT modelling platform relies on a set of assumptions and simplifications of reality in order to simulate the gut in an efficient and interpretable way for a particular purpose. This means that many of the complexities that operate on the populations of microbes in the gut are either purposefully ignored or simply unknown, and that the predictive power of eGUT may suffer as a result.

In order to test whether eGUT-based models are able to provide realistic and accurate predictions of the population dynamics of microbes in the gut, the process of testing and validation is used to compare eGUT's results against results that are either analytically derived, or come from other gut models. Immediate comparison with the most complex models, such as mouse guts would be difficult, given that there are so many potential reasons for mismatches between eGUT and a mouse gut, from numerical errors in eGUT's solver, to problems with microbial growth models, to the numerous biological subtleties of the mouse gut microbiome that are either unknown, or would require additional research to be modelled in eGUT. It is therefore important to start by testing the simplest functions and processes within eGUT, and proceed

gradually to comparisons with more physically and biologically complex models. Having conducted numerical tests of eGUT and compared it to other *in silico* models using the Benchmark 3, validation against a simple *in vitro* model of the gut is the natural next step.

One particular *in vitro* model currently in development at the University of Birmingham is MIMic - Model of the Intestinal Microbiome - a model currently being developed by the Horniblow lab at the University of Birmingham Medical School. This is a chemostat-based model with monitoring and control capabilities, including pH control, temperature control, flow rate control and a connection to a gas supply for the maintenance of specific anaerobic conditions.

As the first instance of *in vitro* validation of eGUT, a simple, well-mixed bioreactor system in MIMic is used to validate a well-mixed (bioreactor) model in eGUT. Both systems are used to model competition between *E. coli* Nissle 1917 (*Ec*N) and *Salmonella enterica* serovar Typhimurium (*S*Tm).

While this only tests the well-mixed chemostat module of the eGUT modelling platform, it is the first validation of eGUT to be carried out using an *in vitro* model. Testing a simple system first enables us (the developers of eGUT) to ensure that the simplest parts of the model work as expected and reflect results obtained in the lab, before testing more complex systems. Furthermore, modelling the system initially in a non-spatial system allows us to investigate whether the colonisation resistance of *Ec*N can exist independent of spatial structure, thus help to tease apart different

aspects of the competition between *Ec*N and *S*Tm. Future work can build on the developments described here to develop more complex validations of eGUT, especially the mucosa module, and to expand on the understanding of *Ec*N and *S*Tm gained in this study.

4.1.2 Model Species

Escherichia coli Nissle 1917 (*Ec*N) is a probiotic strain of *E. coli* that is licensed as a medicinal product in Germany and several other European countries, under the trademark "Mutaflor" (Sonnenborn and Schulze, 2009; Mutaflor, n.d.). It was discovered by Alfred Nissle, a German physician and microbiologist, in 1917, in the faeces of a German soldier who survived a Shigella outbreak while stationed in the Balkans (Sonnenborn, 2016). Upon isolating the strain, Nissle found that the strain had strong antagonistic properties, inhibiting the growth of other taxa within the Enterobacteriaceae, including *Salmonella* (Sonnenborn and Schulze, 2009). More recent work on *Ec*N, however has shown that the strain's genome does contain the pathogenicity island, *pks*, which encodes proteins necessary for the synthesis of colibactin, a genotoxin which causes DNA damage. The synthesis of colibactin by *Ec*N has also been demonstrated in a mouse model (Nougayrède et al., 2021).

Salmonella enterica serovar Typhimurium (*S*Tm) is an enteric pathogen that causes diarrhoeal disease in humans and other animals. It is typically a food-borne pathogen

and is one of the most common causes of food-poisoning outbreaks across the world (Fàbrega and Vila, 2013). Though *S*Tm does not cause systemic infection (typhoid fever) in humans, it does so in mice, making it an important model for human typhoid fever.

Since its discovery by Alfred Nissle, studies have continued to find inhibitory effects of *EcN* on various pathogenic Enterobacteriaceae strains (Massip et al., 2019; Reissbrodt et al., 2009; Sassone-Corsi et al., 2016). Many of the mechanisms discovered so far for this probiotic effect involve competition for iron. For example, Deriu et al., (2013) found that *Ec*N possesses a high-affinity iron uptake pathway not possessed by *S*Tm, that has a higher affinity to iron than host-excreted lipocalin 2, allowing it to outcompete *S*Tm in the low-iron environment of the inflamed gut. Other iron-related antagonistic mechanisms involve microcin production (Massip et al., 2019; Sassone-Corsi et al., 2016). In an iron-limited environment, Enterobacteriaceae compete for iron by producing iron-sequestering siderophores. However, *Ec*N can produce microcins, which mimic siderophores, containing a site that binds to siderophore receptors on other Enterobacteriaceae. Once these microcins are taken up, they can exhibit toxic effects, such as forming pores in the cell membrane, killing the cell (Huang et al., 2021).

Although the above mechanisms relating to iron competition almost certainly play a role in the probiotic effect of *Ec*N, there is evidence for other mechanisms, including competition for carbon source galactitol (Eberl et al., 2021), electron acceptors, oxygen and tetrathionate (Litvak et al., 2019), and the micronutrient zinc (Behnsen et al., 2017)

as well as the production of colibactins (Massip et al., 2019). Furthermore, the dynamics due to chemical interactions cannot be easily separated from those that are reliant on spatial structure and neighbour-neighbour interactions, as most of the above studies were carried out in mice, making small-scale phenomena such as the interactions between neighbouring microbes difficult to ascertain. One study carried out in continuous flow cultures has shown that several E. coli strains can exert antagonistic effects on *S*Tm, suggesting that spatial structure may not be a key factor in colonisation resistance (Ushijima and Seto, 1991). However, this experiment was not carried out with the *Ec*N strain.

Using a simpler system, such as a well-mixed chemostat allows individual metabolic interactions to be tested in a tightly controlled setting, where spatial structures are no longer relevant and concentrations of nutrients are experienced uniformly by all cells. This simple type of system also provides an opportunity to validate the simplest aspects of the eGUT model, and its ability to accurately model a chemostat.

In the human colon, an important energy source for microbes comes from the degradation of mucins – the molecules that make up the bulk of mucus. Mucins are large glycoproteins, with a polypeptide backbone, attached to which are glycan chains comprised of various monomers, mostly O-linked (Derrien et al., 2010). N-acetyl-glucosamine (GlcNAc) is an amino-sugar and one of the chief components of human mucins (Robbe et al., 2004). It is thought to be an abundant carbon and energy source for microbes in the human colon (Bunesova et al., 2018).

In this investigation, *Ec*N and *S*Tm were each grown in chemostats, with GlcNAc as the energy source, in cross-invasion studies, with one species as the resident in each chemostat, and one as the invader added at a later time. Populations of the two species were monitored over time. The aim of this experiment was two-fold:

1. To investigate the probiotic effect of *Ec*N and its antagonism of the growth of pathogenic Enterobacteriaceae. If *Ec*N was able to outcompete the *S*Tm strain, this would suggest that the mechanism of antagonism is, at least in part, independent of the gut setting.

2. To test eGUT's ability to model the system in the MIMic model. If the same outcome could be observed in eGUT, this would help to validate the accuracy and usability of parts of eGUT, specifically, the simulation of well-mixed systems as dimensionless compartments.

In order to model the same system in eGUT, experiments were carried out to estimate the growth parameters of the two species, and an eGUT Model of *E. coli-Salmonella* Competition (eMESC) was developed to emulate the MIMic experiments.

4.2 Materials & Methods

4.2.1 Bacterial strains

Escherichia coli Nissle 1917 stock was kindly given to us by Prof. Chris Thomas as a glycerol stock, with permission from Ardeypharm GmBH, Herdecke, Germany. *Salmonella enterica* serovar Typhimurium LT2 was kindly provided by Dr Maria Laura Ciusa in Prof. Laura Piddock's lab, University of Birmingham.

In order to create fresh stocks, both strains were grown on LB agar plates and then individual colonies were picked with a plastic loop and spread on fresh LB plates to ensure that isolated colonies on the resulting plate would be pure. The colonies on this new plate were then used to create new glycerol stocks with cryobeads. These cryobeads were used to create new growing stocks when starting new experiments by inoculation on LB plates.

To test the identities of our bacterial stocks, we performed API 20E tests (bioMérieux UK Ltd., Basingstoke) before each replicate of our experiment. In each case, the test results matched those expected from our strains.

4.2.2 Media

The medium used for liquid culture, and as the chemostat inflow medium in the MIMic experiments, was CP medium, as originally described in (Plugge, 2005),

supplemented with L-threonine and N-acetyl glucosamine (GlcNAc), as described in (van der Ark et al., 2018) (Table 4.1) and in some cases, a further addition of 20 mg/l sodium dithionite after autoclaving to maintain the anaerobic state. Resazurin was also added to the medium in order to monitor oxygen levels in the medium. CP medium ingredients other than GlcNAc and the vitamin solution were combined and autoclaved at 121°C for 15 minutes. After autoclaving, heat-sensitive ingredients in the medium – GlcNAc and the vitamin solution, were sterilised by filtering through 0.2 μm membrane filter and added to the autoclaved medium in a laminar flow hood to maintain sterility. When preparing medium for the MIMic devices, sodium dithionite was also added at this stage.

Table 4.1 – Ingredient list for the final modified CP medium, including L-threonine, GlcNAc and resazurin. For the full lists of components in the vitamin solution and the acid trace element solution, see Plugge (2005).

Component	**Concentration**
L-threonine	6 g l^{-1}
N-acetyl glucosamine (GlcNAc)	5.53025 g l^{-1}
Potassium dihydrogen phosphate (monobasic)	0.4 g l^{-1}
Disodium phosphate	0.53 g l^{-1}
Magnesium chloride heptahydrate	0.109 g l^{-1}
Calcium chloride	0.11 g l^{-1}
Sodium chloride	0.3 g l^{-1}
Resazurin	0.5 mg l^{-1}
Sodium bicarbonate	4 g l^{-1}
Sodium sulphide	0.25 g l^{-1}
Acid trace element solution	1 ml l^{-1}
Vitamin solution	1 ml l^{-1}

Individual liquid cultures were grown in CP medium in an anaerobic cabinet (Don Whitley Scientific Anaerobic Workstations) in 50 ml polypropylene Falcon tubes. Falcon tubes were left with the lid loose inside the cabinets in order to allow oxygen to diffuse out of the medium before inoculation.

For agar plate cultures, lysogeny broth (LB) agar, MacConkey agar and xylose lysine deoxycholate (XLD) agar were all used. For all plates, 25 ml of liquid agar were poured in a class II biological safety cabinet to minimise risk of contamination of the plates. They were stored wrapped in cling film or parafilm in a refrigerator for a maximum of one month. LB agar was used for general propagation of plate cultures and for plating of liquid cultures containing just one microbial strain. MacConkey and XLD agars were used to differentiate the two strains in order to facilitate plate counting during the MIMic experiments. On MacConkey, the two strains can be distinguished by colour, while on XLD, only *S*Tm can grow (Table 4.2).

Table 4.2 - Plate culture media and their characteristics.

Medium	**Appearance of *E. coli* colonies**	**Appearance of *S. enterica* colonies**	**Use in experiment**
LB (lysogeny broth)	Pale yellow	Pale yellow	Only used before cross-invasion, due to inability to distinguish *Ec*N from *S*Tm
MacConkey	Red-pink	Yellow	Used both before and after cross-invasion. Able to distinguish *Ec*N and *S*Tm by colour
XLD	No growth	Black	Used both before and after cross-invasion. Distinguishes species by inhibiting growth of *Ec*N.

4.2.3 MIMic

The MIMic apparatus is composed of a borosilicate glass bioreactor (Electrolab Biotech Ltd, Tewkesbury, UK) topped with a stainless-steel lid. Five polyether ether ketone (PEEK) tubes traverse the lid, of which one is connected to the gas supply, one to the medium supply and one to the waste outlet. The other two are sealed off with short, clamped lengths of Masterflex tubing. The lid also contains a port for the insertion of a pH meter, which is connected to a pH control/monitoring unit. A further port with a stainless-steel screw-sealed stopper can be opened in order to take samples and inoculate the bioreactor. One final attachment to the lid is a hollow borosilicate glass tube which also traverses the lid, and into which a temperature probe is placed. The temperature probe is attached to a temperature controller, which powers a heating jacket in order to maintain a constant temperature of 37°C in the device.

The PEEK tubes in the lid of the MIMic device are connected to the gas supply, medium inflow and waste outflow by Masterflex platinum-cured silicone tubes with internal diameter 3.1mm. In the case of the medium inflow and waste outflow, these Masterflex tubes are connected to Marprene tubes, a thermoplastic peristaltic pump tubing (Watson-Marlow) via Masterflex polypropylene Luer connectors. While most of the lid's PEEK tubes only extend around 2 cm into the chemostat headspace, the outflow tube extends further into the device to a height corresponding to a liquid

volume of 250 ml. Fluid is removed by the outflow tube whenever it rises above this level, setting a constant volume within the bioreactor.

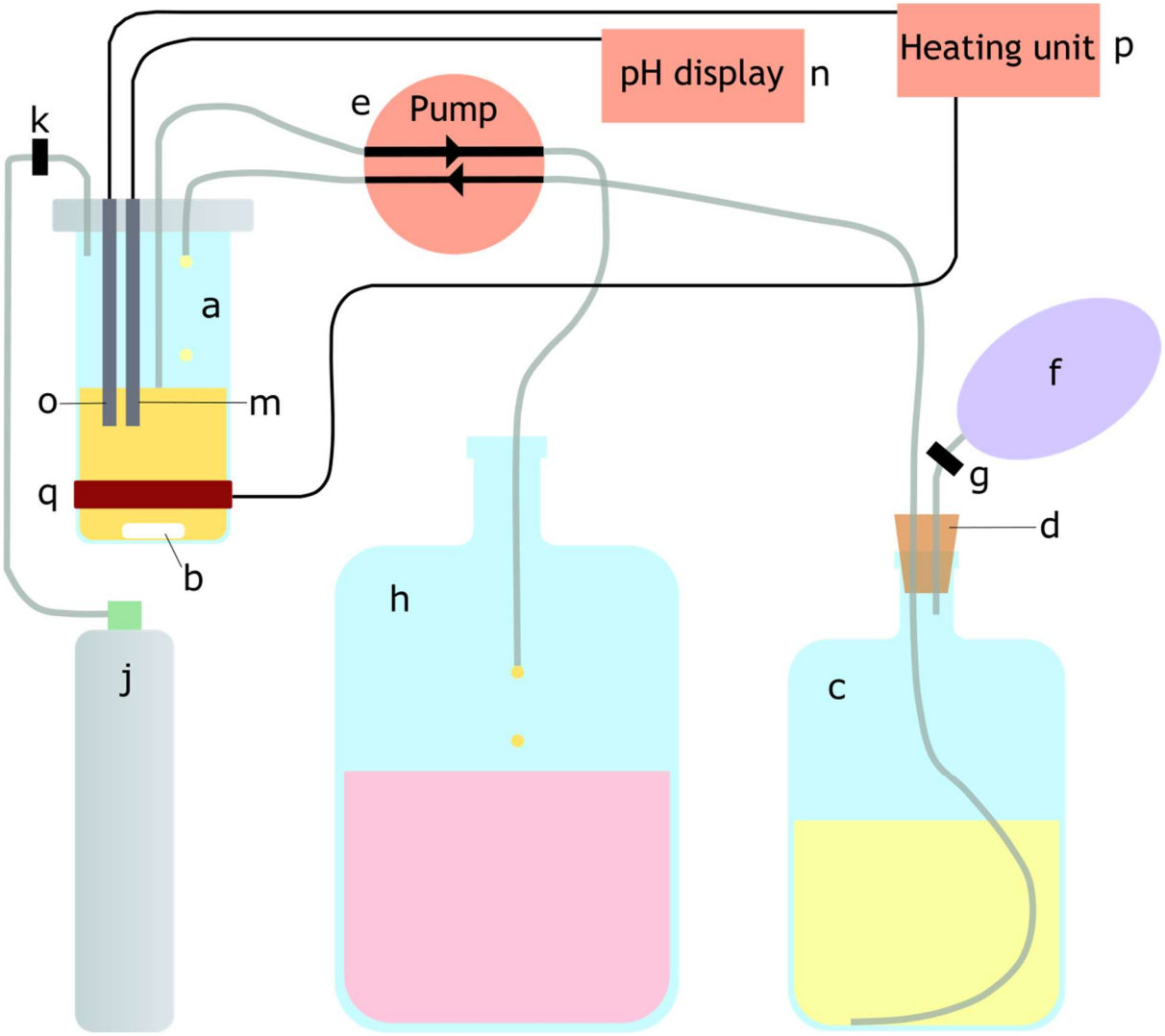

Figure 4.1 - Bacteria were grown in bioreactors (a) in a liquid volume of 250 ml, continuously stirred with a magnetic stirrer (b). Fresh medium was delivered to the bioreactors from a medium bottle containing 10 L of CP medium (c), which was sealed with a rubber bung (d). Medium was pumped from the medium bottle into the bioreactor by a pump (e). Headspace left by the removal of medium is replaced by nitrogen gas from a balloon (f), passing through a 0.2 μm filter (g). Liquid culture is removed from the bioreactor by a rigid PEEK tube which rests at a set height above the bioreactor base, maintaining the correct liquid volume within the bioreactor. This removal is driven by the pump (e), which pumps excess culture

liquid into a waste bottle (h), containing Virkon. A positive pressure is maintained in the headspace of the bioreactor by a 95:5 mix of N2 and CO2, supplied by a cylinder (j), passing through a 0.2 μm filter (k). A pH meter (m) connects to the bioreactor lid and extends into the culture, measuring the culture pH. It is wired to a pH display unit (n). A thermometer (o) connects to the bioreactor lid and extends into the culture, measuring the culture temperature. It is wired to a heating unit (p), which displays the temperature and controls a heating jacket (q) to maintain a temperature of 37°C. Thermoplastic marprene tubes from Watson Marlow were used inside the pump, with wider tubes used in the outflow than the inflow to ensure the liquid volume in the bioreactor remained at 250 ml. This is represented in the diagram as black lines with arrows. All other tubing in the system is represented by grey lines. Wired connections are represented as thin black lines.

Marprene pump tubing for the inflow had an internal diameter of 1.52 mm, while pump tubing for the waste outflow had an internal diameter of 2.38 mm. The difference in bore size ensured that in the case of minuscule differences in flow rate between tubes, the inflow rate would never be higher than the outflow rate, and the device would not fill up with liquid culture. It also reduced clogging of the outflow. For a detailed diagram of the MIMic setup, see Figure 4.1.

4.2.4 MIMic Experimental Procedure

For the cross-invasion experiments, the bacteria were grown in the MIMic devices in a well-mixed liquid volume of 250 ml which was stirred constantly at a rate of 120 rpm

using a magnetic stirrer. The temperature probe and jackets were used to maintain a temperature of 37°C For each device, 10 L of CP medium was stored in a large glass medium bottle stoppered with a rubber bung and connected to the medium inflow tubing using Masterflex tubing.

The device was connected to a gas supply of 95% Nitrogen, 5% CO_2 at a slight pressure of ~50 mbar. This maintained an anaerobic environment in the chemostats and a positive pressure when opening ports. The maintenance of the anaerobic state of the system was confirmed by the resazurin in the medium which stayed colourless inside the devices. However, the resazurin in the medium bottles turned pink over time, suggesting that the seal provided by the rubber bung lids were not completely airtight. The presence of the CO_2 in the gas phase, along with the sodium bicarbonate in the liquid phase acted as a pH buffer system, and the pH remained at approximately 6.4 for the duration of the experiments after inoculation without pH regulation.

During each experimental replicate, two MIMic devices were set up. Prior to inoculation, the whole bioreactor including the medium bottle, was connected and the tubes clamped to prevent medium flowing through the system. The whole system was then autoclaved at 121°C and 1 atm for 15 minutes.

The whole system was then resealed with clamps and placed into another laminar flow hood where it remained for the duration of the experiment. The media feed tubes were attached to the peristaltic pump, and the sealed tubes for gas input were connected to a gas cylinder via a sterile 0.2 μm membrane filter. Finally, clamps were

removed, and the pump, magnetic stirrer plates and temperature control and pH monitor were turned on. pH was monitored but not controlled throughout the experiment.

After the pump was started, the devices were allowed to fill up with medium, and monitored for any sign of contamination. After 24 hours, samples were taken from the devices, and two 50 µl samples of undiluted medium from each device were spread onto LB agar plates and incubated at 37°C to check whether the devices remained sterile.

If no colonies were observed to grow on the sample plates, and no turbidity was observed inside the devices, each device was inoculated with one of the two strains. From this point, daily sampling took place, including OD measurements, plate counting, and monitoring of the pH. Once the system was judged to have reached a steady state population (generally after ~7 days), cross-invasion was carried out and daily sampling continued until the medium was spent.

4.2.5 Plate Counting

During both the MIMic sampling procedure and the OD-CFU calibration experiments, agar plates were inoculated with samples diluted to various degrees. Samples were first diluted in series of 10-fold dilutions, before plating. In the MIMic experiment, samples taken before the cross-invasion were plated on LB agar, while samples taken

after the cross-invasion were plated on MacConkey and XLD agar (Table 4.2). Two different types of agar plating were carried out. On "spread plates", 50 µl of a suitably diluted sample were spread on an agar plate using a Drigalski spreader, while on "dot plates", a series of different dilutions were spotted together, by placing 5 µl sample "dots" into different regions of the plate (Miles and Misra method but with 5 µl instead of 20 µl).

For each dilution, three of these dots were placed (Figure 4.2). The four dilution levels chosen for each sample were chosen based on OD values.

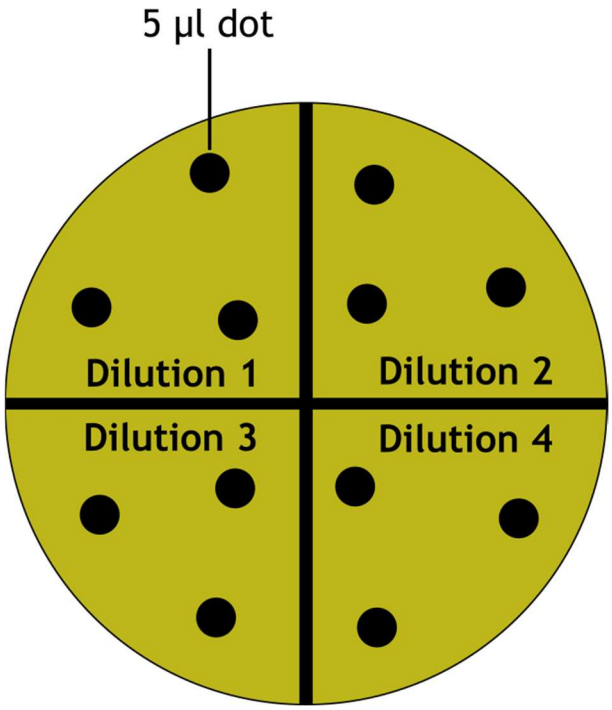

Figure 4.2 - A schematic of the procedure for Miles and Misra "dot plates". Three 5 µl spots of each of four different dilutions are pipetted onto the plate, with each quadrant containing dots from a particular dilution.

4.2.6 Peristaltic Pump

A 12-channel Watson-Marlow 205S peristaltic pump was used to power both the medium inflow and waste outflow of the MIMic devices. A pumping speed of 2 RPM was used, which corresponded to an inflow rate of 250 ml day^{-1}, providing a retention time (complete volume change) of 24 hours.

4.2.7 Inoculation

Three days before planned inoculation of the MIMic devices, the cryobead stocks were used to inoculate LB agar plates. The following day, single colonies were picked from these plates and spread onto new LB agar plates. The following day, single colonies from these plates were inoculated into liquid CP medium and then grown overnight in an anaerobic chamber. When the devices were ready for inoculation, 10 ml of *Ec*N or *S*Tm overnight cultures were used to inoculate a MIMic device.

4.2.8 Cross-invasion

Cross-invasion was used to introduce an "invader" into each of the MIMic devices. For each of the two strains, a small amount of the other strain was introduced. A 2.5 ml inoculum (1% of the volume) was taken from each device and inoculated into the other device.

4.2.9 Daily sampling and growth monitoring

Each day, a 5 ml sample was taken from each chemostat. The OD, pH and temperature were recorded on a daily basis to ensure that the culture was not undergoing rapid, unexpected changes.

4.2.10 OD Measurements

Optical density was measured at three wavelengths (400 nm, 750 nm and 900 nm) using a spectrophotometer, using pure CP medium as a blank. These wavelengths were chosen in order to avoid the absorption peak of the pink resorufin, the reduced form of resazurin, around 600 nm (Bueno et al., 2002), given that resazurin was an ingredient in the CP medium. Any sample with an OD above 0.6 in any wavelength was diluted using deionised water by a factor of 2. This was repeated as needed until the highest OD was below 0.6.

4.2.11 Growth curves

To establish the growth characteristics of the two microbial species, growth in an anaerobic environment was monitored by growing *Ec*N and *S*Tm in CP. Samples of the growing liquid cultures were taken at regular intervals, with shorter intervals when the cultures appeared to be in exponential growth. The OD of the samples was measured according the "OD Measurements" section. For one of the growth curves,

plate counting was also carried out in order to establish a calibration between OD and CFU/mL on each type of agar.

Growth curves were always started early in the morning and continued for approximately 12 hours. Further measurements were taken the following day, approximately 24 hours after the start, in an attempt to observe any potential peak OD that had not been reached during the previous day.

4.2.12 Yield estimations

In order to estimate the biomass yield of *Ec*N and *S*Tm in CP medium, larger cultures were grown in CP medium in an anaerobic chamber and the OD monitored to ensure that the culture was in exponential growth phase during sampling, when cells are in balanced growth and have a constant composition. This ensured that the ratio of cell number to biomass should remain constant. During exponential phase, two 50 ml samples were taken from each culture at two time points, and these were placed into 50 ml Falcon tubes that had previously been dried in an incubator at 80°C and weighed. These samples were then centrifuged at 7000 *g* for 30 minutes. The supernatants were decanted, and the cells were resuspended in the volatile ammonium acetate buffer (50 mM, pH 6.5). The cells were once again centrifuged at 7000 *g* for 30 minutes, and the supernatant decanted. The tubes were then returned to the incubator at 80°C for two days, and repeatedly weighed until their weights were constant. The difference in mass after complete drying was used to estimate the mass

of the bacterial pellet. As ammonium and acetate are both volatile, these are lost during the drying process and not contribute to the weight as salts of other buffers would. A set of control Falcon tubes underwent the same drying regime but did not receive samples, nor undergo centrifugation. Tubes were continually weighed until the average change in experimental tube mass was not significantly different from zero according to a 1-sample t-test. This point was reached after 28.33 hours of drying post-sampling.

4.2.13 Statistics

OD-CFU calibration

In order to establish the ratio between the various OD measurements at different wavelengths of light and the colony forming units (CFU) per mL, samples from one of the growth curves were taken and plated at a range of dilutions after OD measurements, as described in the section "Growth Curves". Linear regression was performed in R on these data, with OD as the predictor variable, and CFU as the outcome. As the linear model should theoretically pass through the origin (0,0), if intercepts in the default linear regression were not significant, then a second linear regression was carried out with a fixed (0,0) intercept for the final slope estimate.

Growth parameter estimation

Growth curves were fitted to logistic growth models using non-linear regression in R x64 4.1.0, using the nlstools package. The logistic growth equation is

$$P = \frac{KP_0 e^{\mu_{max} t}}{K + P_0(e^{\mu_{max} t} - 1)}$$

Where P is the population size, K is the carrying capacity, P_0 is the initial population size at time 0, μ_{max} is the maximum specific growth rate and t is time. Non-linear regression was carried out, with initial values for P_0, μ_{max} and K estimated by inspection of the curves.

In order to estimate the value of various parameters of growth kinetics, including the maximum growth rate, μ_{max} and carrying capacity, K, the growth data from growth curves was first converted from OD measurements to population estimates using the OD-CFU calibration, and then non-linear regression was carried out using a logistic growth model. Since this model does not include a model of the death phase of batch cultures, there was a chance that measurements on day 2 of the growth curve would be taken during the death phase, and thus lead to inaccurate parameter estimates. Thus, only measurements on day 1 of the growth curve were used in the initial non-linear regression. This avoided any artefacts arising from batches reaching carrying capacity and then beginning the death phase before another measurement was taken. However, in some cases, these initial non-linear regressions predicted carrying

capacities lower than measurements from day 2. In this case, the regression was repeated including any results from day 2 up to the maximum OD.

4.2.14 eGUT Model of *Ec*N-*S*Tm Competition

To model the MIMic system in eGUT, the eGUT model of *Ec*N-*S*Tm competition (eMESC) was set up in eGUT with dimensionless (well-mixed) compartment of volume 250 ml, and an inflow rate of 250 ml day^{-1}. The inflowing medium contained GlcNAc at a concentration of 5.53025 g l^{-1} and Monod equations were used to model growth kinetics for each strain, based on the results from the growth parameter estimation. Initial masses of each species were based on the estimated steady state CFU values at the point of cross-invasion, and cell masses estimated from the literature (see section 4.3.3).

eMESC was run with timesteps of 1 hour and cell numbers were recorded each hour.

4.3 Results

4.3.1 OD-CFU Calibration

For each pair of wavelength and species, two linear models of the relationship between CFU and OD were constructed in R, with CFU as the outcome, and OD as the predictor. One with an intercept that was free to vary, and one with a fixed intercept at (0, 0). In all cases, the model with the fixed intercept had a higher R^2 value than the one without a fixed intercept (Table 4.3). Furthermore, in the linear models with fitted intercepts, the intercepts were never significant. As a sample containing no cells is equivalent to the blanks used in the OD measurements (pure CP medium), the OD of a sample with a population of 0 should theoretically be 0. Thus, the models with fixed intercepts were used from this point onwards.

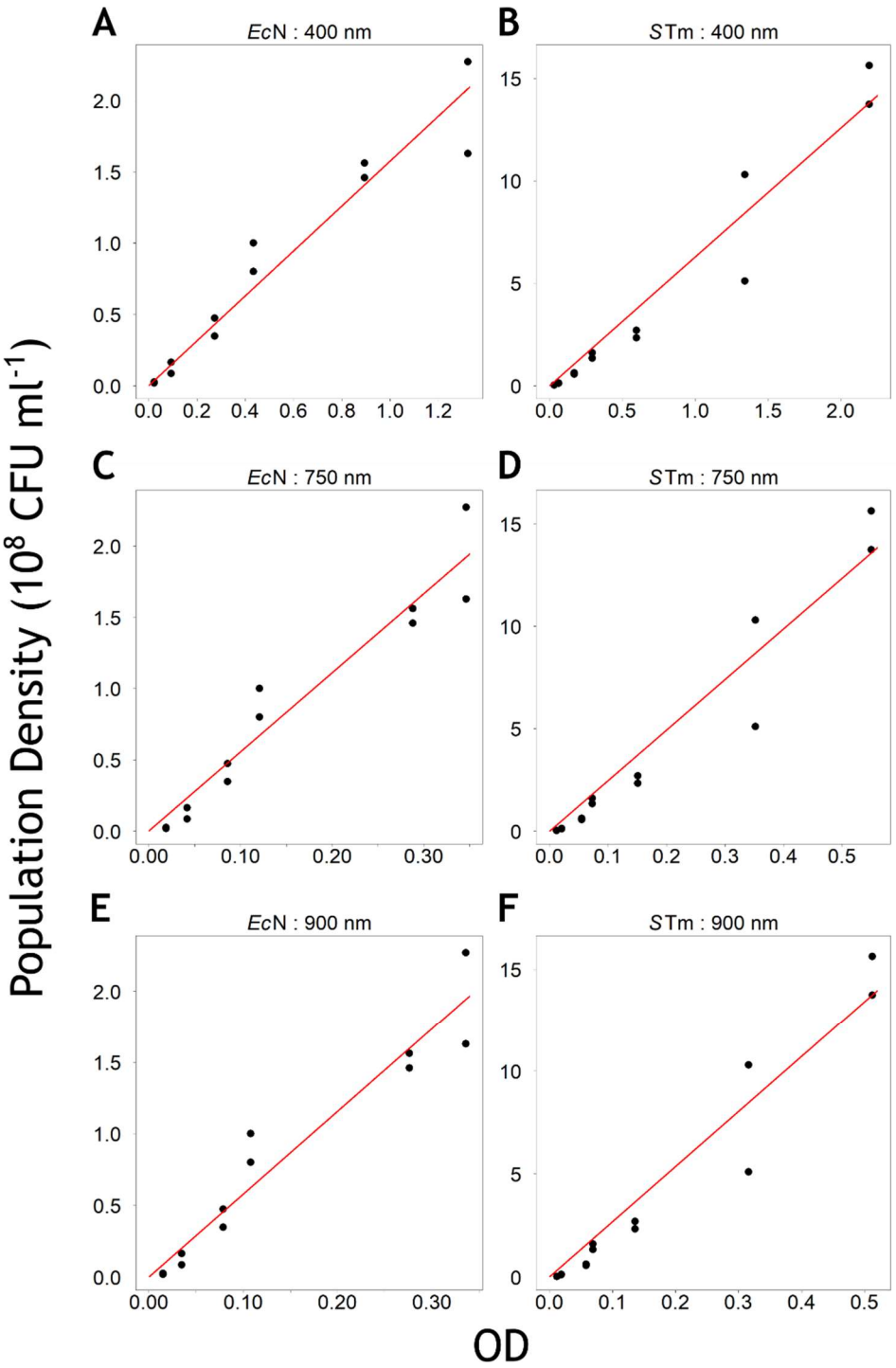

Figure 4.3 – OD-CFU calibrations based on growth measurements of *S*Tm and *Ec*N cultures growing in CP medium in an anaerobic chamber. Regression lines were calculated using R with a fixed intercept at (0, 0). Calibration regressions were carried out for OD measurements at three light wavelengths – 400nm (A, B), 750nm (C, D) and 900nm (E, F), and for both *Ec*N (A, C, E) and *S*Tm (B, D, F). Numerical results are given in Table 4.3. *Data in this figure collected by the author, Mousomi Chakravorty and Migle Savukynaite.*

Table 4.3 – Results of linear regression of OD-CFU calibration data shown in Fig. 4.2. OD measurements were taken at wavelengths of 400 nm, 750 nm and 900 nm and CFU was measured using plate counts. The standard linear model had a variable intercept, while the single regressor model had a fixed intercept at (0,0). *Data in this figure collected by the author, Mousomi Chakravorty and Migle Savukynaite.*

	Species	***Ec*N**			***S*Tm**		
	Wavelength (nm)	**400**	**750**	**900**	**400**	**750**	**900**
Standard linear model	**Intercept (10^8 CFU ml^{-1})**	0.05	-0.04	0.001	-0.61	-0.75	-0.79
	Intercept p-value	0.588	0.689	0.987	0.199	0.136	0.11
	Slope (10^8 CFU ml^{-1} OD^{-1})	1.52	5.70	5.78	6.69	26.68	29.07
	Slope p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2	0.94	0.94	0.93	0.95	0.95	0.95
	Adjusted R^2	0.93	0.93	0.92	0.94	0.94	0.95
Single regressor model	**Slope (10^8 CFU ml^{-1} OD^{-1})**	1.58	5.55	5.79	6.29	24.68	26.80
	Slope p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	R^2	0.97	0.97	0.97	0.96	0.96	0.96
	Adjusted R^2	0.97	0.97	0.97	0.96	0.95	0.96

4.3.2 Estimation of growth kinetics

Three replicates of the growth curve experiment were carried out. For each growth curve, ODs were converted to population size using the OD-CFU calibration, and non-

linear regression was performed to parameterise a logistic growth model for each species and each experiment (Figure 4.4). Fitting of the logistic model of *Ec*N population size to the third replicate growth curve failed to converge when day 2 results were excluded, so for this data set, all measurements were used. The K value for the *Ec*N model of the second replicate was not significant and had a 95% confidence interval greater than its own value. This result was discarded from further analysis.

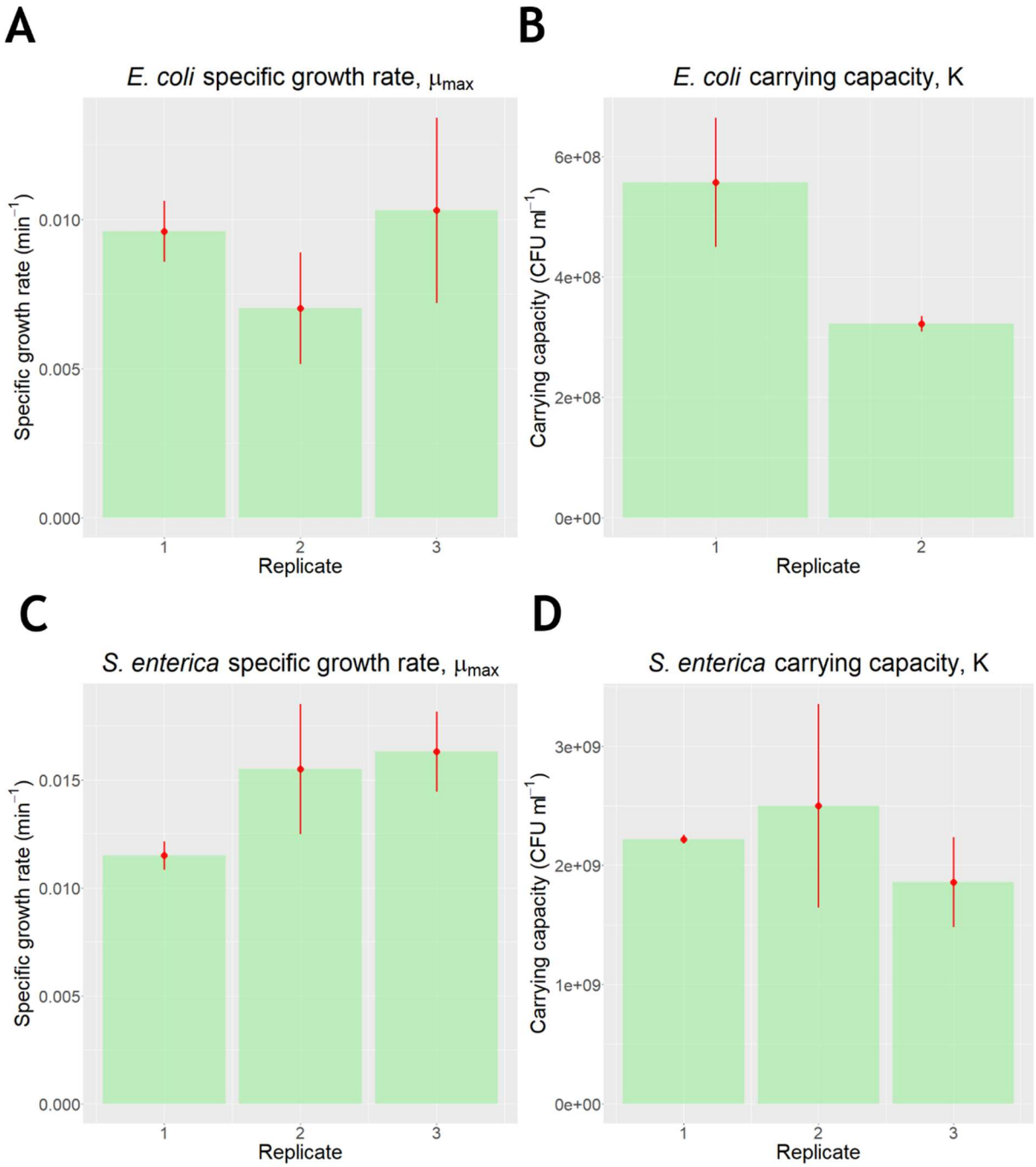

Figure 4.4 – Estimates for the values of μ_{max} and K for the logistic growth models fitted to growth experiments in CP medium. Error bars indicate the 95% confidence intervals. In the case of the μ_{max} parameter for *Ec*N, the result from replicate 2 was discarded, as it was two orders of magnitude greater than the other estimates, and had a confidence interval that extended below zero.

4.3.3 Yield estimations

To estimate biomass yields during growth of *E. coli* and *S. enterica*, cultures of both bacteria were grown in CP medium and then 50ml samples were taken, centrifuged and dried. The samples were put into dried falcon tubes that had been incubated at 80℃ for 24 hours, and were returned to the incubator with bacterial pellets, a control group of falcon tubes were also incubated before and after sampling, and the changes in mass in the experimental tubes were compared to those in the control tubes.

The mean change in mass of control Falcon tubes from immediately prior to sampling to after 28.33 hours drying was -3.35 mg, however a one-sample t-test showed that this value was not significantly different from 0.0 ($p = 0.3103$), so no adjustment was applied to the mass changes in the experimental tubes. The mass changes in the experimental tubes are shown in Table 4.4.

Table 4.4 – Results of the cell biomass estimation experiment. Population estimates were calculated from ODs using the OD-CFU calibration and then the mean population estimate taken from the three OD wavelengths. The biomass per CFU is calculated by dividing the mass change by the population estimate.

	***Ec*N 1st sample**	***Ec*N 2nd sample**	***S*Tm 1st sample**	***S*Tm 2nd sample**
Population estimate (CFU μl^{-1})	39396.63	111242.1	379949.1	1006742.0
Starting tube weight (g)	13.1489	13.1497	13.1340	13.1053
Final tube weight (g)	13.1726	13.1792	13.1610	13.1218
Mass change (mg)	23.7	29.5	27.0	16.5
Biomass per CFU (pg)	12.0315	5.3037	1.4212	0.3278

The biomass per CFU values vary across multiple orders of magnitude and are much higher than expected, as well as differing markedly from values in the literature. For this reason, information on cell dry mass for both species was sought from literature to make an alternative estimate.

Fagerbakke et al (1996) measured the dry mass of *E. coli* cells growing in exponential phase at 37°C and used X-ray microanalysis to measure the masses of almost all elements in the cells. Total dry mass was estimated by adding together the masses of

all elements measured and estimating an additional mass fraction from hydrogen. This yielded an estimated dry mass per cell of 710±80 fg.

Schaechter et al. (1958) found a dry mass per cell of 360 fg in *S. enterica* serovar Typhimurium growing at a similar rate to the cells growing in CP medium in this study (9.63×10^{-3} min^{-1}) and also at 37°C.

As *S*Tm and *Ec*N are close relatives within the Enterobacteriaceae (Alnajar and Gupta, 2017) and the experimental evidence is not clear enough to assert a difference between the two in terms of cell dry mass, the dry mass per cell was henceforth assumed to be 500 fg for both species during exponential growth. This would suggest carrying capacities of 0.219650 gl^{-1} for *Ec*N and 1.097000 gl^{-1} for *S*Tm. Given the concentration of GlcNAc in the CP medium, 5.53025gl^{-1}, this gives an estimated yield of 0.03972 g biomass/g GlcNAc for *Ec*N and 0.19836 g biomass/g GlcNAc for *S*Tm.

GlcNAc metabolism in *E. coli* and *S. enterica* is governed by the phosphotransferase system (PTS) for uptake, it can either be imported through the GlcNAc-specific PTS (IINag) or the mannose PTS (IIDMan, IICMan and IIBAMan). Both of these systems yield GlcNAc 6-phosphate, which can be converted to GlcN6P through deacetylation by the enzyme NagA and then to fructose-6-phosphate by enzyme NagB (Brinkkötter et al., 2000). Information on the kinetic parameters for the growth of *E. coli* and *S. enterica* are very sparse. However, the K_m value for E. coli nagB is 0.71 mM (Álvarez-Añorve et al., 2005) and given no evidence to the contrary it was assumed that the enzyme of S. enterica had the same K_m value. Although the K_m value for growth on a

substrate typically differs from the individual K_m values of enzymes in a pathway, this was the only value available in the literature, so the simulated K_m values for both *S*Tm and *Ec*N were set to 0.71 mM (0.1571 g l^{-1}).

4.3.4 MIMic results

Three replicates of the paired cross-invasion experiments in MIMic were conducted. All replicates showed a clear trend by which *Ec*N managed to successfully outcompete *S*Tm when it was the resident, resisting *S*Tm invasion, and also managed to invade MIMic devices when *S*Tm was the resident, but did not completely replace *S*Tm within the time frame of the experiment.

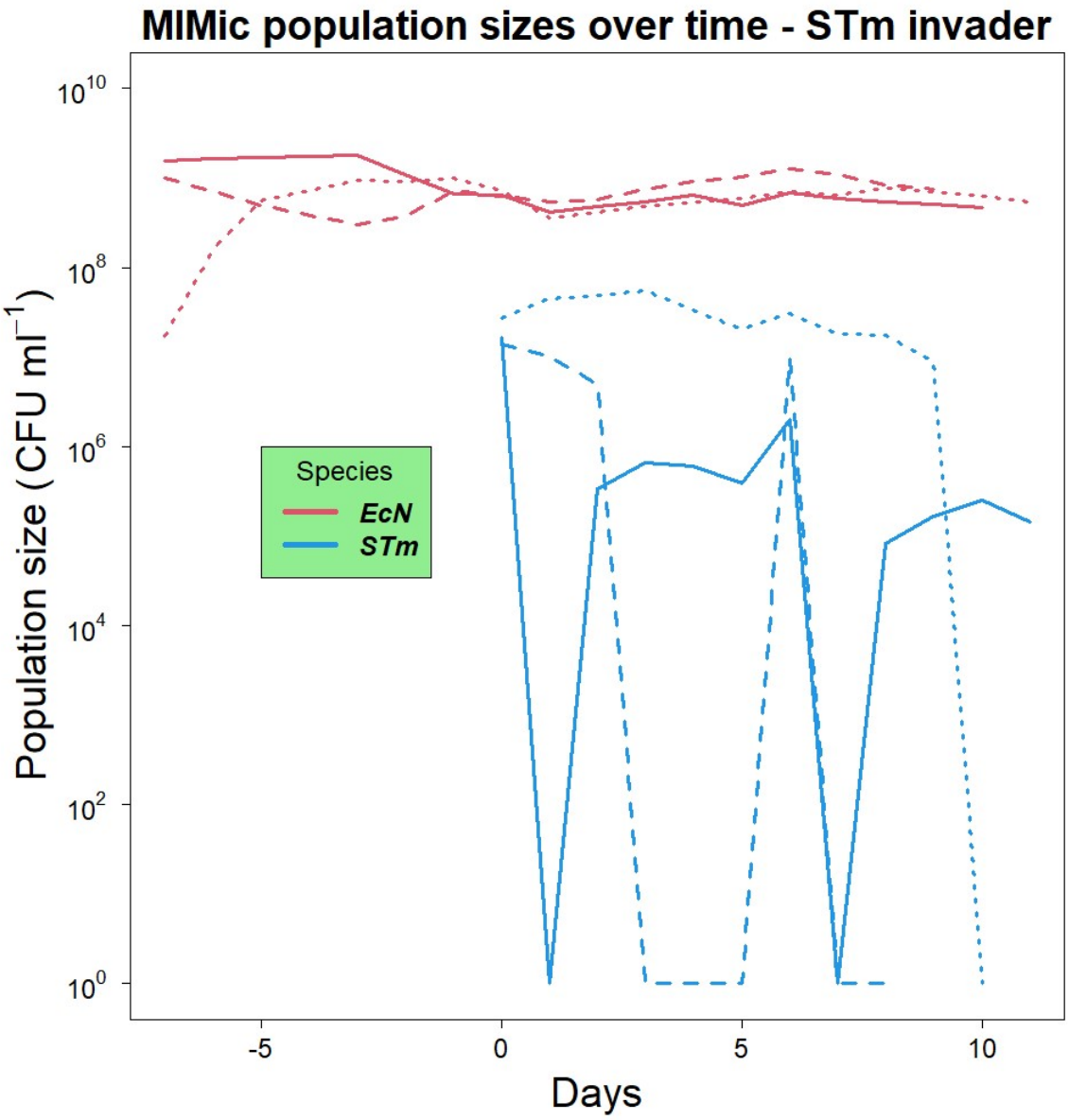

Figure 4.5 – Viable counts of *S*Tm and *Ec*N in a chemostat with resident *Ec*N and *S*Tm as an invader. Each replicate is represented using a different line style, with replicate 1 using solid lines, replicate 2 long dashes, and replicate 3 short dashes. The figure shows that *S*Tm is unable to invade a chemostat inhabited by a resident *Ec*N population. Day 0 represents the start of cross-invasion. *Data in this figure collected by the author, Mousomi Chakravorty and Migle Savukynaite.*

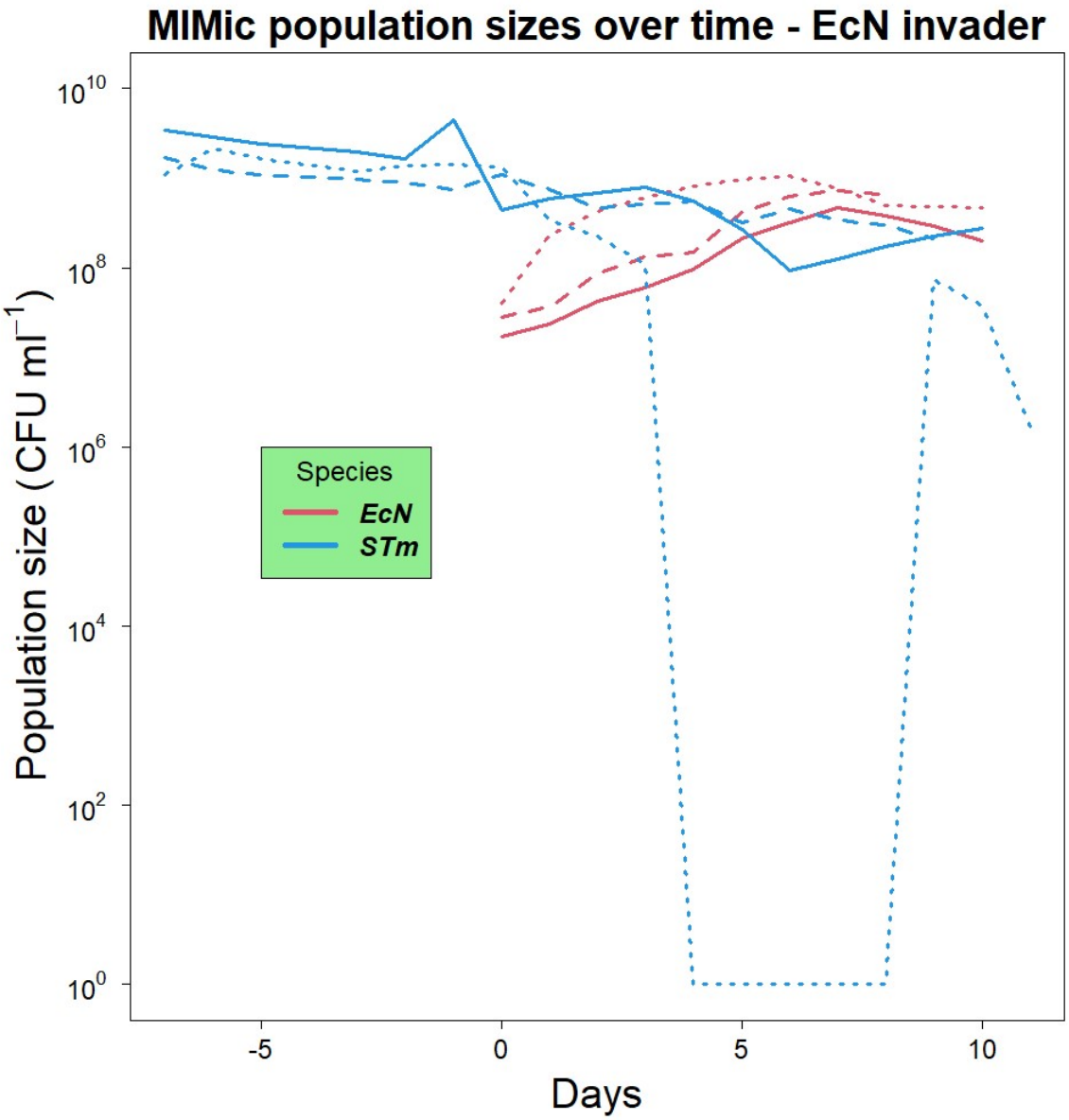

Figure 4.6 – The populations of *S*Tm and *Ec*N in a chemostat with resident *S*Tm and *Ec*N as an invader (inverse of Fig. 4.4). Each replicate is represented using a different line style, with replicate 1 using solid lines, replicate 2 long dashes, and replicate 3 short dashes. The figure shows that *Ec*N can invade a chemostat with resident *S*Tm, although the *S*Tm is not completely replaced. Day 0 represents the start of cross-invasion. *Data in this figure collected by the author, Mousomi Chakravorty and Migle Savukynaite.*

4.3.5 eMESC Results

Two variants of the eMESC model were simulated, one with *S*Tm as a resident, and *Ec*N at low numbers, and one with *Ec*N resident and *S*Tm at low numbers. For the characteristics of the two species, the variable estimates from the growth curve and yield experiments were used. In order to ascertain a suitable division mass for the simulation, several preliminary simulations of the system were run with different division masses to investigate the relationship between division mass and average cell mass (Table 4.5). The results showed a consistent relationship, with average cell mass being ~69% of the division mass. With the average cell mass estimate of 500 fg for both species, this would mean a division mass of 725 fg. However, in order to enable the simulation to proceed with reasonable speed, the division masses were increased 10^8-fold, to 72.5 μg, and agent numbers in the results were increased 10^8-fold to convert back to CFU. This means each agent in the simulation represents 10^8 cells in the MIMic device. The final values of all variables used in the simulation are given in Table 4.6. In order to determine the initial CFU values for the two species in each simulation, the average CFU values for each species in each invasion at day 0 are taken from the experimental results and converted to estimates for the whole 250 ml volume, before the number of individuals is scaled down by a factor of 10^8 to get the number of agents. The growth of both species is modelled using Monod kinetics, with each agent's growth proceeding at rate μ, the rate of increase in biomass, such that

$$\mu = \mu_{max} \frac{S}{K_m + S} m$$

where μ_{max} is the maximum possible growth rate, S is the concentration of GlcNAc, K_S is the affinity to GlcNAc and m is the agent's mass. All other agent characteristics are given in Table 4.6.

Table 4.5 – Mean cell masses at a selection of division masses in preliminary runs of the MIMic simulation. Division mass is shortened to "DV" in columns 2 and 3.

Division mass (μg)	Mean cell mass as percentage of DV	Standard deviation as percentage of DV
20	69.12	2.16
40	69.37	5.39
80	69.51	5.34

Table 4.6 – Parameters used in the MIMic simulations

	***Ec*N**	***S*Tm**
μ_{max} (min^{-1})	8.983×10^{-3}	1.445×10^{-2}
Yield (g biomass / g GlcNAc)	3.972×10^{-2}	1.9836×10^{-1}
K_S	0.1571 g l^{-1}	
Division mass	72.5 μg	
Initial cell number in *Ec*N invasion of *S*Tm	71	2399
Initial cell number in *S*Tm invasion of *Ec*N	1653	48

The results obtained in the eMESC model show the opposite trend to the MIMic results, with *Ec*N not able to invade the chemostat with resident *S*Tm, but *S*Tm persisting in the chemostat with resident *Ec*N, and presumably taking over given enough time (Figures 4.7, 4.8).

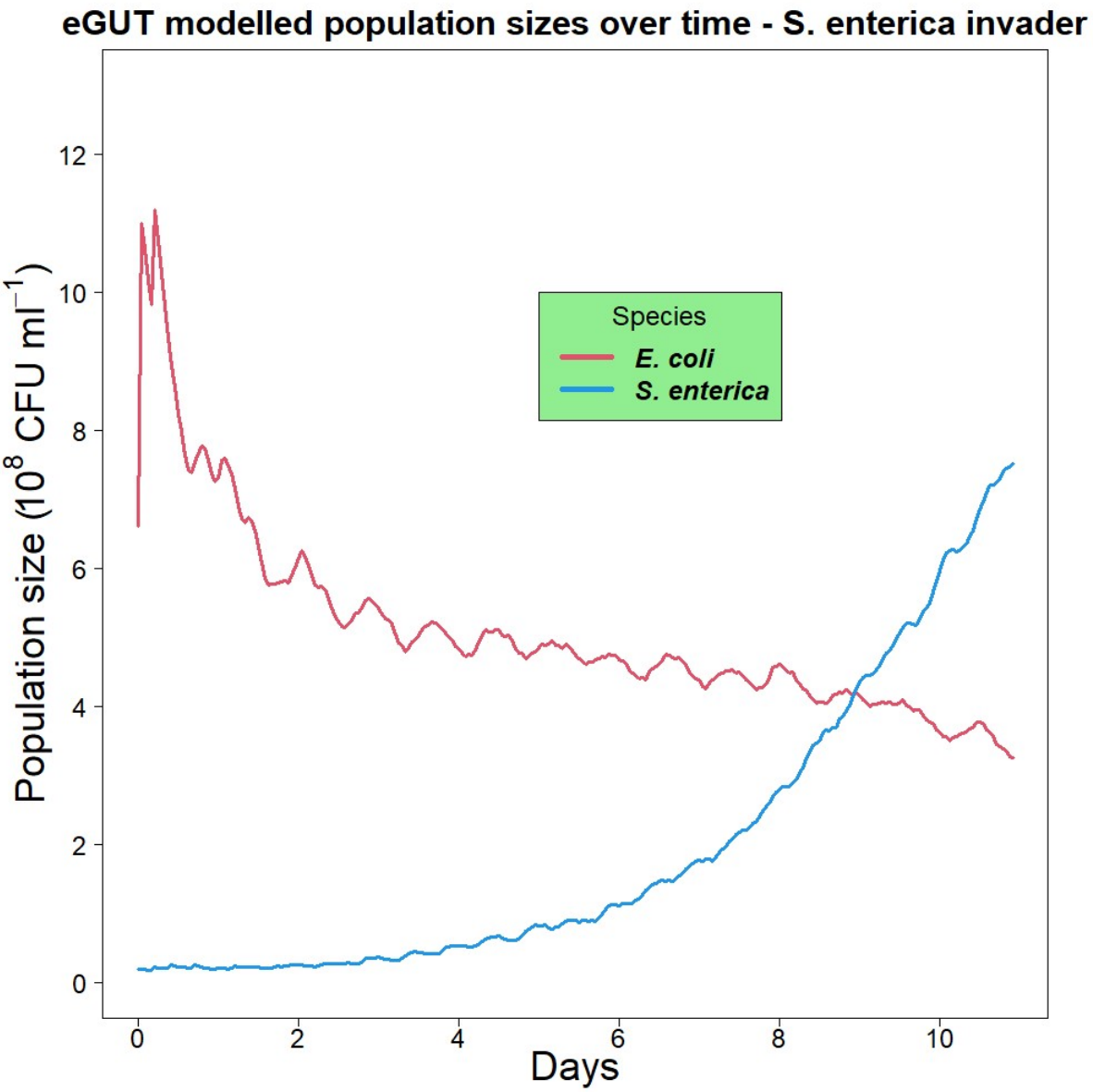

Figure 4.7 – The simulated populations of *Ec*N and *S*Tm in the eGUT model. Here a chemostat is modelled with an initially high *Ec*N population, representing the resident E. coli, and a low population of *S*Tm, representing the invader. *S*Tm is able

to persist and gradually increase in numbers, while the *Ec*N population decreases slowly after an initial drop.

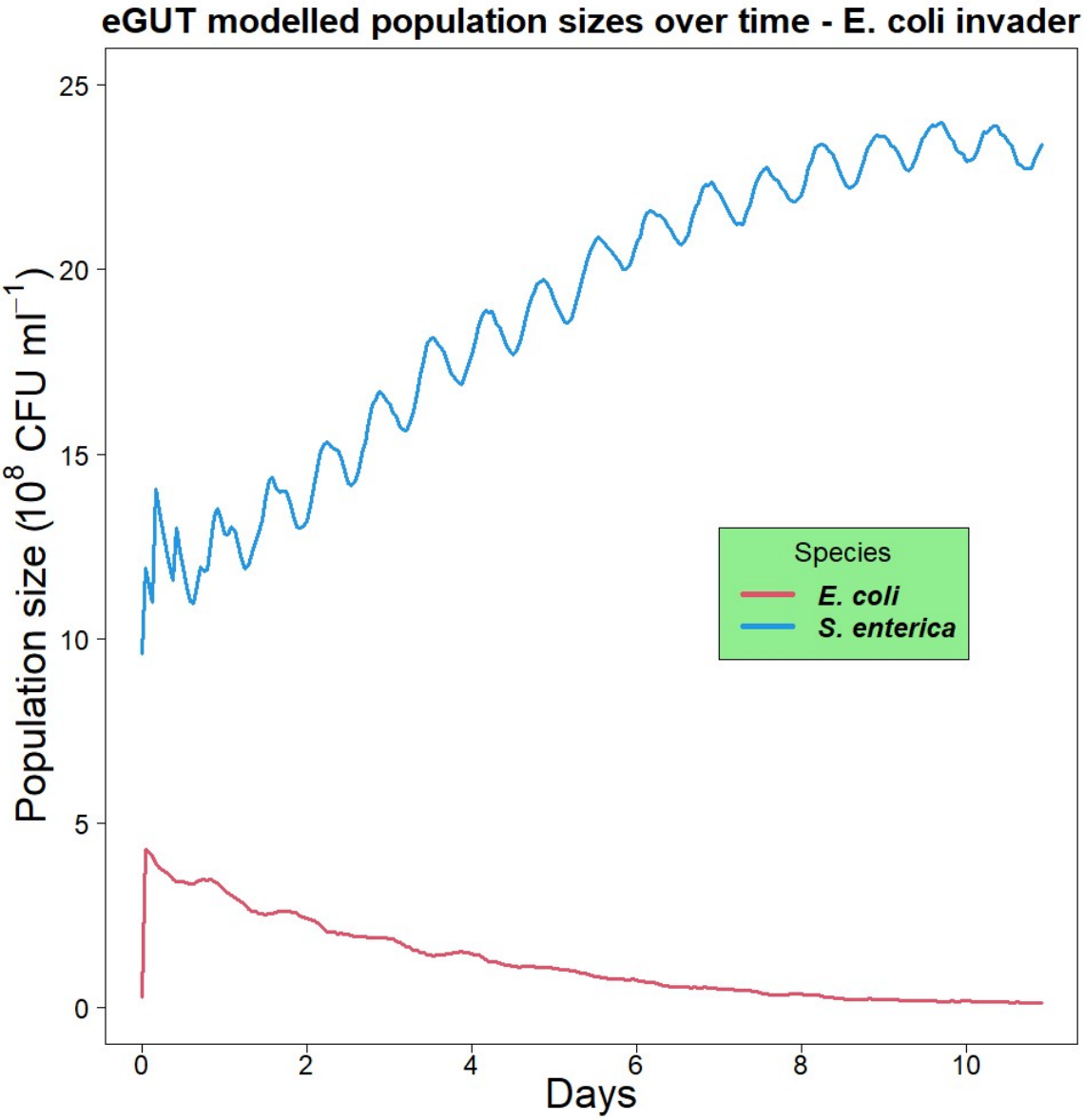

Figure 4.8 – The simulated populations of *S*Tm and *Ec*N in the eGUT model. Here, a high *S*Tm population (resident) and a low *Ec*N population (invader) are simulated. In this case, the *Ec*N population decreases asymptotically to zero.

As the results from MIMic and eGUT differed both quantitatively and qualitatively, the K_m value for GlcNAc consumption by *S*Tm was varied in the simulated system, to probe the sensitivity of the system. This showed that a ~2-fold increase in *S*Tm's K_m value would lead to a reversal of the outcome, with *Ec*N outcompeting *S*Tm (Figure 4.8).

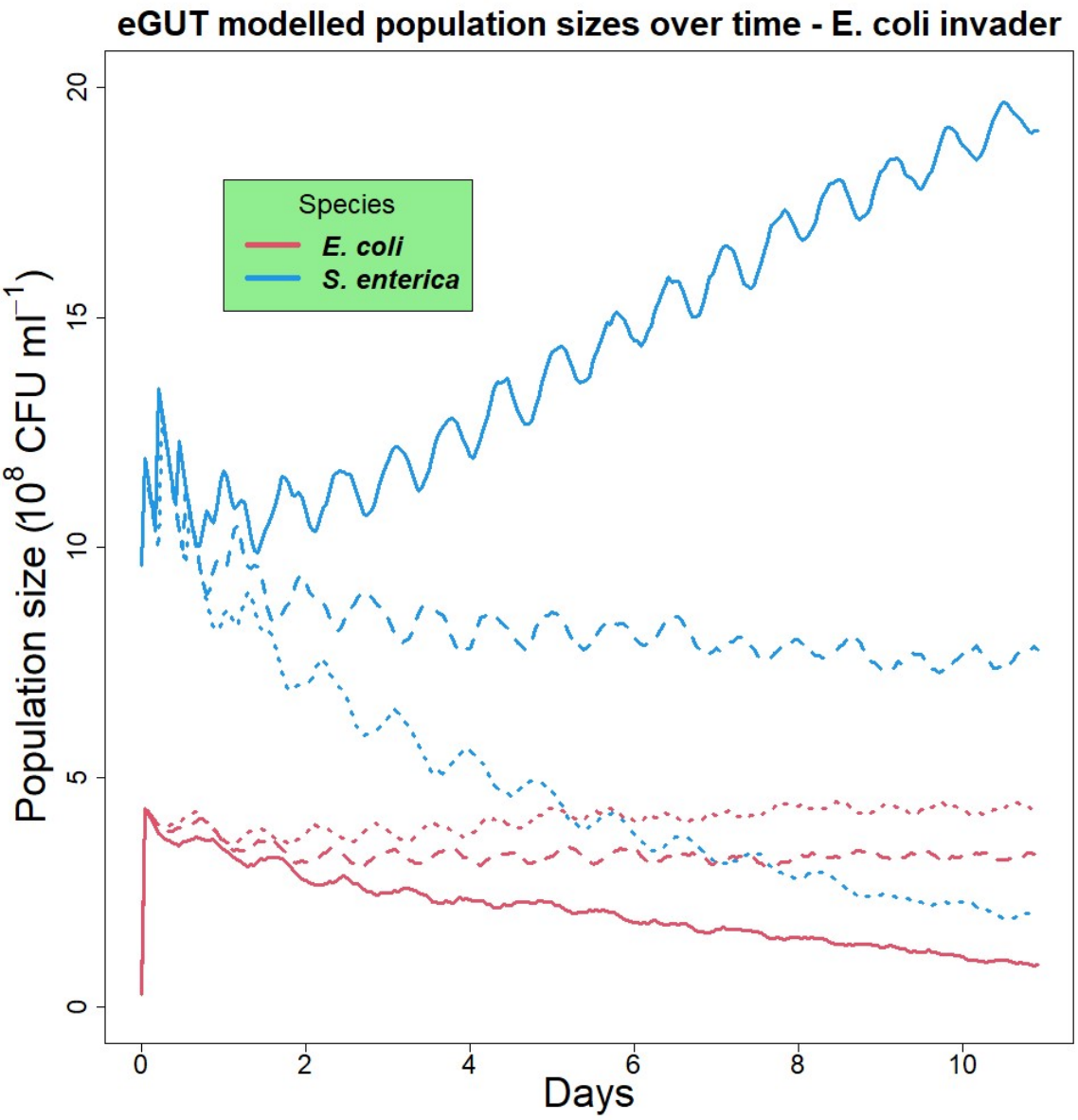

Figure 4.9 – Simulated populations of *Ec*N and *S*Tm in eGUT with varied values of K_m for *S*Tm and with *Ec*N as an invader and *S*Tm as resident. The solid lines correspond to a K_m value of 0.2 gl^{-1}, the long-dashed lines correspond to a K_m value of 0.25 gl^{-1} and the short-dashed lines correspond to a K_m value of 0.3 gl^{-1}.

4.4 Discussion

One of the most important mechanisms by which commensal gut microbes are thought to protect the gut from invasion by pathogens, or more generally colonization resistance, is through their occupation of potential attachment sites in the intestinal mucosa. Attachment to the mucosal surface is the best way for an invading microbial population to avoid washout, but the presence of a similar microbe in the gut, already occupying attachment sites gives the invading species/strain an inherent disadvantage, as any available attachment sites are more likely to be colonised by the numerous and nearby residents than any of the non-resident cells. (Freter et al., 1986).

The MIMic model is unlike Freter's mathematical model in that it is a well-mixed system, without attachment sites, which removes any inherent advantage enjoyed by a resident microbe. Despite this, the results from the experiments in MIMic show a clear pattern of *Ec*N outcompeting *S*Tm in a well-mixed, GlcNAc-fed system. This suggests that the competitive advantage of *Ec*N over *S*Tm that has been observed in the guts of humans, piglets and mice may not rely on spatial structure or residence advantages, implying *Ec*N's invasion resistance does not solely operate on the basis outlined by Freter et al. (1986).

Although residence effects may play a role in the probiotic effects exhibited by *Ec*N in the gut, the results from MIMic suggest that *Ec*N also has a biochemical advantage over *S*Tm in a well-mixed environment where the energy source is the mucus-derived amino sugar, N-acetylglucosamine. In other words, the ability for *Ec*N to resist

invasion by *S*Tm, but also to be able to invade a chemostat dominated by *S*Tm, are evidence of metabolic advantage allowing it to consume and grow on GlcNAc more quickly than *S*Tm.

However, the results from eMESC show the opposite trend. Here, *S*Tm outcompetes *Ec*N, whether it is an invader or a resident. It is unlikely that the difference observed between the two models is due to the effects of chance. In the case of MIMic, similar results were obtained in all three replicates. Furthermore, in eMESC, growth processes are deterministic, being simply determined by reaction rate calculations. The two stochastic processes in the chemostat model set up used are cell division, in which the proportion of biomass going to each daughter cell is subject to random variation within a small range, and dilution from the system, which removes agents with a set probability. Rerunning the eMESC model with a different random number seed does not change the results.

Thus, there is clearly an important mismatch between the results from MIMic and those from eMESC. One possible explanation for this difference is wrong growth kinetics used in eMESC, as various kinetic parameters were estimated based on literature values, rather than being measured in the lab. This includes the K_m values for both *S*Tm and *Ec*N, which were estimated from a K_m value for just one enzyme in the GlcNAc metabolism pathway in E. coli. Another estimated parameter was the average cell mass. Although experimental values were measured for this parameter, they varied widely and across orders of magnitude, as well as differing significantly

from current values in the literature, meaning falling back on literature values was the most sensible option.

This demonstrates one of the key limitations of eGUT. In order to accurately model a given species, a large amount must already be know about its metabolism, and its growth in the particular medium being used.

Further experiments to ascertain the true values of these parameters would be a useful step towards further validation of eGUT, and establishment of whether differences in GlcNAc metabolism can fully explain the observed population dynamics in vitro. A change in affinity can lead to a qualitatively opposite result, as can be seen from the outcome of changing the K_m value for the modelled *S*Tm population in eMESC. Thus estimating the values of this parameter for *Ec*N and *S*Tm could provide a lot of clarity regarding the mechanism of *Ec*N probiotic effect. Measuring K_m in the laboratory would require the running of chemostats with one species at a variety of different dilution rates and measuring the steady state biomass levels. Furthermore, more precise biomass measurements could show a difference between *Ec*N and *S*Tm in average mass per CFU, which may provide an alternative explanation as to why the eMESC results differ from the MIMic ones. Measuring biomass more accurately may require more precise methods, such as filtering, in order to ascertain a more reliable estimate of the relationship between CFU and dry mass.

Clearly there is a lot of information lacking regarding the kinetics of the growth of these two bacterial strains on GlcNAc, and it may be that *S*Tm has a slightly higher

K_m value than *Ec*N. Indeed, rerunning the simulation with a K_m value of 0.3 g/l for Salmonella leads to the outcome observed in MIMic, with E. coli able to outcompete *S*Tm. However, other mechanisms, such as the production of toxins may play a role.

As the CP medium contains 7.5 μM $FeCl_2$, it is unlikely that competition for iron played a role in the results for MIMic, given that in the paper by Sassone-Corsi et al. (2016), a medium with 1 μM iron citrate was considered iron rich, and not limiting for iron. Thus, we can rule out certain mechanisms, such as competition for iron, and chemical attack via microcin production. However, other mechanisms may exist that have not been identified to date.

In order to cause disease, *S*Tm must traverse the mucus boundary that lies between the gut lumen and the epithelium (Lu and Walker, 2001). GlcNAc is one of the most abundant monomers within mucin in the human colon, and is constantly released by the breakdown of mucins by microbes within the gut mucosa (Desai et al., 2016; Robbe et al., 2004). It therefore represents an important carbon source for gut microbes, and the genes required for its breakdown are found across a very wide range of bacterial families (Ravcheev and Thiele, 2017). For any pathogen population that must penetrate through the mucus layer to cause disease, GlcNAc is also likely to be an important resource. Simplifying the gut system to a chemostat with GlcNAc as the primary energy source allows many of the complex and interdependent factors governing population dynamics in the gut to be removed to focus on individual elements of the interaction between a probiotic and a pathogen. Here, this focus has

identified a difference between the two species that is likely to be explained by metabolic differences between the two species in growth on GlcNAc. Further investigation of the affinities of the two species for GlcNAc and the relationships between CFU and biomass could confirm this relationship and explain the phenomena observed in MIMic.

eGUT served as a testing ground for hypotheses to explain the results found in MIMic, such as whether the observations in MIMic could be explained by a particular set of growth kinetic parameters. Unfortunately, the paucity of data on growth kinetics in this instance was likely the cause of a lack of agreement between the two models. However, the exploratory nature of eGUT means it can still make predictions and guide further laboratory research. For example, eGUT can predict a ratio of K_m values to expect if the results in MIMic are caused by differences in growth kinetics. This prompts further work to ascertain more accurate estimates for the various growth parameters of these two species.

EcN is currently used as a probiotic in various countries, and the results from this study confirm its ability to antagonise pathogens from the Enterobacteriaceae. In particular, the results from MIMic suggest that competitive exclusion plays a role in this probiotic effect. However, this is likely to work synergistically with the priority effect described by Freter et al. (1986). The fact that *Ec*N could not replace the *S*Tm population within the 10-11 days of the time courses of the MIMic runs suggests that *Ec*N could not be used as a therapy for someone who was already infected with an

enteropathogen. The invasion of the system is slow in MIMic, taking more than 11 days, and this is in a system without the priority effects that are active in the gut. In addition, the MIMic system is well-mixed, has a smaller volume than the gut and does not include any of the background microbiota that in a real gut would be interacting with both species. Thus, we cannot draw strong conclusions about the use or efficacy of *Ec*N as a probiotic from these results alone.

The cyclic oscillations observed in some of the eGUT simulations are likely the result of synchronised cell cycles. Although subpopulations of agents with different sizes were used to initialise the simulations, there were still discrete populations, rather than a random distribution of masses. Although there is variation introduced to daughter cell masses during cell division, this may not have been enough to counter the synchronised cell cycles, leading to the cycles observed. This is particularly noticeable in a non-spatial simulation, as all cells of a given species experience the same concentrations of solutes and thus grow at the same rate.

CHAPTER 5 - CONCLUSIONS AND FUTURE OUTLOOK

5.1 The eGUT Modelling Platform

My contribution to developing, verifying and validating eGUT, a modelling platform that allows researchers to model the gut microbiome, has been described in this thesis. As an agent-based model, eGUT enables users to build models from the ground up, utilising knowledge on the characteristics and growth kinetics of individual microbes in order to study emergent phenomena in the gut's microbial community, such as competition, cross-feeding and colonisation resistance. The eGUT work builds on, and has contributed to, the development of the underlying or core functionality of iDynoMiCS 2.0, an agent-based model of biofilms that has been developed in parallel with eGUT by co-workers and myself as this core functionality was required by eGUT. While eGUT builds on top of iDynoMiCS 2.0, it has several key capabilities that are central to modelling the gut microbiome and that distinguish it from iDynoMiCS 1.

Firstly, while iDynoMiCS 1 is purely a model of biofilms growing on inert surfaces, eGUT includes a model of the gut mucosa, an active and semi-permeable 'surface'. The epithelial layer in eGUT consists of a monolayer of agents with a dedicated new spawner processes to populate the epithelium with a range of differentiated cell types. In addition, transport reactions were implemented to model both diffusion and active transport into and out of the epithelial cells, including transport through the epithelial cells, for example from mucosal compartment of the gut into a connected

compartment that can represent the blood stream. In addition to uptake and secretion, Epithelial cells can produce solutes according to various reaction kinetics and they can produce particles. This is meant for modelling mucus secretion by goblet cells which form part of the epithelial surface but could have other uses. In order to turn these mucus particles into a confluent but somewhat penetrable layer of mucus, a pattern-oriented modelling approach was adopted to find suitable parameters describing the mechanical interactions of the particles given the scarcity of measurements of such parameters. Since a broad region of parameter space matched the requirements, it was not possible to narrow down parameters let alone fit them as no quantitative empirical data were available.

Secondly, the gastrointestinal tract consists of several compartments along its axis with distinct characteristics. Moreover, the gut does not exist in isolation in the body and interacts with other organs in multiple ways, for example with liver, muscle and adipose tissue in the regulation of blood glucose levels. These organs are connected by the circulation of blood. Thus, the capacity to simulate multiple compartments that are connected with each other was important to develop. The general idea is to model other organs as well-mixed compartments that can be described with a system of ODEs and connect these organs with the mucosa via a 'blood' compartment that transports compounds. Regarding modelling multiple sections of the GI tract, several spatially mixed lumen compartments can be connected with each other while each lumen compartment can be connected with a spatially structured mucosa

compartment. Connecting compartments required implementing transfer of solutes and agents between compartments in one or both directions, by diffusion of solutes or stochastic exchange of agents.

Thirdly, another key advantage of eGUT, developed alongside iDynoMiCS 2.0, is its modular nature. The vast majority of agent characteristics, such as cell morphology, cell division, growth kinetics, density, EPS secretion, attraction forces to other agents and much more can be combined in whatever ways the user wishes, meaning novel species and even synthetic bacteria can be modelled just as easily as well-known model species. In iDynoMiCS 1, the predecessor of eGUT and iDynoMiCS 2.0, agents gained additional functionality by subclassing from simpler, more generic superclasses. This led to the creation of a large 'zoo' of agents with various special features including many features that were not required in each case. This zoo became increasingly confusing and difficult to maintain with lots of redundant code. The decision was thus taken to develop eGUT and iDynoMiCS 2.0 from scratch.

Finally, further key new capabilities of eGUT and iDynoMiCS 2.0 that I contributed to by implementing, reviewing and testing code as well as implementing code for inspecting the PDE solver's inner workings were the numerical algorithms, unit conversion capability, the new graphical user interface and numerous small additions.

Much of this development was undertaken in order to enable those scientists who have no experience in coding agent-based models to specify the gut system their lab works on. eGUT is meant to be a platform for domain experts to specify their system

without being dependent on a funded collaboration with an experienced modeller. The latter is of course an option, but one which is not always feasible or quick enough. Moreover, specifying the model by the domain expert avoids potential miscommunication between domain expert and modeller, although it creates possibilities of poor choices by the domain expert. eGUT is now at a stage that should allow non-modeller users to specify and simulate various aspects of gut population dynamics.

Continued development of eGUT alongside iDynoMiCS 2.0 will ensure that any capabilities that can be used in both platforms will be shared, and could allow for further diversification, and the development of other specialised models sharing a common core with iDynoMiCS 2.0 or eGUT, without having to "reinvent the wheel" by redesigning another agent-based model structure from the ground up.

5.2 Verification and validation of eGUT

To ensure results and predictions from model simulations can be relied upon to result only from the assumptions put into the model and are numerically correct, a variety of tests of numerical algorithms have been carried out, including writing code to inspect the inner workings of these algorithms.

This verification of the correct implementation of the eGUT platform had to be done using specific models as examples as generic models cannot be tested quantitatively. The range of examples was meant to cover expected usage. First, quantitative comparison with analytical solutions available for simpler systems such as a well-mixed chemostat or a thin, uniform biofilm layer was performed and demonstrated that eGUT produced correct output. This final outcome hides the considerable effort that had to go into finding bugs that were apparent from the results of these verification efforts. For verification of more complex models without analytical solutions, comparison with independent implementations of similar mathematical models was the approach chosen. For this, the Benchmark 3 test cases were chosen as these were set up for this purpose by an IWA task group and used for testing the predecessor, iDynoMiCS 1. These BM3 test cases simulate a typical wastewater treatment biofilm scenario with two competing and also cross-feeding bacterial species consuming and producing several resources, which enables testing of the interactions of agents with the environment. There is a wide range of models or modelling platforms included in the BM3 study, from continuum models via Cellular

Automata to particle-based models, a subset of which are agent-based models although the individuality of agents was limited or removed in eGUT for the sake of the BM3 comparison. The fact that the results from eGUT, especially the steady state solute concentrations in the bulk liquid, are similar to those from other models is promising. Again, this suggests that the numerical core of the program works as expected, and that eGUT will make quantitatively similar predictions to other biofilm models. It should be noted that there is quite some variation between different models, which, given their different nature, is not too surprising even though they model the same processes and are based on the same physical laws. Since eGUT is most similar to other agent-based models, especially iDynoMiCS 1, their specification corresponded more closely and thus results were more similar. this was especially true for biomass density, where both eGUT and iDynoMiCS 1 used a biomass density of 1.5×10^{-2} pg μm^{-3}. Based on the values of total biomass in the system, it seems that other models, including other agent-based models, did not use this value. Unfortunately, we were not able to find details in the description of these models to be certain of the way in which biofilm density was specified and what biofilm density profiles emerged in these simulations. Thus, while it is clear that tuning of the parameters in eGUT will yield results that match the other modelling platforms more closely, a deeper understanding of the causes of the differences may remain elusive.

For experimental validation, we also started with a very simple system, a bioreactor operated as a chemostat with two competing bacteria in a well-mixed environment

and a stable steady state. The results from the MIMic validation were less clear. The results from the MIMic chemostat model demonstrated that *E. coli* Nissle had a competitive advantage over *S. enterica* serovar Typhimurium, both as resident and as invader, in all three replicate runs. These results were consistent with several other experimental studies that also identified the underlying mechanisms. However, these mechanisms do not actually apply to the MIMic model as iron was replete and only one electron acceptor, oxygen, may have leaked into the system and only one carbon and energy source was provided. In contrast, the eGUT simulation results predicted *S. enterica* would outcompete *E. coli* as both a resident and an invader. This can be understood from the growth kinetics estimated from very limited published information. Given the lack of measurements, it was assumed that the K_m values for the growth of both species was the same. Hence, the outcome of the simulation in eGUT depended entirely on the value of μ_{max}. As the estimated μ_{max} for *S. enterica* was higher than that for *E. coli*, eGUT failed to reproduce the results observed in MIMic. However, the difficulty encountered in finding and/or estimating the growth kinetics for these species on the carbon and energy source N-acetylglucosamine (GlcNAc) make it somewhat unsurprising that there is a mismatch here. Further laboratory work to measure the various growth parameters of these two species would be very valuable in facilitating a more accurate and rigorous validation of the eGUT platform.

5.3 Future Work

Although the development of the core capabilities of eGUT is complete, and verification against analytical results and other *in silico* models has eliminated all issues that have a noticeable effect on tested outcomes, further work on eGUT should continue to make it even more user-friendly and as flexible as possible as well as embark on a programme of experimental validation from simpler to more complex test cases.

One of the most important factors in user satisfaction and, presumably, willingness to use a piece of software is the quality of its user guides (Gök et al., 2019). eGUT's user guidance currently gives only a bare-bones overview of the complexity of the software and may leave users confused or lacking information at some point in their interactions with the platform. It is therefore vitally important for further work to be carried out on improving and expanding the user guidance for eGUT, in order to make it easy for new users to become familiar with the platform. Ideally this should include not only a manual or wiki, but also videos for beginners, and an active group of developers available to answer questions regarding the platform.

Even more important will be further work on experimental validation of eGUT as a platform for specifying and simulating models that can predict microbial population dynamics in the gut and their interactions with the human body. Currently, eGUT has passed certain key tests, but not yet completed any experimental validation. A sensible approach is to start validation in simple *in vitro* systems, moving to more complex *in*

vitro systems, for example, including a mucosal compartment or spatial structure. This should be followed by *in vivo* validation to convince other researchers that eGUT is a reliable and appropriate modelling platform. Agent-based modelling is an exciting and growing field (Hellweger et al., 2016a) able to explain phenomena as diverse as swarming animals (Kim and Shin, 2006) and altruistic microbes in biofilms (Kreft, 2004). However, in a world where there is an ever-increasing attention being paid to the dynamics of the gut, and development of probiotic and prebiotic treatments, eGUT should aim to not just explain or identify phenomena, but to make accurate and useful predictions that can lead to the identification of potential probiotics, prebiotics or other medically significant contributions. In order to be trusted to provide such breakthroughs, a comprehensive package of validation experiments will be needed.

REFERENCES

Allan, E.S., Winter, S., Light, A.M., Allan, A., 1996. Mucosal enzyme activity for butyrate oxidation; no defect in patients with ulcerative colitis. Gut 38, 886–893. https://doi.org/10.1136/gut.38.6.886

Allegretti, J.R., Mullish, B.H., Kelly, C., Fischer, M., 2019. The evolution of the use of faecal microbiota transplantation and emerging therapeutic indications. The Lancet 394, 420–431. https://doi.org/10.1016/S0140-6736(19)31266-8

Allen, A., 1989. Gastrointestinal Mucus, in: Handbook of Physiology, The Gastrointestinal System, Salivary, Gastric, Pancreatic, and Hepatobiliary Secretion. American Physiological Society, Bethesda, MD, pp. 359–382. https://doi.org/10.1002/cphy.cp060319

Allen, S.J., Martinez, E.G., Gregorio, G.V., Dans, L.F., 2010. Probiotics for treating acute infectious diarrhoea. Cochrane Database Syst. Rev. 2010, CD003048. https://doi.org/10.1002/14651858.CD003048.pub3

Alnajar, S., Gupta, R.S., 2017. Phylogenomics and comparative genomic studies delineate six main clades within the family Enterobacteriaceae and support the reclassification of several polyphyletic members of the family. Infect. Genet. Evol. 54, 108–127. https://doi.org/10.1016/j.meegid.2017.06.024

Altenhoefer, A., Oswald, S., Sonnenborn, U., Enders, C., Schulze, J., Hacker, J., Oelschlaeger, T.A., 2004. The probiotic Escherichia coli strain Nissle 1917 interferes with invasion of human intestinal epithelial cells by different enteroinvasive bacterial pathogens. FEMS Immunol. Med. Microbiol. 40, 223–229. https://doi.org/10.1016/S0928-8244(03)00368-7

Álvarez-Añorve, L.I., Calcagno, M.L., Plumbridge, J., 2005. Why Does Escherichia coli Grow More Slowly on Glucosamine than on N-Acetylglucosamine? Effects of Enzyme Levels and Allosteric Activation of GlcN6P Deaminase (NagB) on Growth Rates. J. Bacteriol. 187, 2974–2982. https://doi.org/10.1128/JB.187.9.2974-2982.2005

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, I.I., Umesaki, Y., Itoh, K., Honda, K., 2011. Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. Science 331, 337–341.

Atuma, C., Strugala, V., Allen, A., Holm, L., 2001. The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. Am. J. Physiol.-Gastrointest. Liver Physiol. 280, G922–G929. https://doi.org/10.1152/ajpgi.2001.280.5.G922

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist, H., Vavilin, V.A., 2002. The IWA Anaerobic Digestion Model No 1 (ADM 1). Water Sci. Technol. 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., Kaleta, C., 2017. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. PLOS Comput. Biol. 13, e1005544. https://doi.org/10.1371/journal.pcbi.1005544

Behnsen, J., Liu, J., Valeri, M., Hoover, E., Tjokrosurjo, J., Montaldo, N.P., Treacy-Abarca, S., Garibay, O., Gilston, B.A., Edwards, R.A., Chazin, W., Skaar, E.P., Raffatellu, M., 2017. Probiotic *Escherichia coli* Nissle 1917 Uses Zinc Transporters and the

Siderophore Yersiniabactin to Acquire Zinc in the Inflamed Gut and Outcompete *Salmonella Typhimurium*. FASEB J. 31, 622.6-622.6. https://doi.org/10.1096/fasebj.31.1_supplement.622.6

Blaak, E.E., Canfora, E.E., Theis, S., Frost, G., Groen, A.K., Mithieux, G., Nauta, A., Scott, K., Stahl, B., van Harsselaar, J., van Tol, R., Vaughan, E.E., Verbeke, K., 2020. Short chain fatty acids in human gut and metabolic health. Benef. Microbes 11, 411–455. https://doi.org/10.3920/BM2020.0057

Blanton, L.V., Barratt, M.J., Charbonneau, M.R., Ahmed, T., Gordon, J.I., 2016. Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics. Science 352, 1533–1533. https://doi.org/10.1126/science.aad9359

Blick, R., Burns, K.C., 2009. Network properties of arboreal plants: Are epiphytes, mistletoes and lianas structured similarly? Perspect. Plant Ecol. Evol. Syst. 11, 41–52. https://doi.org/10.1016/j.ppees.2008.10.002

Brandt, A., 1977. Multi-level adaptive solutions to boundary-value problems. Math. Comput. 31, 333–390. https://doi.org/10.1090/S0025-5718-1977-0431719-X

Brinkkötter, A., Klöss, H., Alpert, C., Lengeler, J.W., 2000. Pathways for the utilization of N-acetyl-galactosamine and galactosamine in Escherichia coli. Mol. Microbiol. 37, 125–135. https://doi.org/10.1046/j.1365-2958.2000.01969.x

Brozek, J., Grande, F., Anderson, J.T., Keys, A., 1963. Densitometric analysis of body composition: Revision of some quantitative assumptions. Ann. N. Y. Acad. Sci. 110, 113–140. https://doi.org/10.1111/j.1749-6632.1963.tb17079.x

Bueno, C., Villegas, M.L., Bertolotti, S.G., Previtali, C.M., Neumann, M.G., Encinas, M.V., 2002. The Excited-State Interaction of Resazurin and Resorufin with Aminesin Aqueous Solutions. Photophysics and Photochemical Reaction¶. Photochem. Photobiol. 76, 385–390. https://doi.org/10.1562/0031-8655(2002)0760385TESIOR2.0.CO2

Bunesova, V., Lacroix, C., Schwab, C., 2018. Mucin Cross-Feeding of Infant Bifidobacteria and Eubacterium hallii. Microb. Ecol. 75, 228–238. https://doi.org/10.1007/s00248-017-1037-4

Callaway, T.R., Edrington, T.S., Anderson, R.C., Harvey, R.B., Genovese, K.J., Kennedy, C.N., Venn, D.W., Nisbet, D.J., 2008. Probiotics, prebiotics and competitive exclusion for prophylaxis against bacterial disease. Anim. Health Res. Rev. 9, 217–225. https://doi.org/10.1017/S1466252308001540

Cammarota, G., Masucci, L., Ianiro, G., Bibbò, S., Dinoi, G., Costamagna, G., Sanguinetti, M., Gasbarrini, A., 2015. Randomised clinical trial: faecal microbiota transplantation by colonoscopy vs. vancomycin for the treatment of recurrent Clostridium difficile infection. Aliment. Pharmacol. Ther. 41, 835–843. https://doi.org/10.1111/apt.13144

Clausen, M.R., Mortensen, P.B., 1994. Kinetic studies on the metabolism of short-chain fatty acids and glucose by isolated rat colonocytes. Gastroenterology 106, 423–432. https://doi.org/10.1016/0016-5085(94)90601-7

Clegg, R.J., Kreft, J.-U., 2017. Reducing discrepancies between 3D and 2D simulations due to cell packing density. J. Theor. Biol. 423, 26–30. https://doi.org/10.1016/j.jtbi.2017.04.016

Cockburn, D.W., Koropatkin, N.M., 2016. Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease. J. Mol. Biol.,

Glycosylation in Cellular Processes and Disease 428, 3230–3252. https://doi.org/10.1016/j.jmb.2016.06.021
De Preter, V., Geboes, K.P., Bulteel, V., Vandermeulen, G., Suenaert, P., Rutgeerts, P., Verbeke, K., 2011. Kinetics of butyrate metabolism in the normal colon and in ulcerative colitis: the effects of substrate concentration and carnitine on the β-oxidation pathway. Aliment. Pharmacol. Ther. 34, 526–532. https://doi.org/10.1111/j.1365-2036.2011.04757.x
Deriu, E., Liu, J.Z., Pezeshki, M., Edwards, R.A., Ochoa, R.J., Contreras, H., Libby, S.J., Fang, F.C., Raffatellu, M., 2013. Probiotic Bacteria Reduce Salmonella Typhimurium Intestinal Colonization by Competing for Iron. Cell Host Microbe 14, 26–37. https://doi.org/10.1016/j.chom.2013.06.007
Derrien, M., van Passel, M.W., van de Bovenkamp, J.H., Schipper, R.G., de Vos, W.M., Dekker, J., 2010. Mucin-bacterial interactions in the human oral cavity and digestive tract. Gut Microbes 1, 254–268. https://doi.org/10.4161/gmic.1.4.12778
Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., Young, V.B., Henrissat, B., Wilmes, P., Stappenbeck, T.S., Núñez, G., Martens, E.C., 2016. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell 167, 1339-1353.e21. https://doi.org/10.1016/j.cell.2016.10.043
Diether, N.E., Willing, B.P., 2019. Microbial Fermentation of Dietary Protein: An Important Factor in Diet–Microbe–Host Interaction. Microorganisms 7. https://doi.org/10.3390/microorganisms7010019
Dimijian, G.G., 2000. Evolving together: the biology of symbiosis, part 1. Proc. Bayl. Univ. Med. Cent. 13, 217–226.
Dockery, J., Klapper, I., 2001. Finger formation in biofilm layers. SIAM J. Appl. Math. 62, 853–869.
Donaldson, G.P., Lee, S.M., Mazmanian, S.K., 2016. Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 14, 20–32. https://doi.org/10.1038/nrmicro3552
Duncan, K., Carey-Ewend, K., Vaishnava, S., 2021. Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer. Gut Microbes 13, 1874815. https://doi.org/10.1080/19490976.2021.1874815
Ebeling, W., 2004. Nonlinear Brownian Motion - mean square displacement. Condens. Matter Phys. 7, 539–550.
Eberl, C., Weiss, A.S., Jochum, L.M., Raj, A.C.D., Ring, D., Hussain, S., Herp, S., Meng, C., Kleigrewe, K., Gigl, M., Basic, M., Stecher, B., 2021. E. coli enhance colonization resistance against Salmonella Typhimurium by competing for galactitol, a context-dependent limiting carbon source. Cell Host Microbe 29, 1–13. https://doi.org/10.1016/j.chom.2021.09.004
Fàbrega, A., Vila, J., 2013. Salmonella enterica Serovar Typhimurium Skills To Succeed in the Host: Virulence and Regulation. Clin. Microbiol. Rev. 26, 308–341. https://doi.org/10.1128/CMR.00066-12
Fagerbakke, K., Heldal, M., Norland, S., 1996. Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. Aquat. Microb. Ecol. 10, 15–27. https://doi.org/10.3354/ame010015

Ficara, M., Pietrella, E., Spada, C., Della Casa Muttini, E., Lucaccioni, L., Iughetti, L., Berardi, A., 2020. Changes of intestinal microbiota in early life. J. Matern. Fetal Neonatal Med. 33, 1036–1043. https://doi.org/10.1080/14767058.2018.1506760

Fierer, N., Lennon, J.T., 2011. The generation and maintenance of diversity in microbial communities. Am. J. Bot. 98, 439–448. https://doi.org/10.3732/ajb.1000498

Fletcher, J.R., Pike, C.M., Parsons, R.J., Rivera, A.J., Foley, M.H., McLaren, M.R., Montgomery, S.A., Theriot, C.M., 2021. Clostridioides difficile exploits toxin-mediated inflammation to alter the host nutritional landscape and exclude competitors from the gut microbiota. Nat. Commun. 12, 462. https://doi.org/10.1038/s41467-020-20746-4

Fons, A.G., Tuomo Karjalainen, Michel, 2000. Mechanisms of Colonisation and Colonisation Resistance of the Digestive Tract Part 2: Bacteria/Bacteria Interactions. Microb. Ecol. Health Dis. 12, 240–246. https://doi.org/10.1080/089106000750060495

Freter, R., 1984. Interdependence of mechanisms that control bacterial colonization of the large intestine. Microecol. Ther. 14, 89–96.

Freter, R., Brickner, H., Fekete, J., Vickerman, M.M., Carey, K.E., 1983. Survival and implantation of *Escherichia coli* in the intestinal tract. Infect. Immun. 39, 686–703.

Freter, R., Brickner, H., Temme, S., 1986. An understanding of colonization resistance of the mammalian large intestine requires mathematical analysis. Microecol. Ther. 16, 147–155.

Gök, O., Ersoy, P., Börühan, G., 2019. The effect of user manual quality on customer satisfaction: the mediating effect of perceived product quality. J. Prod. Brand Manag. 28, 475–488. https://doi.org/10.1108/JPBM-10-2018-2054

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W.F., Veldhuyzen van Zanten, S.J.O., 2006. Differences between Tissue-Associated Intestinal Microfloras of Patients with Crohn's Disease and Ulcerative Colitis. J. Clin. Microbiol. 44, 4136–4141. https://doi.org/10.1128/JCM.01004-06

Grimm, V., 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecol. Model. 115, 129–148.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecol. Model. 198, 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J., Giske, J., Railsback, S.F., 2010. The ODD protocol: a review and first update. Ecol. Model. 221, 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a "multi-scope" for predictive systems ecology. Philos. Trans. R. Soc. B Biol. Sci. 367, 298–310. https://doi.org/10.1098/rstb.2011.0180

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., Deangelis, D.L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A.S.A., Milles, A., Nabe-Nielsen, J., Polhill, J.G., Radchuk, V., Rohwäder, M.S., Stillman, R.A., Thiele, J.C., Ayllón, D., 2020. The ODD protocol for describing agent-based and other simulation

models: A second update to improve clarity, replication, and structural realism. J. Artif. Soc. Soc. Simul. 23.
Guan, C., Ye, C., Yang, X., Gao, J., 2010. A review of current large-scale mouse knockout efforts. genesis 48, 73–85. https://doi.org/10.1002/dvg.20594
Gunawardena, J., 2014. Time-scale separation – Michaelis and Menten's old idea, still bearing fruit. FEBS J. 281, 473–488. https://doi.org/10.1111/febs.12532
Gutiérrez, M., Gregorio-Godoy, P., Pérez del Pulgar, G., Muñoz, L.E., Sáez, S., Rodríguez-Patón, A., 2017. A New Improved and Extended Version of the Multicell Bacterial Simulator gro. ACS Synth. Biol. 6, 1496–1508. https://doi.org/10.1021/acssynbio.7b00003
Hadi, A., Alizadeh, K., Hajianfar, H., Mohammadi, H., Miraghajani, M., 2020. Efficacy of synbiotic supplementation in obesity treatment: A systematic review and meta-analysis of clinical trials. Crit. Rev. Food Sci. Nutr. 60, 584–596. https://doi.org/10.1080/10408398.2018.1545218
Hall, L., 2019. Manipulating the early life microbiota.
Hart, S.F.M., Mi, H., Green, R., Xie, L., Pineda, J.M.B., Momeni, B., Shou, W., 2019. Uncovering and resolving challenges of quantitative modeling in a simplified community of interacting cells. PLOS Biol. 17, e3000135. https://doi.org/10.1371/journal.pbio.3000135
Hellweger, F.L., Clegg, R.J., Clark, J.R., Plugge, C.M., Kreft, J.-U., 2016a. Advancing microbial sciences by individual-based modelling. Nat. Rev. Microbiol. 14, 461–471. https://doi.org/10.1038/nrmicro.2016.62
Hellweger, F.L., Fredrick, N.D., McCarthy, M.J., Gardner, W.S., Wilhelm, S.W., Paerl, H.W., 2016b. Dynamic, mechanistic, molecular-level modeling of cyanobacteria: *Anabaena* and nitrogen interaction. Environ. Microbiol. 18, 2721–2731. https://doi.org/10.1111/1462-2920.13299
Herp, S., Brugiroux, S., Garzetti, D., Ring, D., Jochum, L.M., Beutler, M., Eberl, C., Hussain, S., Walter, S., Gerlach, R.G., Ruscheweyh, H.J., Huson, D., Sellin, M.E., Slack, E., Hanson, B., Loy, A., Baines, J.F., Rausch, P., Basic, M., Bleich, A., Berry, D., Stecher, B., 2019. Mucispirillum schaedleri Antagonizes Salmonella Virulence to Protect Mice against Colitis. Cell Host Microbe. https://doi.org/10.1016/j.chom.2019.03.004
Hertz, H., 1882. Ueber die Berührung fester elastischer Körper. 1882, 156–171. https://doi.org/10.1515/crll.1882.92.156
Hewitt, M., Ellison, C.M., Cronin, M.T.D., Pastor, M., Steger-Hartmann, T., Munoz-Muriendas, J., Pognan, F., Madden, J.C., 2015. Ensuring confidence in predictions: A scheme to assess the scientific validity of in silico models. Adv. Drug Deliv. Rev., In silico ADMET predictions in pharmaceutical research 86, 101–111. https://doi.org/10.1016/j.addr.2015.03.005
Hill, D.B., Button, B., Rubinstein, M., Boucher, R.C., 2022. Physiology and pathophysiology of human airway mucus. Physiol. Rev. 102, 1757–1836. https://doi.org/10.1152/physrev.00004.2021
Huang, K., Zeng, J., Liu, X., Jiang, T., Wang, J., 2021. Structure of the mannose phosphotransferase system (man-PTS) complexed with microcin E492, a pore-forming bacteriocin. Cell Discov. 7, 1–5. https://doi.org/10.1038/s41421-021-00253-6

Hugenholtz, F., de Vos, W.M., 2018. Mouse models for human intestinal microbiota research: a critical evaluation. Cell. Mol. Life Sci. 75, 149–160. https://doi.org/10.1007/s00018-017-2693-8

Ianiro, G., Maida, M., Burisch, J., Simonelli, C., Hold, G., Ventimiglia, M., Gasbarrini, A., Cammarota, G., 2018. Efficacy of different faecal microbiota transplantation protocols for Clostridium difficile infection: A systematic review and meta-analysis. United Eur. Gastroenterol. J. 6, 1232–1244. https://doi.org/10.1177/2050640618780762

Isolauri, E., Salminen, S., Ouwehand, A.C., 2004. Probiotics. Best Pract. Res. Clin. Gastroenterol. 18, 299–313. https://doi.org/10.1016/j.bpg.2003.10.006

Johansson, M.E.V., Larsson, J.M., Hansson, G.C., 2011. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc. Natl. Acad. Sci. U. S. A. 108, 4659–4665.

Jørgensen, B.B., Revsbech, N.P., 1985. Diffusive boundary layers and the oxygen uptake of sediments and detritus. Limnol. Oceanogr. 30, 111–122. https://doi.org/10.4319/lo.1985.30.1.0111

Kalbermatter, C., Fernandez Trigo, N., Christensen, S., Ganal-Vonarburg, S.C., 2021. Maternal Microbiota, Early Life Colonization and Breast Milk Drive Immune Development in the Newborn. Front. Immunol. 12.

Kettle, H., Louis, P., Holtrop, G., Duncan, S.H., Flint, H.J., 2014. Modelling the emergent dynamics and major metabolites of the human colonic microbiota. Environ. Microbiol. 17, 1615–1630. https://doi.org/10.1111/1462-2920.12599

Kilburn, D.G., Lilly, M.D., Webb, F.C., 1969. The Energetics of Mammalian Cell Growth. J. Cell Sci. 4, 645–654. https://doi.org/10.1242/jcs.4.3.645

Kim, D.H., Shin, S., 2006. Self-organization of Decentralized Swarm Agents Based on Modified Particle Swarm Algorithm. J. Intell. Robot. Syst. 46, 129–149. https://doi.org/10.1007/s10846-006-9047-3

Kostic, A.D., Howitt, M.R., Garrett, W.S., 2013. Exploring host–microbiota interactions in animal models and humans. Genes Dev. 27, 701–718. https://doi.org/10.1101/gad.212522.112

Kreft, J.-U., 2009. Mathematical Modeling of Microbial Ecology: Spatial Dynamics of Interactions in Biofilms and Guts, in: Jaykus, L.-A., Wang, H.H., Schlesinger, L.S. (Eds.), Food-Borne Microbes: Shaping the Host Ecosystem. ASM Press, Washington, DC, pp. 347–377.

Kreft, J.-U., 2004. Biofilms promote altruism. Microbiology 150, 2751–2760. https://doi.org/10.1099/mic.0.26829-0

Kreft, J.-U., Picioreanu, C., Wimpenny, J.W.T., van Loosdrecht, M.C.M., 2001. Individual-based modelling of biofilms. Microbiology 147, 2897–2912. https://doi.org/10.1099/00221287-147-11-2897

Lai, C.Y., Sung, J., Cheng, F., Tang, W., Wong, S.H., Chan, P.K.S., Kamm, M.A., Sung, J.J.Y., Kaplan, G., Chan, F.K.L., Ng, S.C., 2019. Systematic review with meta-analysis: review of donor features, procedures and outcomes in 168 clinical studies of faecal microbiota transplantation. Aliment. Pharmacol. Ther. 49, 354–363. https://doi.org/10.1111/apt.15116

Lai, S.K., Wang, Y.-Y., Wirtz, D., Hanes, J., 2009. Micro- and macrorheology of mucus. Adv. Drug Deliv. Rev. 61, 86–100. https://doi.org/10.1016/j.addr.2008.09.012

Lan, A., Andriamihaja, M., Blouin, J.-M., Liu, X., Descatoire, V., Desclée de Maredsous, C., Davila, A.-M., Walker, F., Tomé, D., Blachier, F., 2015. High-protein diet differently modifies intestinal goblet cell characteristics and mucosal cytokine expression in ileum and colon. J. Nutr. Biochem. 26, 91–98. https://doi.org/10.1016/j.jnutbio.2014.09.007

Lardon, L.A., Merkey, B.V., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., Smets, B.F., 2011. iDynoMiCS: next-generation individual-based modelling of biofilms. Environ. Microbiol. 13, 2416–2434. https://doi.org/10.1111/j.1462-2920.2011.02414.x

Li, B., Taniguchi, D., Gedara, J.P., Gogulancea, V., Gonzalez-Cabaleiro, R., Chen, J., McGough, A.S., Ofiteru, I.D., Curtis, T.P., Zuliani, P., 2019. NUFEB: A massively parallel simulator for individual-based modelling of microbial communities. PLOS Comput. Biol. 15, e1007125. https://doi.org/10.1371/journal.pcbi.1007125

Li, N., Zuo, B., Huang, S., Zeng, B., Han, D., Li, T., Liu, T., Wu, Z., Wei, H., Zhao, J., Wang, J., 2020. Spatial heterogeneity of bacterial colonization across different gut segments following inter-species microbiota transplantation. Microbiome 8, 161. https://doi.org/10.1186/s40168-020-00917-7

Lin, C., Culver, J., Weston, B., Underhill, E., Gorky, J., Dhurjati, P., 2018. GutLogo: Agent-based modeling framework to investigate spatial and temporal dynamics in the gut microbiome. PLOS ONE 13, e0207072. https://doi.org/10.1371/journal.pone.0207072

Litvak, Y., Byndloss, M.X., Bäumler, A.J., 2018. Colonocyte metabolism shapes the gut microbiota. Science 362, eaat9076. https://doi.org/10.1126/science.aat9076

Litvak, Y., Mon, K.K.Z., Nguyen, H., Chanthavixay, G., Liou, M., Velazquez, E.M., Kutter, L., Alcantara, M.A., Byndloss, M.X., Tiffany, C.R., Walker, G.T., Faber, F., Zhu, Y., Bronner, D.N., Byndloss, A.J., Tsolis, R.M., Zhou, H., Bäumler, A.J., 2019. Commensal Enterobacteriaceae Protect against Salmonella Colonization through Oxygen Competition. Cell Host Microbe 25, 128-139.e5. https://doi.org/10.1016/j.chom.2018.12.003

Lu, H.-P., Lai, Y.-C., Huang, S.-W., Chen, H.-C., Hsieh, C., Yu, H.-T., 2014. Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. Sci. Rep. 4, 6185. https://doi.org/10.1038/srep06185

Lu, L., Walker, W.A., 2001. Pathologic and physiologic interactions of bacteria with the gastrointestinal epithelium. Am. J. Clin. Nutr. 73, 1124S-1130S. https://doi.org/10.1093/ajcn/73.6.1124S

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., Dore, J., 2006. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 55, 205–211. https://doi.org/10.1136/gut.2005.073817

Marchesi, J.R., Ravel, J., 2015. The vocabulary of microbiome research: a proposal. Microbiome 3, 31. https://doi.org/10.1186/s40168-015-0094-5

Massip, C., Branchu, P., Bossuet-Greif, N., Chagneau, C.V., Gaillard, D., Martin, P., Boury, M., Sécher, T., Dubois, D., Nougayrède, J.-P., Oswald, E., 2019. Deciphering the interplay between the genotoxic and probiotic activities of Escherichia coli Nissle 1917. PLOS Pathog. 15, e1008029. https://doi.org/10.1371/journal.ppat.1008029

Matyjaszkiewicz, A., Fiore, G., Annunziata, F., Grierson, C.S., Savery, N.J., Marucci, L., di Bernardo, M., 2017. BSim 2.0: An Advanced Agent-Based Cell Simulator. ACS Synth. Biol. 6, 1969–1972. https://doi.org/10.1021/acssynbio.7b00121

Miller, H.R.P., Nawa, Y., 1979. Nippostrongylus brasiliensis: Intestinal goblet-cell response in adoptively immunized rats. Exp. Parasitol. 47, 81–90. https://doi.org/10.1016/0014-4894(79)90010-9

Morgenroth, E., Wilderer, P.A., 2000. Influence of detachment mechanisms on competition in biofilms. Water Res. 34, 417–426. https://doi.org/10.1016/S0043-1354(99)00157-8

Munoz-Tamayo, R., Laroche, B., Walter, E., Dore, J., Leclerc, M., 2010. Mathematical modelling of carbohydrate degradation by human colonic microbiota. J. Theor. Biol. 266, 189–201. https://doi.org/10.1016/j.jtbi.2010.05.040

Mutaflor, n.d. Medicinal product safety [WWW Document]. Mutaflor® – Unique Probiotic Act. Ingred. Escherichia Coli Strain Nissle 1917 Hum. Gut Health. URL https://www.mutaflor.com/mutaflor-with-escherichia-coli-strain-nissle-1917-for-human-gut-health/medicinal-product-safety.html (accessed 2.7.23).

Naylor, J., Fellermann, H., Ding, Y., Mohammed, W.K., Jakubovics, N.S., Mukherjee, J., Biggs, C.A., Wright, P.C., Krasnogor, N., 2017. Simbiotics: A Multiscale Integrative Platform for 3D Modeling of Bacterial Populations. ACS Synth. Biol. 6, 1194–1210. https://doi.org/10.1021/acssynbio.6b00315

Nguyen, T.L.A., Vieira-Silva, S., Liston, A., Raes, J., 2015. How informative is the mouse for human gut microbiota research? Dis. Model. Mech. 8, 1–16. https://doi.org/10.1242/dmm.017400

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-gut microbiota metabolic interactions. Science 336, 1262–1267. https://doi.org/10.1126/science.1223813

Noguera, D.R., Picioreanu, C., 2004. Results from the multi-species Benchmark Problem 3 (BM3) using two-dimensional models. Water Sci. Technol. 49, 169–176.

Noguera, D.R., Pizarro, G.E., Regan, J.M., 2004. Modeling Biofilms, in: Microbial Biofilms. John Wiley & Sons, Ltd, pp. 222–249. https://doi.org/10.1128/9781555817718.ch13

Nougayrède, J.-P., Chagneau, C.V., Motta, J.-P., Bossuet-Greif, N., Belloy, M., Taieb, F., Gratadoux, J.-J., Thomas, M., Langella, P., Oswald, E., 2021. A Toxic Friend: Genotoxic and Mutagenic Activity of the Probiotic Strain Escherichia coli Nissle 1917. mSphere 6, e00624-21. https://doi.org/10.1128/mSphere.00624-21

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248. https://doi.org/10.1038/nbt.1614

Ovalle, W., Nahirney, P., 2021. Lower Digestive System, in: Netter's Essential Histology. Elsevier.

Panigrahi, P., Parida, S., Nanda, N.C., Satpathy, R., Pradhan, L., Chandel, D.S., Baccaglini, L., Mohapatra, A., Mohapatra, S.S., Misra, P.R., Chaudhry, R., Chen, H.H., Johnson, J.A., Morris, J.G., Paneth, N., Gewolb, I.H., 2017. A randomized synbiotic trial to prevent sepsis among infants in rural India. Nature 548, 407–412.

Panigrahi, P., Parida, S., Pradhan, L., Mohapatra, S.S., Misra, P.R., Johnson, J.A., Chaudhry, R., Taylor, S., Hansen, N.I., Gewolb, I.H., 2008. Long-term Colonization of a Lactobacillus plantarum Synbiotic Preparation in the Neonatal Gut. J. Pediatr. Gastroenterol. Nutr. 47, 45–53. https://doi.org/10.1097/MPG.0b013e31815a5f2c

Panzer, A.R., Lynch, S.V., 2015. Influence and effect of the human microbiome in allergy and asthma. Curr. Opin. Rheumatol. 27, 373–380. https://doi.org/10.1097/BOR.0000000000000191

Picioreanu, C., Kreft, J.-U., van Loosdrecht, M.C.M., 2004. Particle-based multidimensional multispecies biofilm model. Appl. Environ. Microbiol. 70, 3024–3040. https://doi.org/10.1128/AEM.70.5.3024-3040.2004

Pizarro, G.E., Garcia, C., Moreno, R., Sepulveda, M.E., 2004. Two-dimensional cellular automaton model for mixed-culture biofilm. Water Sci. Technol. 49, 193–198. https://doi.org/10.2166/wst.2004.0839

Plugge, C.M., 2005. Anoxic Media Design, Preparation, and Considerations, in: Methods in Enzymology, Environmental Microbiology. Academic Press, pp. 3–16. https://doi.org/10.1016/S0076-6879(05)97001-8

Press, W.H., Teukolsky, S., Vetterling, W.T., Flannery, B.P., 1993. Numerical Recipes in C. Cambridge University Press, Cambridge.

Ragsdale, S.W., Pierce, E., 2008. Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation. Biochim. Biophys. Acta 1784, 1873–1898. https://doi.org/10.1016/j.bbapap.2008.08.012

Ravcheev, D.A., Thiele, I., 2017. Comparative Genomic Analysis of the Human Gut Microbiome Reveals a Broad Distribution of Metabolic Pathways for the Degradation of Host-Synthetized Mucin Glycans and Utilization of Mucin-Derived Monosaccharides. Front. Genet. 8.

Reichert, P., 1994. Aquasim: A tool for simulation and data analysis of aquatic systems. Water Sci. Technol. 30, 21–30.

Reichert, P., Wanner, O., 1997. Movement of solids in biofilms: significance of liquid phase transport. Water Sci. Technol. 36, 321–328.

Reissbrodt, R., Hammes, W.P., Dal Bello, F., Prager, R., Fruth, A., Hantke, K., Rakin, A., Starcic-Erjavec, M., Williams, P.H., 2009. Inhibition of growth of Shiga toxin-producing Escherichia coli by nonpathogenic Escherichia coli. FEMS Microbiol. Lett. 290, 62–69. https://doi.org/10.1111/j.1574-6968.2008.01405.x

Rittmann, B.E., Schwarz, A.O., Eberl, H., Morgenroth, E., Peréz, J., Van Loosdrecht, M.C.M., Wanner, O., 2004. Results from the multi-species Benchmark Problem (BM3) using one-dimensional models. Water Sci. Technol. 49, 163–168.

Robbe, C., Capon, C., Coddeville, B., Michalski, J.-C., 2004. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. Biochem. J. 384, 307–316. https://doi.org/10.1042/BJ20040605

Robinson, N.B., Krieger, Katherine, Khan, F.M., Huffman, W., Chang, M., Naik, A., Yongle, R., Hameed, I., Krieger, Karl, Girardi, L.N., Gaudino, M., 2019. The current state of animal models in research: A review. Int. J. Surg. 72, 9–13. https://doi.org/10.1016/j.ijsu.2019.10.015

Roediger, W.E., 1980. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut 21, 793–798. https://doi.org/10.1136/gut.21.9.793

Rosenthal, N., Brown, S., 2007. The mouse ascending: perspectives for human-disease models. Nat. Cell Biol. 9, 993–999. https://doi.org/10.1038/ncb437

Rudge, T.J., Steiner, P.J., Phillips, A., Haseloff, J., 2012. Computational modeling of synthetic microbial biofilms. ACS Synth. Biol. 1, 345–352. https://doi.org/10.1021/sb300031n

Sadahiro, S., Ohmura, T., Yamada, Y., Saito, T., Taki, Y., 1992. Analysis of length and surface area of each segment of the large intestine according to age, sex and physique. Surg. Radiol. Anat. 14, 251–257. https://doi.org/10.1007/BF01794949

Sakaguchi, E., 2003. Digestive strategies of small hindgut fermenters. Anim. Sci. J. 74, 327–337. https://doi.org/10.1046/j.1344-3941.2003.00124.x

Salzman, N.H., 2011. Microbiota–immune system interaction: an uneasy alliance. Curr. Opin. Microbiol., Host–microbe interactions 14, 99–105. https://doi.org/10.1016/j.mib.2010.09.018

Sassone-Corsi, M., Nuccio, S.-P., Liu, H., Hernandez, D., Vu, C.T., Takahashi, A.A., Edwards, R.A., Raffatellu, M., 2016. Microcins mediate competition among Enterobacteriaceae in the inflamed gut. Nature 540, 280–283.

Schaechter, M., Maaløe, O., Kjeldgaard, N.O., 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. J. Gen. Microbiol. 19, 592–606. https://doi.org/10.1099/00221287-19-3-592

Schink, B., 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. 61, 262–280.

Sender, R., Fuchs, S., Milo, R., 2016. Revised estimates for the number of human and bacteria cells in the body. bioRxiv 036103. https://doi.org/10.1101/036103

Shah, P., Fritz, J.V., Glaab, E., Desai, M.S., Greenhalgh, K., Frachet, A., Niegowska, M., Estes, M., Jäger, C., Seguin-Devaux, C., Zenhausern, F., Wilmes, P., 2016. A microfluidics-based in vitro model of the gastrointestinal human-microbe interface. Nat. Commun. 7, 11535. https://doi.org/10.1038/ncomms11535

Shen, J., Obin, M.S., Zhao, L., 2013. The gut microbiota, obesity and insulin resistance. Mol. Aspects Med., Obesity 34, 39–58. https://doi.org/10.1016/j.mam.2012.11.001

Smith, W.P.J., Davit, Y., Osborne, J.M., Kim, W., Foster, K.R., Pitt-Francis, J.M., 2017. Cell morphology drives spatial patterning in microbial communities. Proc. Natl. Acad. Sci. 114, E280–E286. https://doi.org/10.1073/pnas.1613007114

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottière, H.M., Doré, J., Marteau, P., Seksik, P., Langella, P., 2008. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc. Natl. Acad. Sci. U. S. A. 105, 16731–16736. https://doi.org/10.1073/pnas.0804812105

Sokol, H., Seksik, P., Furet, J.P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., Cosnes, J., Corthier, G., Marteau, P., Doré, J., 2009. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm. Bowel Dis. 15, 1183–1189. https://doi.org/10.1002/ibd.20903

Sommer, F., Bäckhed, F., 2016. Know your neighbor: Microbiota and host epithelial cells interact locally to control intestinal function and physiology. BioEssays. https://doi.org/10.1002/bies.201500151

Sonnenborn, U., 2016. Escherichia coli strain Nissle 1917—from bench to bedside and back: history of a special Escherichia coli strain with probiotic properties. FEMS Microbiol. Lett. 363. https://doi.org/10.1093/femsle/fnw212

Sonnenborn, U., Schulze, J., 2009. The non-pathogenic Escherichia coli strain Nissle 1917 – features of a versatile probiotic. Microb. Ecol. Health Dis. 21, 122–158. https://doi.org/10.3109/08910600903444267

Steinert, M., Hentschel, U., Hacker, J., 2000. Symbiosis and pathogenesis: evolution of the microbe-host interaction. Naturwissenschaften 87, 1–11. https://doi.org/10.1007/s001140050001

Stokes, C.R., 2017. The development and role of microbial-host interactions in gut mucosal immune development. J. Anim. Sci. Biotechnol. 8, 12. https://doi.org/10.1186/s40104-016-0138-0

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., Barratt, M.J., VanArendonk, L.G., Zhang, Q., Province, M.A., Petri Jr, W.A., Ahmed, T., Gordon, J.I., 2014. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature. https://doi.org/10.1038/nature13421

Theriot, C.M., Koenigsknecht, M.J., Carlson Jr, P.E., Hatton, G.E., Nelson, A.M., Li, B., Huffnagle, G.B., Z Li, J., Young, V.B., 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat. Commun. 5. https://doi.org/10.1038/ncomms4114

Troeger, C., Forouzanfar, M., Rao, P.C., Khalil, I., Brown, A., Reiner, R.C., Fullman, N., Thompson, R.L., Abajobir, A., Ahmed, M., Alemayohu, M.A., Alvis-Guzman, N., Amare, A.T., Antonio, C.A., Asayesh, H., Avokpaho, E., Awasthi, A., Bacha, U., Barac, A., Betsue, B.D., Beyene, A.S., Boneya, D.J., Malta, D.C., Dandona, L., Dandona, R., Dubey, M., Eshrati, B., Fitchett, J.R.A., Gebrehiwot, T.T., Hailu, G.B., Horino, M., Hotez, P.J., Jibat, T., Jonas, J.B., Kasaeian, A., Kissoon, N., Kotloff, K., Koyanagi, A., Kumar, G.A., Rai, R.K., Lal, A., Razek, H.M.A.E., Mengistie, M.A., Moe, C., Patton, G., Platts-Mills, J.A., Qorbani, M., Ram, U., Roba, H.S., Sanabria, J., Sartorius, B., Sawhney, M., Shigematsu, M., Sreeramareddy, C., Swaminathan, S., Tedla, B.A., Jagiellonian, R.T.-M., Ukwaja, K., Werdecker, A., Widdowson, M.-A., Yonemoto, N., Zaki, M.E.S., Lim, S.S., Naghavi, M., Vos, T., Hay, S.I., Murray, C.J.L., Mokdad, A.H., 2017. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Infect. Dis. 17, 909–948. https://doi.org/10.1016/S1473-3099(17)30276-1

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., Gordon, J.I., 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–1031.

Ushijima, T., Seto, A., 1991. Selected Fecal Bacteria and Nutrients Essential for Antagonism of Salmonella-Typhimurium in Anaerobic Continuous-Flow Cultures. J. Med. Microbiol. 35, 111–117.

van der Ark, K.C.H., Aalvink, S., Suarez-Diez, M., Schaap, P.J., de Vos, W.M., Belzer, C., 2018. Model-driven design of a minimal medium for *Akkermansia muciniphila* confirms mucus adaptation. Microb. Biotechnol. 11, 476–485. https://doi.org/10.1111/1751-7915.13033

van der Waaij, D., Vries, J.M.B., der Wees, J.E.C.L., 1971. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. J. Hyg. (Lond.) 69, 405–411.

Vermeiren, J., Van den Abbeele, P., Laukens, D., Vigsnaes, L.K., De Vos, M., Boon, N., Van de Wiele, T., 2012. Decreased colonization of fecal Clostridium coccoides/Eubacterium rectale species from ulcerative colitis patients in an in vitro

dynamic gut model with mucin environment. FEMS Microbiol. Ecol. 79, 685–696. https://doi.org/10.1111/j.1574-6941.2011.01252.x
Wanner, O., Eberl, H.J., Morgenroth, E., Noguera, D.R., Picioreanu, C., Rittmann, B.E., Van Loosdrecht, M.C.M., 2006. Mathematical modeling of biofilms. IWA Publishing, London.
Wanner, O., Gujer, W., 1986. A multispecies biofilm model. Biotechnol. Bioeng. 28, 314–328. https://doi.org/10.1002/bit.260280304
Wanner, O., Reichert, P., 1996. Mathematical modeling of mixed-culture biofilm. Biotechnol. Bioeng. 49, 172–184.
Wilensky, U., 2001. Modeling Nature's Emergent Patterns with Multi-agent Languages. Proc. EuroLogo 2001 16.
Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Järnerot, G., Engstrand, L., Tysk, C., Jansson, J.K., 2009. Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. Inflamm. Bowel Dis. 15, 653–660. https://doi.org/10.1002/ibd.20783
Wilson, M., 2008. Bacteriology of Humans: an Ecological Perspective. Blackwell, Oxford.
Zhu, D., Sorg, J.A., Sun, X., 2018. Clostridioides difficile Biology: Sporulation, Germination, and Corresponding Therapies for C. difficile Infection. Front. Cell. Infect. Microbiol. 8.