Process-based modelling of microbial community dynamics in the human colon

Helen Kettle, Petra Louis and Harry J. FLint

August 10, 2022

Abstract

The human colon contains a dynamic microbial community whose composition has important implications for human health. In this work we build a process-based model of the colonic microbial ecosystem and compare with general empirical observations and the results of in-vivo experiments. Based our previous work (**?**), the microbial model consists of 10 microbial functional groups, 4 substrates and 10 metabolites; to this we add the interaction with a human host to give simulations of the in-situ colonic microbial ecosystem. This model incorporates absorption of short chain fatty acids (SCFA) and water by the host through the gut wall, variations in incoming dietary substrates (in the form of "meals" whose composition varies in time), bowel movements, feedback on microbial growth from changes in pH resulting from SCFA production, and multiple compartments to represent the proximal, transverse and distal colon. We verify our model against a number of observed criteria, e.g. total SCFA concentrations, SCFA ratios, mass of bowel movements, pH and water absorption over the transit time; and then run simulations investigating the effect of colonic transit time, and the composition and amount of indigestible carbohydrate in the host diet, which we compare with in-vivo studies. Gut microbiota are highly complex and poory understood yet our work shows that it is nevertheless possible to develop predictive models of the key components of the dynamics of this ecological system. The code is available as an R package (microPopGut) to aid future research.

The human colon harbours a dense and diverse community of microbiota whose interactions with the host can have a profound effect on human health (e.g. **?**, **?**). Due to the location of this community within its host, data collection and experimentation are problematic. Information on this system must come from volunteer experiments in which diet and stool samples are monitored or from laboratory experiments using the microbes found in stool samples. Another approach is to put current knowledge into a mathematical framework and run simulations of the system to test our understanding and identify knowledge gaps. To this end a number of mathematical models of this system have been developed - e.g. **?**, **?**, **?**, **?**, **?**.

When developing a model, a number of assumptions about the system are made in order to reduce complexity/dimensionality so that the model is easier to parameterise, run and analyse. Some modellers choose to reduce the microbial complexity and focus on the physics of the gut (e.g. **?**, **?**), some try to achieve a

balance of both (e.g. **?**) and some choose to develop the microbial community (e.g. **?**). The model described here focuses on the microbial community dynamics and on interactions with the host, with a fairly simple model of the colon. We include the simulation of 'meals' (of random composition and size) arriving at the colon and look at the effects of bowel movements, both of which, as far as we are aware, have not been previously incorporated into such models. Having developed a complex model of human gut microbiota in a fermentor system (**?**), and publicly available software (microPop - an R package for modelling microbial communities (**?**)) we now incorporate this 10-group microbial ecosystem model (Table **??**) into a model of the human gut in order to simulate the effects of diet and host on the microbial composition and subsequent short chain fatty acid (SCFA) production.

Approximately 95% of the SCFA produced by the microbes during growth are absorbed by the host through the gut wall and it is the ratio of the 3 main SCFAs (acetate, butyrate and propionate) which is known to have a significant effect on human health. Thus, we prioritise information on the values of these ratios in our model verification. Similarly approximately 90% of the water flowing into the colon is absorbed. Changes in the volume of water have a significant effect on the concentration of the molecules in the colon which in turn affects pH which then affects microbial growth, all of which are included in our model.

Due to its shape within the body, the colon is commonly divided into 3 different regions - the proximal, transverse and distal sections running from beginning to end (Fig. **??**A and B). The availability of substrate, microbial growth and hence pH vary along the colon, therefore, although our model is not spatial we simulate these three regions explicitly, with flow from one to another. Furthermore, as well as incorporating varying substrate inflow in the form of meals we also add in the release of mucins along the length of the colon which can be microbially broken down to release proteins and carbohydrates, allowing for further microbial growth away from the beginning of the colon where the substrates enter. A graphical summary of the model is shown in Fig. **??**.

We use the following criteria to verify our model captures the main features established for the system:

1. Total SCFA (TSCFA) concentration in the proximal, transverse and distal compartments should be around 123, 117 and 80 mM respectively according to sudden death human autopsies (**?**)

2. Acetate:Propionate:Butyrate ratios are similar (around 3:1:1) in all regions of the colon and around 60:20:20 mM (**?**)

3. Over 95% of SCFA are absorbed by the host (**?**)

4. Approx. 90% of incoming water is absorbed by the host (**?**)

5. pH in the proximal, transverse and distal compartments should be around 5.7, 6.2 and 6.6 respectively (**?**)

6. Normal daily fecal output in Britain is 100-200 g d^{-1} of which 25-50 g is solid matter (i.e. 50-175 g d^{-1} is water). Bacteria make up about 55% of the solid matter i.e. 14-28 g d^{-1} of microbes emitted (**?**).

Figure 1: Colon schematic plus table of typical values for physical properties (length, volume, pH and TSCFA) and plots of summarised model simulations for average TSCFA and pH for comparison with typical values.

Figure 2: Model system with the microbial ecosystem comprising 10 microbial functional groups (Table **??**) which consume substrates (RS, NSP and protein) and water. The microbes produce metabolites some of which are consumed by other MFGs ('cross-feeding'). SCFA and water are absorbed through the colon wall (at a different specific rates). The system shown within the dashed line is repeated in each of the modelled regions of the colon (proximal, transerve, and distal) with the contents of the previous region, flowing into the next. The first compartment (proximal) has inflow from the small intestine - this can be constant inflow or simulated meals whose composition varies randomly in time. The third model compartment (distal) has outflow to stool which can be constant or evacuation via bowel movements can be simulated. pH varies with the TSCFA concentration and affects the rate of microbial growth differently for each MFG.

7. TSCFA concentration decreases with transit time (**?**)

After model verification we examine the effects of including meals, bowel movements and fixed/varying pH into the model. We then use the model to look at how carbohydrate composition (based on the fractions of resistant starch (RS) and non-starch polysaccharides (NSP)) and total carbohydrate affect the microbial community and SCFA composition. The simulations are then compared with in-vivo data from human volunteer experiments.

Although gut microbiota are highly complex and not fully understood, here we show that it is nonetheless possible to develop predictive models of key components of this ecological system. Our results show promise and we believe this model represents a significant step forward in this field. We refer to the model as "microPopGut" and to aid future research the code is available as an R-package on github (https://github.com/HelenKettle/microPopGut) and instructions on how to use the package are given in the supplementary file 'gettingStartedWithMicroPopGut.pdf'.

Results

Standard Model

Having established the default model settings and parameter values which give the best fit to our criteria (see Table **??** for colon parameters and Supp. Info. (section 3) for microbial group parameters) we then investigate the effects of different model configurations, e.g. with/without bowel movements, meals and variable pH, for a range of transit times. Simulations with meals have a random

Table 1: Microbial functional groups included in the model (and the R package microPop (**?**)) and described by **?**. Users should be aware that the parameter values given in the data frames in the software will almost certainly change with increasing knowledge of gut microbiota and in some cases are simply a "best guess".

microPop Name	Abbr.	Description	Examples
Bacteroides	B	Acetate-propionate-succinate group	*Bacteroides* spp., *Prevotella* spp., *Akkermansia muciniphila* (Verrucomicrobia)
NoButyStarchDeg	NBSD	Non-butyrate-forming starch degraders	*Ruminococcaceae* related to *Ruminococcus bromii*. Also includes certain *Lachnospiraceae*
NoButyFibreDeg	NBFD	Non-butyrate-forming fibre degraders	*Ruminococcaceae* related to *Ruminococcus albus*, *Ruminococcus flavefaciens*. Also includes certain *Lachnospiraceae*
LactateProducers	LP	Lactate producers	*Actinobacteria*, especially *Bifidobacterium* spp., *Collinsella aerofaciens*
ButyrateProducers1	BP1	Butyrate Producers	*Lachnospiraceae* related to *Eubacterium rectale*, *Roseburia* spp.
ButyrateProducers2	BP2	Butyrate Producers	Certain *Ruminococcaceae*, in particular *Faecalibacterium prausnitzii*
PropionateProducers	PP	Propionate producers	*Veillonellaceae* e.g. *Veillonella* spp., *Megasphaera elsdenii*
ButyrateProducers3	BP3	Butyrate Producers	*Lachnospiraceae* related to *Eubacterium hallii*, *Anaerostipes* spp.
Acetogens	A	Acetate Producers	Certain *Lachnospiraceae*, e.g. *Blautia hydrogenotrophica*
Methanogens	M	Methanogenic archaea	*Methanobrevibacter smithii*

component therefore the model is run for a number of different starting seed values. Due to the random fluctuations these simulations will not reach steady state therefore the summary values are taken as the mean from day 7 (to remove the effect of the initial conditions) to the end of the simulation (28 days) and are averaged over multiple seeds.

Table **??** gives summary results of the model simulations without bowel movements but with varying pH for each bowel region. Fig. **??** shows results from more simulations but for the distal colon only. Fig. **??**a shows that although bowel movements make a difference to the total biomass and the TSCFA they do not have a large effect on the community composition or the SCFA ratios. Thus in the interests of model simplicity we decide to not include bowel movements in later simulations. However, varying pH with TSCFA can be seen to make a very large difference to the microbial community (Fig. **??**b) and also improves the SCFA ratios with respect to our verification criteria. The addition of meals makes a significant difference which increases with increasing transit time (Fig. **??**c). In Fig. **??** the time series output from the model shows how the meals-inflow allows the community to experience large shifts over time (on a much longer time scale than the variations in the input), as opposed to the

fixed state approached using a constant substrate inflow.

Fig. **??**C shows the average pH and TSCFA for the proximal, transverse and distal compartments. It can be seen that blue (meals) and red (continuous inflow) dots show the same basic trends. The decrease in TSCFA with transit time has been shown experimentally (**?**); in section 2 of the Supp. Info. we suggest a mathematical explanation for this based on the supposition that the specific rate of absorption of water through the gut wall is slower than that for SCFA.

Regarding Table **??**, for some criteria, e.g. pH, the continuous inflow setting gives results closer to our verification values, but in other cases, e.g. A:B:P in distal colon, simulating meals gives closer results. Note that we consider a transit time of 1 day the most typical of the three transit times, and the one that should be compared with our verification criteria, the others are included to show the variation in results. Ideally TSCFA should be 123, 117 and 80 mM for prox., trans., dist. but the best match we have to this is for a 3 d transit time and continuous inflow. This is most likely due to the fact that our model has fixed rates of specific absorption of SCFA and water throughout the colon. However, our TSCFA values are within a reasonable range and display the general trend of decreasing TSCFA from the proximal to distal colon. The microbe output, i.e. the outflow of fecal microbes is steady at around 20 g d^{-1} in all cases which fits well in the verification range (14-28 g d^{-1}). The water fraction is the ratio of the rate of fecal water over the rate of water flowing into the colon, since 90% of water is absorbed this should be 0.1. This is approximately correct for our 1 d simulations (0.14) but, as expected, when transit time increases this decreases significantly. In summary, comparing these simulation results with our list of model verification criteria shows that in general our model is fit for purpose, and that the inclusion of meals-inflow and varying pH improve our simulations.

Table 2: Summary of model results (for comparison with our list of criteria) for 3 different transit times, with meals or continuous inflow and with pH varying with TSCFA. Microbe output is the mass of microbes leaving the colon per day and the water fraction is amount of water leaving the colon per day divided by the amount entering. All simulations were run for 28 days and the results shown are the average over days 7-28. The results for the simulations with meals are averaged over 4 random seeds. 'A:B:P dist' refers to the Acetate:Butyrate:Propionate ratio (mM) in the distal colon.

	Meals			Continuous inflow		
transit time	1d	2d	3d	1d	2d	3d
TSCFA prox (mM)	115.3	110.3	105.2	124.4	123.5	122.1
TSCFA trans (mM)	102.1	88.1	83.4	75.5	96.1	111.1
TSCFA dist (mM)	107.6	64.8	69.0	89.5	62.0	83.3
A:B:P dist (mM)	62:28:17	31:23:10	34:23:12	56:27:7	38:18:6	59:17:7
pH prox	6.0	6.1	6.2	5.9	5.9	5.9
pH trans	6.2	6.5	6.6	6.7	6.4	6.1
pH dist	6.2	6.9	6.8	6.5	6.9	6.6
microbe output (g d^{-1})	20.2	20.1	20.1	20.0	20.1	20.0
water fraction	0.14	0.04	0.02	0.14	0.04	0.02

(a) Bowel Movements (b) pH (c) Transit time

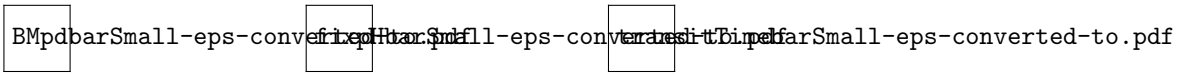

Figure 3: Summary results (averaged over days 7-28 and over random seeds) for the distal compartment for continuous inflow or fluctuating inflow (i.e. 'meals') for continuous outflow from colon or for 2 bowel movements per day ('2 BM/d'). The RS fraction is 0.78 (i.e. 78% of the dietary carbohyrate is resistant starch and 22% is NSP) and the transit time is 0.93 d for a), 1.25 d for b) and at 0.25, 0.5, 1, 1.5, 2, 2.5 and 3 days for c). The top row shows the biomass of each group, the bottom row shows the SCFA.

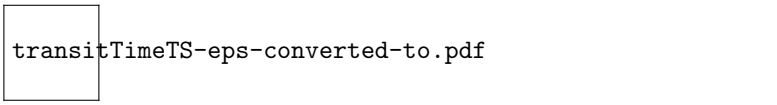

Figure 4: Simulation results for the distal compartment for continuous inflow (first plot on each row) or fluctuating inflow (i.e. 'meals') for transit times of 1 d (top row) and 2 d (bottom row) and for 2 random seeds. Modelled pH varies with TSCFA and the RS fraction is 0.78. There are no bowel movements (i.e. outflow is continuous). See Table **??** for microbial groups.

Model Experiments

We now use our model to simulate two scenarios – firstly, the effects of decreasing total carbohydrate intake and secondly, the effects of changing carbohydrate composition (whilst keeping total intake fixed) on the microbial commnuity and associated SCFA production. Comparing our simulations with data from human volunteer experiments is not straightforward since in order to run our model, ingested food must be translated to substrates reaching the colon. This is problematic due to unknown water consumption and transit times and uncertainties associated with the absorption rates of the ingested carbohydrate and protein higher up the digestive tract. Thus we do not attempt to reproduce human experiments but rather we run simulations based on variations to our standard model set up and then compare our results qualitatively with available data.

Effects of total dietary carbohydrate

In this model experiment we investigate the effects of decreasing carbohydrate on the microbial community. Here we compare our results qualtitatively with the human dietary study of **?** which explored the impacts of carefully controlled decreases in carbohydrate intake upon weight loss and microbial fermentation products in obese subjects using 3 diets – a maintenance (M) diet, a high protein, moderate carbohydrate diet (HPMC) and a high protein, low carbohydrate diet (HPLC) (see Fig. **??** for details). This is of course, the composition for ingested food, which is not easily translated into substrate concentrations entering the colon. However, we can look at the general trends in SCFA and microbial composition with changing colonic carbohydrate intake rate. Thus, in these model experiments we keep protein inflow to the colon at 10 g d^{-1} (our default

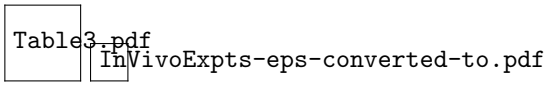

Figure 5: Table on left shows the dietary intake for two human studies (**?** and **?**. PI, CI, SI and NSPI refer to ingested dietary protein, carbohydrate, starch and NSP. Note, starch value for the high RS diet in the **?** study included 26 g commercial RS. Bar plots show SCFA data from these studies.

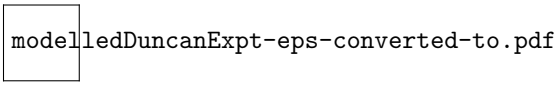

Figure 6: Simulated Biomass and SCFA results for increasing carbohydrate inflow. Simulations are run with continuous substrate inflow (cts) and with 'meals' for a transit time of 1.5 days. The results are the average over the last 3 weeks of a 28 day simulation and 'meals' is the average over 4 stochastically-generated simulations.

value) and then increase inflowing carbohydrate from 10 g d^{-1} to 60 g d^{-1} in 10 g d^{-1} intervals. To include the effects of different carbohydrate composition we run the model for an resistant starch (RS) fraction of either 0.2 or 0.78 (the default value), with non-starch polysaccharides (NSP) making up the remaining carbohydrate in each case. Although subject to large uncertainties, we estimate the RS fractions for the **?** experiments of 0-0.6 (M diet), 0-0.68 (HPMC) and 0-0.12 (HPLC) (based on RS is 0–20% of ingested starch (**?**) and bio-available NSP is 75% of ingested NSP (**?**)). Due to the low fibre nature of many of these simulations we run the model with a slightly longer transit time of 1.5 d and for both continuous inflow and meals.

Fig. **??** shows the SCFA results from our model experiment and Fig. **??** shows the results from the in vivo experiment. It is very clear, from both the model and in vivo results that the proportion of butyrate increases as the amount of carbohydrate in the diet increases. Furthermore, both model and in vivo results show an increase in TSCFA with carbohydrate intake rate. Since **?** also look at the relationship between butyrate concentration and grams of carbohydrate eaten per day we plot butyrate against carbohydrate entering the colon (Fig. **??**) to compare with their Fig. 1. In both cases, butyrate concentration increases with incoming carbohydrate. Furthermore, as seen in both the model and the data, the percentage of butyrate increases with carbohydrate intake (Fig. **??**).

In terms of microbial composition, Fig. **??** shows the results from our simulations are reasonably consistent across inflow type (meals or continuous), with B dominating at low carbohydrate intake. When the RS fraction is low (i.e. when carbohydrate is made up of 80% NSP) then NBFD increase with increased C intake. Whereas when C is mostly RS then NBSD and BP1 increase with C. In both cases BP2 increase with increasing C intake.

Effects of carbohydrate composition

Here we use the model to simulate the effects of changing carbohydrate composition on the microbial community composition by changing the ratio of RS to